UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"

SCUOLA DI DOTTORATO DI RICERCA IN FISICA
CICLO XXXIII

MEASUREMENT OF THE CP-VIOLATING PHASE ϕ_s IN MUON-TAGGED $B_s^0 \to J/\psi\,\phi$ DECAYS WITH THE CMS EXPERIMENT

ALBERTO BRAGAGNOLO

SUPERVISORE:
PROF. MARTINO MARGONI

CO-SUPERVISORE:
PROF. FRANCO SIMONETTO

ANNO ACCADEMICO 2019–2020

Alberto Bragagnolo: *Measurement of the CP-violating phase ϕ_s in muon-tagged $B^0_s \to J/\psi\,\phi$ decays with the CMS experiment* © February 2021

TESI DI DOTTORATO DI RICERCA IN FISICA
Università degli Studi di Padova
Dipartimento di Fisica e Astronomia "Galileo Galilei"
Scuola di Dottorato di Ricerca in Fisica

SUPERVISORE:
Prof. Martino Margoni
CO-SUPERVISORE:
Prof. Franco Simonetto

ABSTRACT

This dissertation presents a study of the $B^0_s \to J/\psi\,\phi$ decay performed with the CMS experiment and aimed at measuring the CP-violating phase ϕ_s. This weak phase parametrizes the CP violation in the interference between direct B^0_s meson decays to a CP eigenstate via $b \to c\bar{c}s$ transitions and decays through mixing to the same final state. The Standard Model of particle physics relates ϕ_s to the elements of the Cabibbo–Kobayashi–Maskawa matrix, which are precisely constrained by global fits to experimental data. Due to its clean theoretical prediction, the CP-violating phase ϕ_s is extremely sensitive to new-physics effects that can modify its value via the contribution of non-SM particles to the B^0_s–$\overline{B}^0_s$ mixing. Thus, the precise measurement of this phase is a powerful test to either establish physics phenomena beyond the Standard model or put stringent constraints on them. Moreover, by analyzing the $B^0_s \to J/\psi\,\phi$ decay rate, several other interesting physics parameters can be measured, such as the decay width difference $\Delta\Gamma_s$ between the light and heavy B^0_s mass eigenstates, the average decay width Γ_s, the mass splitting Δm_s, and the CP observable $|\lambda|$ that describes the amount of direct CP violation in the process. Therefore, this measurement consists also of a comprehensive test of the phenomenology of the B^0_s system.

To measure the physics parameters, a time-dependent and flavor-tagged angular analysis of the $B^0_s \to J/\psi\,\phi$ decay rate is performed. The measurement is based on a sample of 48 500 reconstructed $B^0_s \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ events collected in proton-proton collisions at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of $\mathcal{L}_{\text{int}} = 96.4\,\text{fb}^{-1}$. Compared to the previous CMS measurement at $\sqrt{s} = 8$ TeV, this analysis benefits from the increase in the center-of-mass energy from 8 to 13 TeV that nearly doubles the B^0_s production cross section. Events are collected with a trigger that requires a $J/\psi \to \mu^+\mu^-$ candidate plus an additional third muon, which is exploited to infer the flavor of the B^0_s meson at production time. A novel opposite-side muon tagger based on calibrated deep neural networks has been developed to maximize the analysis sensitivity. A high tagging power of ≈10%, corresponding to a tagging efficiency of ≈50%, is achieved, aided by the requirement of an additional muon imposed at the trigger level.

The CP-violating phase ϕ_s is measured to be:

$$\phi_s = -11 \pm 50\,(\text{stat}) \pm 10\,(\text{syst})\ \text{mrad}\,,$$

in agreement both with the Standard Model predictions and the absence of CP violation in the decay-mixing interference. These results are significantly more precise than those from the previous CMS measurement at 8 TeV. This analysis also represents the first CMS measurement of the mass splitting Δm_s and the direct CP observable $|\lambda|$. The results are further combined with those obtained by CMS at $\sqrt{s} = 8$ TeV, yielding $\phi_s = -21 \pm 44\,(\text{stat}) \pm 10\,(\text{syst})$ mrad.

SINTESI

Questa tesi presenta uno studio del decadimento $B^0_s \to J/\psi\,\phi$ effettuato con l'esperimento CMS e mirato alla misura della fase debole ϕ_s. Questa fase parametrizza la violazione di CP nell'interferenza tra decadimenti diretti di mesoni B^0_s in autostati di CP e decadimenti allo stesso stato finale attraverso oscillazioni di sapore. Il Modello Standard della fisica delle particelle mette in relazione ϕ_s con gli elementi della matrice Cabibbo–Kobayashi–Maskawa, i quali sono precisamente determinati da fit globali ai dati sperimentali. Data la sua precisa previsione teorica, la fase ϕ_s è estremamente sensibile ad effetti di nuova fisica, che possono modificare il suo valore tramite il contributo di nuove particelle alle oscillazioni di sapore. Pertanto, la misura precisa di questa fase è un potente test per indagare fenomeni fisici oltre il Modello Standard. Inoltre, analizzando il decadimento $B^0_s \to J/\psi\,\phi$ è possibile misurare diversi altri interessanti parametri fisici, come la differenza di ampiezza di decadimento $\Delta\Gamma_s$ tra gli autostati di massa del mesone B^0_s, l'ampiezza di decadimento media Γ_s, la differenza di massa Δm_s, ed il parametro $|\lambda|$ che descrive la quantità di violazione di CP diretta nel processo. Pertanto, questa misura consiste anche in un test globale della fenomenologia del sistema B^0_s.

L'estrazione dei parametri fisici viene eseguita attraverso un'analisi angolare dipendente dal tempo del decadimento $B^0_s \to J/\psi\,\phi$ e del sapore iniziale del mesone B^0_s. La misura si basa su un campione di 48 500 eventi di segnale ricostruiti in collisioni protone-protone a $\sqrt{s} = 13$ TeV, corrispondenti ad una luminosità integrata di $\mathcal{L}_{\text{int}} = 96.4\,\text{fb}^{-1}$. Rispetto alla precedente misura eseguita da CMS a $\sqrt{s} = 8$ TeV, questa analisi beneficia dell'aumento dell'energia del centro di massa che praticamente raddoppia la sezione d'urto di produzione del mesone B^0_s. Gli eventi sono raccolti con un trigger che richiede un candidato $J/\psi \to \mu^+\mu^-$ ed un terzo muone addizionale, il quale viene sfruttato per dedurre il sapore del mesone B^0_s al momento della sua produzione. Al fine di massimizzare la sensibilità dell'analisi, è stato sviluppato un nuovo algoritmo di *tagging* muonico basato su reti neurali calibrate. L'algoritmo raggiunge un elevato potere di tagging di circa il 10%, corrispondente ad un'efficienza di tagging del 50%, come conseguenza della richiesta del muone aggiuntivo imposto a livello di trigger.

La fase responsabile della violazione di CP è misurata $\phi_s = -11 \pm 50\,(\text{stat}) \pm 10\,(\text{sist})$ mrad, in accordo sia con le previsioni teoriche che con l'assenza di violazione di CP nell'interferenza tra decadimento e oscillazione di sapore. Questi risultati sono significativamente più precisi di quelli della precedente misura effettuata dalla collaborazione CMS a 8 TeV. Questa analisi rappresenta anche la prima misura della differenza di massa Δm_s tra gli autostati di massa del mesone B^0_s e del parametro $|\lambda|$ da parte dell'esperimento CMS. I risultati sono stati ulteriormente combinati con quelli ottenuti da CMS a $\sqrt{s} = 8$ TeV, ottenendo $\phi_s = -21 \pm 44\,(\text{stat}) \pm 10\,(\text{sist})$ mrad.

ACKNOWLEDGMENTS

Achieving this milestone would not have been possible without the help and support of many people.

First and foremost, I would like to express my deepest gratitude to my supervisors, Prof. Martino Margoni and Prof. Franco Simonetto, for their patient guidance during these last three years. Their invaluable support has proven essential in completing this complex work.

I would also like to express my sincere gratitude to Marcella Bona and Fernando Martinez Vidal for their thorough review of this manuscript. Their extensive feedback and sincere curiosity towards my work were both highly motivating and appreciated.

Next, I would like to express my heartfelt gratitude to the colleagues and friends of the CMS team in Padova, for all the time they spent supporting, helping, and teaching me, both in and out of work. Thanks to the CMS collaboration and CERN for providing the resources to make this research possible.

Completing this work would have not been possible without the continuous efforts of the brilliant colleagues who worked with me on this complex analysis. To the whole *BsJpsiPhi* group I owe my sincere gratitude, both on a scientific and human level. In particular, I would like to thank Giacomo for his support and for sharing his knowledge with me. Thanks to Paolo for his tireless efforts in developing the framework and taking care of other heavy-duty tasks. Thanks to Enrico for the enormous contributions that have well exceeded what was expected.

A special thanks to all the Ph.D. students and post-docs I met during my last year at CERN. Thanks to mutual support, we survived that crazy, yet unforgettable year, that was 2020. In particular, thanks to my flatmate Federico, with whom I shared the joys and sorrows of writing a thesis during a global lockdown; thanks for keeping me sane.

Infine, ringrazio la mia famiglia per il loro continuo sostegno e per tutte le opportunità che mi sono state date nel corso della vita. Non avrei mai potuto raggiungere questo traguardo senza di voi.

CONTENTS

MOST USED ACRONYMS

CL	Confidence Level
CPV	CP Violation
DNN	Deep Neural Network
EW	Electroweak
FCCC	Flavor-Changing Charged Currents
FCNC	Flavor-Changing Neutral Currents
HLT	High Level Trigger
L1	Level-1
MLE	Maximum Likelihood Estimation
MVA	Multivariate Analysis
NDF	Number of Degrees of Freedom
NLL	Negative Log-Likelihood
NP	New Physics
OS	Opposite Side
PDF	Probability Density Function
PF	Particle Flow
PV	Primary Vertex
QCD	Quantum Chormodynamics
QED	Quantum Electrodynamics
SM	Standard Model of particle physics
SS	Same Side
SSB	Spontaneous Symmetry Breaking
SV	Secondary Vertex
UML	Unbinned Maximum Likelihood
WA	World Average

INTRODUCTION

In modern physics, the Universe is described as composed of particles, point-like entities that interact with each other through fundamental forces and give rise to more complex structures. The Standard Model of particle physics (SM) is the most successful description of how these elementary particles and their interactions are related to each other. As of today, the SM has been tested extensively and withstood all experimental tests exceptionally well. Despite its overwhelming successes (to say one, the Higgs particle was predicted more than fifty years before its actual discovery), the SM is not a complete theory of fundamental interactions, as it leaves unexplained several known phenomena. To name a few, the abundance of dark matter in the Universe, the gravitational interaction, the origin of neutrino masses, and the large asymmetry between matter and antimatter. These and many other arguments motivate the need for a new, more fundamental, theory to describe the observed Universe.

In particular, the origin of the matter-antimatter asymmetry in the Universe is one of the most long-standing unsolved problems in modern physics. Antimatter particles share the same mass as their matter counterparts but have opposite physical charges (color charge, weak isospin, as well as electric charge), and when they come in contact, they annihilate one another. It is hypothesized that the Universe during its first instants was composed of almost equal amounts of matter and antimatter, leaving open the question of how the observed asymmetry was generated. As noticed by Sakharov half a century ago, the combined violation of the parity (P) and charge conjugation (C) symmetries, that is the violation of the CP symmetry, is one of the three essential conditions to produce an imbalance between matter and antimatter, the other two being the violation of the baryon number and interactions out of thermal equilibrium. CP symmetry was believed to be one of the fundamental conservation laws until its violation was discovered in 1964 in the decays of neutral kaons by Cronin and Fitch, resulting in a Nobel Prize for its discoverers. At the time only three quark types were known (up, down, and strange) and no existing theory could explain the CP violation (CPV) observation. In 1973, Kobayashi and Maskawa successfully provided an explanation within the context of the SM by generalizing the two-dimensional Cabibbo matrix into the three-dimensional Cabibbo–Kobayashi–Maskawa (CKM) matrix, allowing the inclusion of a complex non-trivial phase. This led to the hypothesis and subsequent discovery of the third family of quarks and another Nobel Prize.

The amount of CP violation in the SM, however, is not sufficient to explain the observed matter-antimatter asymmetry. For this reason, the search for CP-violating phenomena is of broad and current interest in high energy physics. The Large Hadron Collider (LHC) at CERN is the most powerful particle accelerator ever built and therefore presents unprecedented opportunities for the study of CP violation and the search for new physics. One of the large experiments it services is CMS, a general-purpose detector able to measure different types of particles and their properties by exploiting its many sub-systems.

Within this theoretical and experimental context, this dissertation presents a new study of the $B_s^0 \to J/\psi\,\phi$ decay aimed at measuring the CP-violating phase ϕ_s. This weak phase parametrizes the CP violation in the interference between direct B_s^0 meson decays to a CP eigenstate via $b \to c\bar{c}s$ transitions and decays through mixing to the same final state. The SM relates ϕ_s to the elements of the CKM matrix, which are precisely constrained by global fits to experimental data, assuming unitarity of the CKM matrix. Due to its clean theoretical prediction, the CP-violating phase ϕ_s is extremely sensitive to new-physics effects that can modify its value via the contribution of non-SM particles to the B_s^0–$\overline{B}_s^0$ mixing. The precise measurement of ϕ_s is therefore a powerful probe to search for beyond Standard Model effects, study the phenomenology of CP violation, and shed additional light on the mystery of the origin of the matter-antimatter asymmetry.

THESIS OUTLINE

The thesis is structured as follows. Part I gives a brief overview of the theoretical background and physics motivations behind the analysis. The SM is covered in Chapter 1, the phenomenology of flavor physics and CP violation is presented in Chapter 2, while the $B_s^0 \to J/\psi\,\phi$ decay and the phase ϕ_s are discussed in Chapter 3. Particular attention is given to the derivation of the SM prediction for ϕ_s and the parameterization of the $B_s^0 \to J/\psi\,\phi$ decay rate.

Part II reviews the experimental setup used to perform the analysis. Chapter 4 is dedicated to the description of the Large Hadron Collider and CMS detector, while Chapter 5 reviews the CMS trigger system, the algorithms used to reconstruct the physics objects, and the basic concepts behind Monte Carlo simulations.

Part III discusses the procedures and techniques used to collect and prepare the data sample for the actual measurement. A general overview of the analysis is given in Chapter 6, while the candidate reconstruction and selection procedure are described in Chapter 7. Subsequently, a detailed description of the development of the flavor tagging algorithms used to infer the initial flavor of the B_s^0 meson is given in Chapter 8, where a brief review on deep neural networks is also presented.

Finally, Part IV presents the CP violation measurement on the selected $B_s^0 \to J/\psi\,\phi$ data sample. Chapter 9 briefly reviews the maximum-likelihood estimation method, before describing in details the fit model and presenting the results. Particular attention is also given to the discussion of resolution and acceptance effects. The description of relevant systematic uncertainties is presented in Chapter 10, which is followed by the final discussion of the measurement results in Chapter 11, where also the combination with the previous CMS results is covered. The thesis concludes in Chapter 12 with a summary of the results and a brief discussion of future prospects of the analysis.

PUBLICATIONS AND BIBLIOGRAPHIC REFERENCES

The analysis presented in this dissertation refers to the official CMS measurement of the CP-violating phase ϕ_s performed at a center-of-mass energy of 13 TeV. The preliminary manuscript is available in Ref. [1], which has been submitted to the journal "Physics Letters B" in July 2020. As a consequence of the journal review process, the results presented in this dissertation may end up being different from the ones of the would-be published article. The results here reported have been first presented in Refs. [2] and [3]. This work is also documented in the internal analysis notes [4] and [5][1].

OWN CONTRIBUTIONS

The measurement presented in this thesis contains contributions from a team of several people, each contributing to various aspects of the analysis. Within such a group, I developed the flavor tagging framework, devising a new approach to estimate the mistag probability on a per-event basis by exploiting a deep neural network. I also gave major contributions to the development of the selection framework, the definition of the maximum-likelihood fit model, and the study of several sources of systematic uncertainties. I also curated the combination with the previous CMS measurement. Besides the data analysis, I extensively contributed to the CMS internal documentation and review process. In addition, I am also part of the ATLAS–CMS–LHCb cross-experiment combination group in charge of combining the three collaboration results and provide the world-average values for all the relevant parameters described in this work.

UNIT CONVENTION

Through this dissertation, the "natural units" convention is implied, unless stated otherwise, i. e., the speed of light in vacuum (c) and the reduced Plank constant ($\hbar$) are normalized to unity. Accordingly, both masses and momenta are measured in GeV.

FOOTNOTE CONVENTION

This work makes extensive use of footnotes. Whenever Arabic numerals might cause misunderstandings, footnotes are labeled with upper case Roman numerals.

1 These analysis notes are not publicly available.

Part I

THEORETICAL BACKGROUND

In modern physics, the Universe is described as composed of particles, point-like entities that interact with each other through fundamental forces and give rise to more complex structures. The Standard Model represents our best understanding of how these particles and three of the four forces are related to each other. Despite being consistent, precise, and very successful, the SM is not a complete theory of fundamental interactions, as it leaves some phenomena unexplained. Among these, the origin of the matter-antimatter asymmetry in the Universe is one of the most long-standing unsolved problems in modern physics. As noticed by Sakharov half a century ago, the violation of the CP symmetry is one of the essential conditions to explain the observed imbalance between matter and antimatter in the Universe. The amount of CP violation in the SM, however, is not sufficient to explain the observed matter-antimatter asymmetry. For this reason, the search for CP-violating phenomena is of broad and current interest in high energy physics. The following chapters provide a summary of the Standard Model theoretical framework, and a brief description of the phenomenology of flavor physics and CP violation relevant for this dissertation. The CP-violating phase ϕ_s, which measurement is the main result of this work, is also presented.

1 THE STANDARD MODEL OF PARTICLE PHYSICS

The Standard Model of particle physics (SM) [6–12] is the most successful description of elementary particles and their interactions. It has been tested extensively and has so far withstood all experimental tests exceptionally well. This chapter presents the content of the SM and a summary of its mathematical description.

1.1 THE STANDARD MODEL PARTICLE CONTENT

In the SM, matter and its interaction are described by twelve spin 1/2 particles, called elementary fermions, their antifermions, twelve gauge bosons, and one scalar Higgs boson. Elementary fermions are further classified into six leptons and six quarks, according to how they interact.

The quarks carry color charge, therefore they interact via strong interaction. As a consequence of the *color confinement* phenomenon, quarks are very strongly bound to one another, forming color-neutral particles called hadrons containing either a quark and an antiquark (mesons) or three quarks (baryons). Recently there have been also various experimental evidences for bounded states of four and five quarks [13–15]. Carrying electric charge and weak isospin, quarks also interact via electromagnetic and weak interactions. Two types of quarks exist: up-type (u, c, t) and down-type (d, s, b), which have an electric charge of $+\frac{2}{3}e$ and $-\frac{1}{3}e$, respectively.

Leptons do not carry color charge and can be divided into charged leptons (e, μ, τ), with electric charge of $-1e$, and neutrinos (ν_e, ν_μ, ν_τ), which are neutral and interact only through weak interaction.

Both quarks and leptons are divided into three generations with increasing mass[1].

In the SM, gauge bosons are the force-carrying particles that mediate the strong, weak, and electromagnetic fundamental interactions. All gauge bosons have spin 1 and they are classified according to the force they mediate:

- The *strong* interaction between color charged particles (i. e., quarks) is mediated by eight massless gluons (g), which also interact among themselves as they have an effective color charge.
- The *weak* interaction between elementary fermions (i. e., quarks and leptons) is mediated by the W^+, W^- and Z massive bosons, with the latter

1 This may not hold for neutrinos, as their mass has not yet been determined.

being electrically neutral and the W bosons having an electric charge of $+1e$ and $-1e$, respectively.

- The *electromagnetic* force between electrically charged particles is mediated by the massless neutral photon γ.

1.2 FIELD CONTENT AND THE STANDARD MODEL LAGRANGIAN

1.2.1 *The Standard Model field content*

The mathematical description of the SM relies on global Poincaré space-time symmetries and the local gauge invariance of its Lagrangian $\mathcal{L}_{SM}$ under the transformation of the symmetry group:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \,. \tag{1.1}$$

Gauge bosons are structured in adjoint representations of the corresponding gauge groups. The octet of gluons G^a_μ for the $SU(3)_C$ group established the interactions of quantum chromodynamics (QCD) by acting on $SU(3)_C$ triplets of quark. The electroweak (EW) gauge bosons are represented by the $SU(2)_L$ triplet W^i_μ and the singlet B_μ of $U(1)_Y$.

Elementary fermions appear as right- and left-handed spinors. The left-handed electron (e_L), muon (μ_L) and tau (τ_L) together with the associated neutrinos ($\nu_{e,L}$, $\nu_{\mu,L}$, $\nu_{\tau,L}$), form $SU(2)_L$ doublets, whereas right-handed neutrinos are sterile (non-interacting) particles, which are not included in the minimal version of the SM, leading to the right-handed $SU(2)_L$ singlets (e_R, μ_R, τ_R). Accordingly, left-handed up (u), charm (c) and top (t) quarks form $SU(2)_L$ doublets with their associated down-type down (d), strange (s) and bottom (b) quarks. The right-handed quarks form six $SU(2)_L$ singlets. All elementary fermions are charged under $U(1)_Y$.

The picture is completed by the scalar Higgs field, which contains two neutral and two electrically charged components that form a complex doublet of $SU(2)_L$.

The field content of the SM is summarized in Tab. 1.1.

1.2.2 *The quantum chromodynamics sector*

The QCD Lagrangian can be described by:

$$\mathcal{L}^{QCD} = \overline{\psi} i \gamma^\mu D_\mu \psi - \frac{1}{4} G^{a\mu\nu} G^a_{\mu\nu} \,, \tag{1.2}$$

where

- ψ is a generic massless quark spinor field;
- γ^μ are the Dirac matrices [16];

Table 1.1: The field content of the Standard Model of particle physics.

Particle	Field			Spin
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	1/2
	u_R	c_R	t_R	1/2
	d_R	s_R	b_R	1/2
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	1/2
	e_R	μ_R	τ_R	1/2
Gauge bosons		G^a_μ		1
		W^i_μ		1
		B_μ		1
Higgs-doublet		$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L$		0

- D_μ is the covariant derivative, defined as

$$D_\mu = \partial_\mu + \frac{i}{2} g_s \lambda_a G^a_\mu \,, \tag{1.3}$$

where λ_a are the Gell-Mann matrices [17] and g_s is the QCD coupling constant;

- $G^a_{\mu\nu}$ is the field tensor of the gluonic fields G^a_μ, defined as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu \,, \tag{1.4}$$

with the last term being responsible for the gluons self-interaction and f_{abc} being the structure constants of $SU(3)_C$.

The QCD coupling constant is found to increase with decreasing energy, leading to both quark confinement and QCD asymptotic freedom. Therefore the strong force between quarks is an interaction that becomes weaker at smaller distances and that becomes stronger as the quarks move apart, preventing the separation of an individual quark.

1.2.3 *The electroweak sector*

The Lagrangian of the electroweak sector of the SM is:

$$\mathcal{L}^{\mathrm{EW}} = \mathcal{L}_{\mathrm{gauge}} + \mathcal{L}_{\mathrm{fermion}}\,, \tag{1.5}$$

$$\mathcal{L}_{\mathrm{gauge}} = -\frac{1}{4}F^{i}_{\mu\nu}F^{i\,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\,, \tag{1.6}$$

$$\mathcal{L}_{\mathrm{fermion}} = \overline{\psi}^{i}_{L,R}\, i\gamma^{\mu}D_{\mu}\psi^{i}_{L,R}\,, \tag{1.7}$$

where

- $F^{i}_{\mu\nu}$ and $B_{\mu\nu}$ are the gauge field tensors of $SU(2)_L$ and $U(1)_Y$, respectively, defined as

  $$F^{i}_{\mu\nu} = \partial_\mu W^{i}_{\nu} - \partial_\nu W^{i}_{\mu} + g\epsilon^{ijk}W^{j}_{\mu}W^{k}_{\nu}\,, \tag{1.8}$$

  $$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu\,, \tag{1.9}$$

  with W^{i}_{μ} being the $SU(2)_L$-gauge field triplet, B_μ being the $U(1)_Y$-gauge field, and ϵ^{ijk} being the three-dimensional Levi-Civita symbol;

- D_μ is the covariant derivative for the $SU(2)_L \otimes U(1)_Y$ SM gauge theory, defined as

  $$D_\mu = \partial_\mu - igT_iW^{i}_{\mu} - ig'\frac{Y}{2}B_\mu\,,$$

  with g and g' being the coupling constants of the $SU(2)_L$ and $U(1)_Y$ interaction, respectively, and T^i and Y being the gauge group generators;

- ψ^{i}_{L} are the left-handed fermion fields with $i = e, \mu, \tau, \nu_e, \nu_\mu, \nu_\tau, u, c, t, d, s, b$.

- ψ^{i}_{R} are the right-handed fermion fields with $i = e, \mu, \tau, u, c, t, d, s, b$.[II]

Mass terms for all fields in the form of $m_f\overline{\psi}\psi$, $m_B^2 B_\mu B^\mu$ or $m_W^2 W_\mu W^\mu$ are strictly forbidden[3], as these would directly spoil the gauge symmetry. In the SM the masses of charged elementary fermions and gauge bosons are generated through electroweak spontaneous symmetry breaking (SSB), in the so-called "Higgs mechanism".

II There are no right-handed neutrinos, nor left-handed antineutrinos.

3 m_f, m_B, and m_W would be the masses of the particles represented by the fields ψ, B_μ, and W_μ, respectively.

1.3 THE HIGGS MECHANISM

In the SM the masses of gauge bosons are generated by spontaneously breaking the corresponding gauge symmetries. The term spontaneous means that the symmetry is not broken explicitly by the interactions, but rather by the asymmetry of the state of lowest energy, referred to as the vacuum state. According to the Goldstone theorem, the spontaneous symmetry breaking (SSB) creates massless bosons, so-called "Goldstone bosons" [18]. If the broken symmetry also corresponds to a gauge symmetry, then the associated Goldstone boson and the massless gauge boson combine to form a massive gauge boson. This is the famous Higgs mechanism, elaborated in 1964 independently by Peter Higgs [19, 20], Robert Brout and François Englert [21], and by Gerald Guralnik, Carl Hagen, and Tom Kibble [22][4]. The simpler way to implement the Higgs mechanism is to add to the theory a weakly coupled spin-0 particle with a potential that is minimized at a non-zero field value. Such a field spontaneously breaks the electroweak $SU(2)_L \otimes U(1)_Y$ symmetry, giving mass to the gauge bosons, while the masses of the charged elementary fermions are generated by "Yukawa couplings" to the newly introduced scalar field.

The simplest choice for the new scalar field is a single complex scalar doublet, the "Higgs field":

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.10}$$

The Higgs Lagrangian can be then written as:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) + \mathcal{L}_{\text{Yukawa}}\,, \tag{1.11}$$

where the covariant derivative is

$$D_\mu = \partial_\mu - \frac{i}{2} g \tau_i W^i_\mu - \frac{i}{2} g' B_\mu\,, \tag{1.12}$$

and $\mathcal{L}_{\text{Yukawa}}$ is the Yukawa Lagrangian, which describes the generation of charged elementary fermion masses[5]. The only renormalizable model where a complex scalar field acquires a nonzero value is the *Mexican-hat* model, with potential:

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2\,, \tag{1.13}$$

where unitarity requires the constants λ and μ^2 to be real and stability demands λ to be positive. The parameter μ^2 is chosen negative, in order to realize the spontaneous symmetry breaking

$$SU(2)_L \otimes U(1)_Y \xrightarrow{\text{SSB}} U(1)_{\text{QED}}\,, \tag{1.14}$$

4 As such, the Higgs mechanism is also called the BEH (Brout–Englert–Higgs) mechanism, or EBHGHK (Englert–Brout–Higgs-Guralnik–Hagen–Kibble) mechanism.

5 The Yukawa Lagrangian is described in Section 1.3.2.

where $U(1)_{QED}$ is the gauge symmetry group of the electromagnetic interactions. This inevitably leads to a degenerate non-zero vacuum state of the Higgs field:

$$\phi^\dagger\phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}, \tag{1.15}$$

with $v \equiv \sqrt{-\frac{\mu^2}{\lambda}}$. In order to have an electrically neutral vacuum, only ϕ^0 can contribute to the vacuum expectation value (VEV) of the Higgs field, leading to:

$$\langle\phi\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.16}$$

Expanding around the minimum of its potential and applying a convenient gauge transformation, called "unitary gauge", the scalar doublet can be written in the form

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ 0 \\ v + H(x) \end{pmatrix}, \tag{1.17}$$

where $H(x)$ is a real scalar field, called the Standard Model Higgs boson field, that represent *radial* excitations of the Higgs VEV.

1.3.1 *Boson masses*

The bosonic mass Lagrangian $\mathcal{L}_m$ can be obtained by expressing the Higgs Lagrangian in Eq. 1.11 in the unitary gauge, omitting the kinetic, interaction and Yukawa terms[6]:

$$\mathcal{L}_m = -\lambda v^2 H^2 + \frac{v^2}{8}\left\{g^2\left[\left(W^1_\mu\right)^2 + \left(W^2_\mu\right)^2\right] + \left(gW^3_\mu - g'B_\mu\right)^2\right\}. \tag{1.18}$$

The bosonic fields can be rotated, without loss of generality, into the physical states:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right) \quad \text{(charged currents)}, \tag{1.19}$$

$$Z_\mu = \cos\theta_W W^3_\mu - \sin\theta_W B_\mu \quad \text{(neutral current)}, \tag{1.20}$$

$$A_\mu = \sin\theta_W W^3_\mu + \cos\theta_W B_\mu \quad \text{(photon)}, \tag{1.21}$$

6 Interaction terms are those with more than two fields.

where θ_W is the so called "Weinberg angle", which is related to the coupling constants g and g' as:

$$\begin{aligned} \cos\theta_W &= \frac{g}{\sqrt{g^2+g'^2}}\,, \\ \sin\theta_W &= \frac{g'}{\sqrt{g^2+g'^2}}\,. \end{aligned} \tag{1.22}$$

After this rotation, the mass Lagrangian is given by:

$$\mathcal{L}_m = -\lambda v^2 H^2 + m_w^2 W_\mu^+ W^{-\,\mu} + \frac{1}{2} m_Z^2 Z_\mu Z^\mu\,, \tag{1.23}$$

where

- the first term gives the mass of the Higgs boson as:

$$m_H \equiv \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}\,, \tag{1.24}$$

- the second and third term are the mass terms for the W and Z gauge bosons, from which the masses are given by:

$$m_W \equiv \frac{gv}{2}\,, \tag{1.25}$$

$$m_Z \equiv \frac{v}{2}\sqrt{g^2+g'^2} = \frac{gv}{2\cos\theta_W}\,. \tag{1.26}$$

The cosine of the Weinberg angle, defined in Eq. 1.22, can therefore be expressed in terms of m_W and m_Z as:

$$\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}} = \frac{m_W}{m_Z}\,. \tag{1.27}$$

The photon field A_μ remains massless after the rotation 1.21.

1.3.2 *Fermion masses*

In the SM the fermion masses are generated via Yukawa couplings in the Yukawa Lagrangian

$$\mathcal{L}_{\text{Yukawa}} = -\sum_{i,j=1}^{3}\left[Y_{ij}^{\ell}\overline{L_{Li}}\phi\ell_{Rj} + Y_{ij}^{u}\overline{Q_{Li}}\widetilde{\phi}q_{Rj}^{u} + Y_{ij}^{d}\overline{Q_{Li}}\phi q_{Rj}^{d} + \text{h.c.}\right], \tag{1.28}$$

where

- i and j are flavor generation indices;
- ℓ_{Ri}, q_{Ri}^d, and q_{Ri}^u with $i = 1, 2, 3$ are the right-handed lepton, down-type quark, and up-type quark singlets, respectively;

- $L_{Li} = (\nu_{Li}, \ell_{Li})^T$ and $Q_{Li} = (q^u_{Li}, q^d_{Li})^T$ with $i = 1, 2, 3$ are the left-handed lepton and quark doublets, respectively;
- ϕ is the Higgs doublet with $\widetilde{\phi} \equiv i\sigma_2\phi^*$ being its complex conjugate[7];
- Y^ℓ, Y^u, and Y^d are 3×3 complex matrices, called "Yukawa coupling matrices", that describe the coupling between the Higgs doublet ϕ and the fermions.

LEPTON SECTOR In the unitary gauge[8] the lepton sector of the Yukawa Lagrangian is given by:[9]

$$\mathcal{L}^{\text{lepton}}_{\text{Yukawa}} = -\left(\frac{v+H}{\sqrt{2}}\right) \sum_{i,j=e,\mu,\tau} Y^\ell_{ij} \overline{\ell_{Li}} \ell_{Rj} + \text{h.c.}\,, \tag{1.29}$$

where the mass terms for the charged fermions are the ones proportional to the VEV (v) of the Higgs doublet. However, since the matrix Y^ℓ is in general nondiagonal, the fermion fields $\ell_i = \ell_{Li} + \ell_{Ri}$ ($i = e, \mu, \tau$) do not have definite masses. The Lagrangian in Eq. 1.29 can also be written in matrix form as

$$\mathcal{L}^{\text{lepton}}_{\text{Yukawa}} = -\left(\frac{v+H}{\sqrt{2}}\right) \overline{\ell_L} Y^\ell \ell_R + \text{h.c.}\,, \tag{1.30}$$

where ℓ_L and ℓ_R are the arrays of charged lepton fields defined as

$$\ell_L \equiv \begin{pmatrix} e_L \\ \mu_L \\ \tau_L \end{pmatrix}, \quad \ell_R \equiv \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix}. \tag{1.31}$$

In order to construct fields with a definite mass, the Yukawa matrix Y^ℓ needs to be diagonalized through a biunitary transformation as

$$Y'^\ell = V^{\ell\dagger}_L Y^\ell V^\ell_R\,, \quad \text{with} \quad Y'^\ell_{ij} = y^\ell_i \delta_{ij} \quad (i, j = e, \mu, \tau)\,, \tag{1.32}$$

where Y'^ℓ is a real diagonal matrix with y^ℓ_i ($i = e, \mu, \tau$) as diagonal elements and $V^\ell_{L,R}$ are two appropriate 3×3 unitary matrices.

---

7 σ_2 is the second Pauli matrix.
8 See Eq. 1.17.
9 From Eq. 1.29 onwards the generation indices are explicitly labeled after the corresponding three generations of elementary fermions.

After the diagonalization in Eq. 1.32, the Lagrangian can be written as:

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}}^{\text{lepton}} &= -\left(\frac{v+H}{\sqrt{2}}\right)\overline{\ell'_L}Y'^{\ell}\ell'_R + \text{h.c.}\,,\\ &= -\sum_{i=e,\mu,\tau}\frac{y_i^{\ell}v}{\sqrt{2}}\overline{\ell'_i}\ell'_i - \sum_{i=e,\mu,\tau}\frac{y_i^{\ell}}{\sqrt{2}}\overline{\ell'_i}\ell'_i H\,,\end{aligned} \tag{1.33}$$

where

$$\ell'_L = V_L^{\ell\dagger}\ell_L \equiv \begin{pmatrix} e'_L \\ \mu'_L \\ \tau'_L \end{pmatrix}, \quad \ell'_R = V_R^{\ell\dagger}\ell_R \equiv \begin{pmatrix} e'_R \\ \mu'_R \\ \tau'_R \end{pmatrix} \tag{1.34}$$

are the left- and right-handed components of the charged lepton fields $\ell'_i \equiv \ell'_{Li} + \ell'_{Ri}$ $(i = e, \mu, \tau)$ with definite masses

$$m_i = \frac{y_i^{\ell}v}{\sqrt{2}} \quad (i = e,\ \mu,\ \tau)\,. \tag{1.35}$$

Primes are used to distinguish the fermionic fields with definite mass, hereafter "mass eigenstates", from the ones with well-defined transformations under the electroweak gauge group, hereafter "flavor eigenstates". The theory does not put any constraint on the couplings y^{ℓ}, which are free parameters and can only be determined experimentally.

In principle, the transformations in Eq. 1.34 should affect the leptonic charged weak current, as they effectively rotate flavor eigenstates into mass eigenstates. However, this is irrelevant for charged leptons, as the massless neutrino fields can be freely redefined to absorb the previously mentioned flavor rotation[10], so that the flavor and mass eigenstates of charged leptons coincide:

$$j^{\mu}_{W,\ell} = \overline{\nu_L}\gamma^{\mu}\ell_L = \overline{\nu_L}\gamma^{\mu}V_L^{\ell}\ell'_L = \overline{\nu'}_L\gamma^{\mu}\ell'_L\,. \tag{1.36}$$

where j^{μ}_W is the leptonic charged weak current written in matrix form[11] and the massless neutrino fields $\nu_L = (\nu_{eL},\ \nu_{\mu L},\ \nu_{\tau L})^T$ have been transformed as

$$\nu_L \to \nu'_L = V_L^{\ell\dagger}\nu_L \tag{1.37}$$

to reabsorb the flavor rotation of the charged lepton fields.

It can also be shown that neutral weak currents do not depend at all on the unitary matrices relating flavor to mass eigenstate, therefore, the expression of the leptonic charged and neutral weak currents in terms of flavor eigenstates is the same as that in terms of mass eigenstates.

10 Any linear combination of massless fields is also a massless field. This implies that in the SM the neutrino flavor eigentates coincide with the mass ones, which is observed to be not true in Nature, as neutrinos mixing and oscillation phenomena are well established.

11 The leptonic charged weak current can be derive from the EW Lagrangian defined in Eq. 1.5.

QUARK SECTOR For quarks, the discussion follows the same reasoning made for leptons. The corresponding Yukawa coupling matrices are diagonalized via biunitary transformations as

$$\begin{aligned} Y'^{d} &= V^{d}_{L} Y^{d} V^{d\dagger}_{R} \qquad \text{with} \quad Y'^{d}_{ij} = y^{d}_{i}\delta_{ij} \quad (i,j = d,s,b)\,, \\ Y'^{u} &= V^{u}_{L} Y^{u} V^{u\dagger}_{R} \qquad \text{with} \quad Y'^{u}_{ij} = y^{u}_{i}\delta_{ij} \quad (i,j = u,c,t)\,, \end{aligned} \tag{1.38}$$

where $V^{u}_{L,R}$ and $V^{d}_{L,R}$ are four appropriate 3×3 complex unitary matrices. As for the leptons, the following arrays are also defined to allow the Lagrangian to be written in matrix form:

$$q^{d}_{L} \equiv \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}, \quad q^{d}_{R} \equiv \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix}, \quad q^{u}_{L} \equiv \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix}, \quad q^{u}_{R} \equiv \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix}. \tag{1.39}$$

After the change of basis in Eq. 1.38, the quark Yukawa terms in the unitary gauge are:

$$\begin{aligned} \mathcal{L}^{\text{quark}}_{\text{Yukawa}} &= -\left(\frac{v+H}{\sqrt{2}}\right)\left[\overline{q'^{d}_{L}} Y'^{d} q'^{d}_{R} + \overline{q'^{u}_{L}} Y^{u} q'^{u}_{R}\right] + \text{h.c.} \\ &= -\left(\frac{v+H}{\sqrt{2}}\right)\left[\sum_{i=d,s,b} y^{d}_{i} \overline{q'^{d}_{Li}} q'^{d}_{Ri} + \sum_{i=u,c,t} y^{u}_{i} \overline{q'^{u}_{Li}} q'^{u}_{Ri}\right] + \text{h.c.}\,, \end{aligned} \tag{1.40}$$

where

$$\begin{aligned} q'^{d}_{L} &= V^{d\dagger}_{L} q^{d}_{L} \equiv \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix}, \quad q'^{d}_{R} = V^{d\dagger}_{R} q^{d}_{R} \equiv \begin{pmatrix} d'_R \\ s'_R \\ b'_R \end{pmatrix}, \\ q'^{u}_{L} &= V^{u\dagger}_{L} q^{u}_{L} \equiv \begin{pmatrix} u'_L \\ c'_L \\ t'_L \end{pmatrix}, \quad q'^{u}_{R} = V^{u\dagger}_{R} q^{u}_{R} \equiv \begin{pmatrix} u'_R \\ c'_R \\ t'_R \end{pmatrix}, \end{aligned} \tag{1.41}$$

are the left- and right-handed components of the quark fields with definite masses

$$m_i = \frac{y^{d}_{i} v}{\sqrt{2}} \quad (i = d,s,b)\,, \tag{1.42}$$

$$m_j = \frac{y^{u}_{j} v}{\sqrt{2}} \quad (j = u,c,t)\,. \tag{1.43}$$

As in the case of leptons, the y^d and y^u couplings are not predicted by the theory and must be obtained experimentally.

For quarks, their mass eigenstates do not coincide with the electroweak ones, since the transformations in Eq. 1.41 cannot be reabsorbed as in the case of leptons. The quark charged weak current can be derived from the electroweak Lagrangian in Eq. 1.5 and written in matrix form as:

$$j^{\mu}_{W,q} = \underbrace{\overline{q^{u}_{L}}\gamma^{\mu}q^{d}_{L}}_{\text{flavor eigenstates}} = \underbrace{\overline{q'^{u}_{L}}V^{u\dagger}_{L}\gamma^{\mu}V^{d}_{L}q'^{d}_{L}}_{\text{mass eigenstates}} = \overline{q'^{u}_{L}}\gamma^{\mu}\underbrace{V^{u\dagger}_{L}V^{d}_{L}}_{\equiv V_{CKM}}q'^{d}_{L}\,. \tag{1.44}$$

Equation 1.44 shows that the quark charged weak current does not depend separately on the matrices V^{u}_{L} and V^{d}_{L}, but only on their product

$$V_{CKM} = V^{u}_{L}V^{d\dagger}_{L}\,. \tag{1.45}$$

The matrix V_{CKM} is the Cabibbo–Kobayashi–Maskawa (CKM) matrix [23, 24] and contains the information of the physical effects of quark mixing.

The CKM matrix is particularly relevant for this dissertation as it is one of the only two sources of CP violation in the SM[12] and it is described in detail in Chapter 2.

12 The other being the PMNS matrix [25, 26], describing the neutrino mixing. Evidence of CP violation in neutrino oscillations has been recently reported by the T2K Collaboration [27].

2 FLAVOR PHYSICS AND CP VIOLATION

The concept that physical systems are invariant under some transformations ("symmetries") plays an important role in the fundamental description of nature. Continuous symmetries[1] lead to conservation laws as for Noether's theorem [28], while discrete ones are of high interest in modern particle physics.

The Standard Model of particle physics has three natural discrete *near*-symmetries: parity (P), charge conjugation (C), and time reversal (T). Composed symmetries, such as CP and CPT, can also be constructed. As one of the three Sakharov's conditions for baryogenesis [29], CP violation (CPV) is an essential ingredient for explaining the observed baryon asymmetry in the Universe. The violation of the CP symmetry has been observed in kaons [30], B mesons [31, 32], and D mesons [33][2]. Evidence of CPV in neutrino oscillations has been recently reported by the T2K Collaboration [27]. In the SM the only discrete symmetry observed to be exact is CPT. In fact, the CPT theorem states that any Lorentz invariant local quantum field theory with an Hermitian Hamiltonian must have CPT symmetry [36–38]. An immediate consequence of the CPT theorem is that if CP is violated then also T must be violated, in order to preserve CPT.

This chapter presents the discrete symmetries of the SM and a summary of the CPV phenomenology in B mesons.

2.1 DISCRETE SYMMETRIES IN THE STANDARD MODEL

P-SYMMETRY The parity transformation inverts the spatial coordinates and momenta, as:

$$\begin{aligned} P : (x, y, z) &\rightarrow (-x, -y, -z)\,, \\ P : (p_x, p_y, p_z) &\rightarrow (-p_x, -p_y, -p_z)\,, \end{aligned} \tag{2.1}$$

where the parity operator P is unitary. This symmetry is conserved by the electromagnetic and strong interactions. However, it is violated maximally in weak interactions [39, 40].

C-SYMMETRY The charge conjugation transformation switches all particles with their corresponding antiparticles, inverting all additive quantum numbers as:

$$C : |\psi\rangle \rightarrow |\overline{\psi}\rangle\,, \tag{2.2}$$

1 Examples of continuous symmetries: time translation, spatial translation, spatial rotation, Poincaré transformation.

2 Only the first observation for each species is cited. For a full list of CPV observations see Refs. [34, 35].

where the charge conjugation operator C is unitary. This is maximally violated in weak interaction: left-handed fermions have different couplings to W and Z from right-handed ones; right-handed neutrinos do not interact with any other particle and are not included in the minimal version of the SM. C is conserved by all the other fundamental interactions.

T-SYMMETRY The time reversal transformation reflects the time axis, as:

$$T : t \to -t\,. \tag{2.3}$$

This is a near-symmetry of the SM Lagrangian[3]. Contrary to P and C, the time reversal operator T is antiunitary. The T-symmetry is (slightly) violated together with CP, as for the CPT theorem. T violation has been observed also directly in decays of B^0 meson [41].

CP-SYMMETRY The CP symmetry is the combination of the parity and charge conjugation symmetries. In other words, CP symmetry states that the laws of physics should be the same if a particle is switched with its antiparticle while its spatial coordinates are inverted. Within the Standard Model, CP is violated at tree level in flavor-changing charged currents of the weak interactions (FCCC). This type of interaction is described in the following section.

2.2 FLAVOR-CHANGING CURRENTS OF THE WEAK INTERACTION

In the Standard Model, FCCC are mediated by the W^+ and W^- bosons. Their coupling is described by the Lagrangian

$$\mathcal{L}_{W^\pm} = -\frac{g}{2\sqrt{2}}\left[\sum_i W^+_\mu \overline{\phi}^u_i \gamma^\mu (1-\gamma^5)\phi^d_i + \sum_i W^-_\mu \overline{\phi}^d_i \gamma^\mu (1-\gamma^5)\phi^\mu_i\right] \tag{2.4}$$

where

- $\phi^{u,d}_i$ are either the quark or the lepton fields:

$$\begin{aligned} \phi^u &= \begin{pmatrix} u \\ c \\ t \end{pmatrix} \qquad \phi^d = \begin{pmatrix} d \\ s \\ b \end{pmatrix} \qquad \text{(quarks)}\,, \\ \phi^u &= \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} \qquad \phi^d = \begin{pmatrix} e^- \\ \mu^- \\ \tau^- \end{pmatrix} \qquad \text{(leptons)}\,; \end{aligned} \tag{2.5}$$

3 In macroscopic systems a preferred direction of time exists. An evident example is the expansion of the Universe.

Figure 2.1: Graphical representation of flavor-changing charged currents mediated by a W^+ (left) and W^- (right) bosons. The strength of the interaction is proportional to $V_{qq'}$ and $V^*_{qq'}$.

- $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^4$ is the so called "fifth" Dirac matrix;
- $\frac{1}{2}(1-\gamma^5) \equiv P_L$ is the left-handed projection operator.

As discussed in Section 1.3.2, the quarks flavor and mass eigenstates do not coincide. For the quark sector, Eq. 2.4 can be rewritten in terms of the mass eigenstates using the CKM matrix V_{CKM} defined in Eq. 1.45[IV]:

$$\mathcal{L}^{\text{quark}}_{W^\pm} = -\frac{g}{\sqrt{2}}\left[\sum_{i,j} W^+_\mu \overline{u}_i \gamma^\mu P_L V_{ij} d_j + \sum_{i,j} W^-_\mu \overline{d}_j \gamma^\mu P_L V^\dagger_{ij} u_i\right]. \tag{2.6}$$

It is easy to see in Eq. 2.6 that up- and down-type quarks, from different generations i and j, couple with each other with strength proportional to the elements V_{ij} of the CKM matrix. A graphical representation of FCCC interaction vertices is shown in Fig. 2.1.

2.3 THE CKM MATRIX

As discussed in the previous section, the CKM matrix is of central importance to describe FCCC interactions[5]. The CKM elements are defined as:

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}, \tag{2.7}$$

where on the left are the down-type quarks flavor eigenstates[6] and on the right the CMK matrix along with the mass eigenstates. As described in Eq. 2.6, the probability of a transition from one quark i to another quark j is proportional to $|V_{ij}|^2$.

IV Hereafter, the subscript *CKM* will be dropped in the most crowded equations.

5 But also to describe some flavor changing *neutral* currents phenomenology, e. g. flavor oscillations (see Section 2.4).

6 The choice of use down-type quarks is a convention. Other conventions are equally valid.

The current best determination of the magnitudes of the CKM elements is [34]:

$$|V_{CKM}| = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{pmatrix}. \tag{2.8}$$

It is immediate to see that the CKM matrix is nearly diagonal, with off-diagonal terms being very small and diagonal terms very close to one. This is reflected by the suppression of transitions between different quark families.

2.3.1 *CKM parametrizations*

The CKM matrix is a unitary 3×3 complex matrix and as such it has $n^2 = 9$ degrees of freedom, which can be reduced down to four, by reabsorbing five quark phases. Two of the most common CKM parametrizations are the so-called "Standard" representation and the one developed by Wolfenstein [42].

The Standard representation uses three Euler angles $\theta_{ij} = (\theta_{12}, \theta_{23}, \theta_{13})$ and one CP-violating phase δ:

$$\begin{aligned} V_{CKM} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \end{aligned} \tag{2.9}$$

where cosines and sines of the angles θ_{ij} are denoted c_{ij} and s_{ij}, respectively. Due to the CPV phase δ appearing in several elements of V_{CKM}, CP violation is expected in all three generations of quarks.

On the other hand, the Wolfenstein representation is based on the observed hierarchy among the parameters and is obtained by expanding the CKM terms in powers of $\lambda = s_{12} \simeq 0.23$. At the order λ^3 the CKM matrix is:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{2.10}$$

where A, ρ and η are real parameters of order unity ($A \simeq 0.79$, $\rho \simeq 0.14$, $\eta \simeq 0.36$ [34]). In this parametrization CP is violated if $\eta \neq 0$. An alternative and widely used convention for the parameters ρ and η is:

$$\overline{\rho} = \rho\left(1 - \frac{\lambda^2}{2} + \mathcal{O}(\lambda^4)\right), \quad \overline{\eta} = \eta\left(1 - \frac{\lambda^2}{2} + \mathcal{O}(\lambda^4)\right). \tag{2.11}$$

2.3.2 *The unitary triangles*

The unitary condition of the CKM matrix[7] leads a set of nine equations. The first three are constrains on the diagonal terms:

$$\sum_k |V_{ik}|^2 = \sum_i |V_{ik}|^2 = 1\,, \tag{2.12}$$

which imply that the sum of the couplings of any of the up-type quarks to all the down-type quarks is the same for all generations (*weak universality*). The remaining six constrains can be written as:

$$\sum_k V_{ik} V^*_{jk} = 0 \quad \text{with fixed } i, j, \tag{2.13}$$

$$\sum_i V_{ij} V^*_{ik} = 0 \quad \text{with fixed } j, k. \tag{2.14}$$

These relations can be represented as six *triangles* in the complex plane, each of which called a "unitary triangle", with area equal to half of the Jarlskog invariant [43][8]:

$$J \equiv \mathrm{Im}(V_{us} V_{cb} V^*_{ub} V^*_{cs}) = c_{12} c^2_{13} c_{23} s_{12} s_{13} s_{23} \sin\delta\,. \tag{2.15}$$

It is immediate to notice that $J \neq 0$ if CP is violated with $\delta \neq (0, \pi)$. The current best determination of the Jarlskog invariant is $J = \left(3.00^{+0.15}_{-0.09}\right) \times 10^{-5}$ [34].

Of particular interest for the measurement presented in this dissertation is the so-called "B^0_s unitary triangle":

$$V_{us} V^*_{ub} + V_{cs} V^*_{cb} + V_{ts} V^*_{tb} = 0\,, \tag{2.16}$$

with angles:

$$\alpha_s = \arg\left(-\frac{V_{ts} V^*_{tb}}{V_{us} V^*_{ub}}\right), \quad \beta_s = \arg\left(-\frac{V_{ts} V^*_{tb}}{V_{cs} V^*_{cb}}\right), \quad \gamma_s = \arg\left(-\frac{V_{us} V^*_{ub}}{V_{cs} V^*_{cb}}\right). \tag{2.17}$$

Note that the angle β_s is defined with V_{ts} in the numerator, in contrast with other unitary triangles where the angle β is defined with V_{tq} in the denominator. This generate a relative minus sign between β_s and β. This B^0_s unitary triangle is shown in Fig. 2.2.

---

7 $V^* V = V V^* = \mathbb{1}_3$.

8 All unitary triangles have the same area.

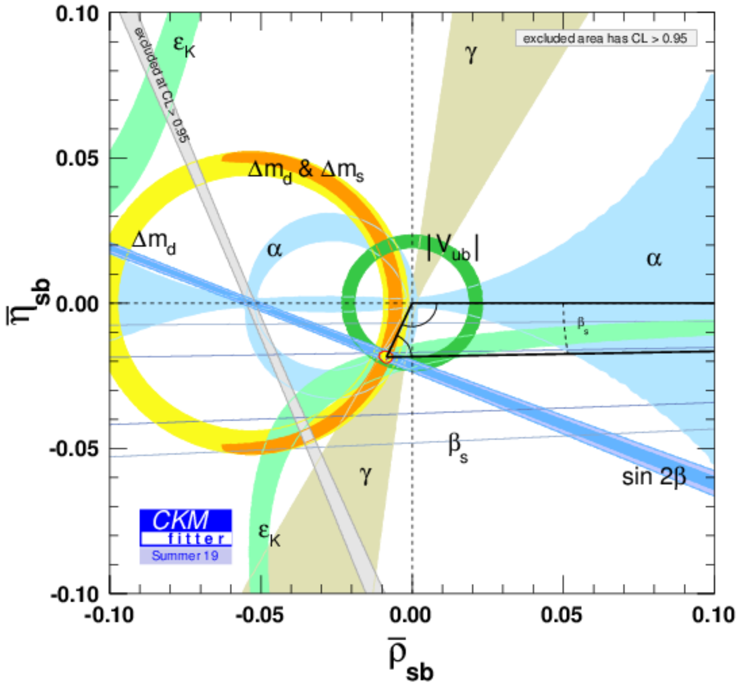


Figure 2.2: Constrains and global fit results for the B_s^0 unitary triangle. The red hashed region of the global combination for the vertex position corresponds to 68% confidence level (CL). Figure from Ref. [44].

2.4 NEUTRAL B MESONS MIXING

Neutral K, D, and $B^0_{(s)}$ mesons are produced in particle colliders in states with well-defined quark content (flavor eigenstates). These are not however eigenstate of the effective, time evolution (weak) Hamiltonian, so that they may evolve in time to their antiparticle, formed by opposite flavor quarks. This phenomenon. known as *flavor oscillation* or *flavor mixing*, is mediated by flavor-changing neutral currents (FCNC). An example of leading mixing diagrams for the B^0_s meson is shown in Fig. 2.3.

This section discusses the oscillations of B^0_s mesons, as it is the case relevant for this dissertation, but the same formalism can be used to characterize the mixing phenomenology of the other neutral meson species.

2.4.1 *Mass and flavor*

A general state of the B^0_s system can be described as a superposition of the two flavor eigenstates

$$|\Psi(t)\rangle = \left|B^0_s(t)\right\rangle + \left|\overline{B}^0_s(t)\right\rangle \tag{2.18}$$

with time evolution governed by the Schrödinger equation:

$$i\frac{d}{dt}\begin{pmatrix}\left|B^0_s(t)\right\rangle \\ \left|\overline{B}^0_s(t)\right\rangle\end{pmatrix} = H\begin{pmatrix}\left|B^0_s(t)\right\rangle \\ \left|\overline{B}^0_s(t)\right\rangle\end{pmatrix}. \tag{2.19}$$

The effective Hamiltonian H can be split using 2×2 Hermitian matrices, as

$$H \equiv M - \frac{i}{2}\Gamma,$$

where:

- M is Hermitian and it is associated with off-shell dispersive intermediate states (e. g. box diagrams);
- $\frac{i}{2}\Gamma$ is antihermitian and associated with on-shell absorptive intermediate states (e. g. $B^0_s \to \pi^+\pi^- \to \overline{B}^0_s$).

Diagonal and off-diagonal terms of M and Γ are associated with the flavor-conserving transitions $B^0_s \leftrightarrows B^0_s$ ($\overline{B}^0_s \leftrightarrows \overline{B}^0_s$) and with the flavor-changing transitions $B^0_s \leftrightarrows \overline{B}^0_s$ ($\overline{B}^0_s \leftrightarrows B^0_s$), respectively. The CPT theorem also requires the diagonal terms of H to be equal, which leads to:

$$\text{CPT} \quad \leftrightarrow \quad \begin{matrix} M_{11} = M_{22} \\ \Gamma_{11} = \Gamma_{22} \end{matrix}, \tag{2.20}$$

where M_{ij} and Γ_{ij} are elements of the M and Γ matrices, respectively.

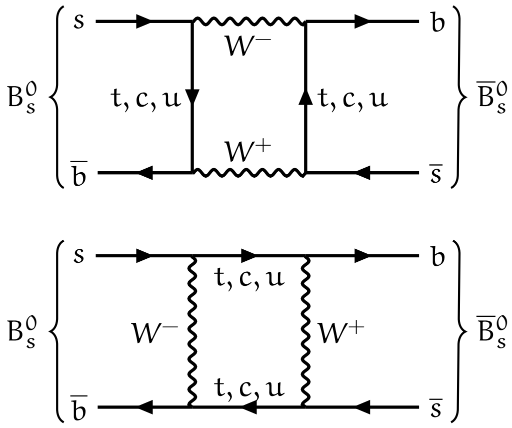


Figure 2.3: Leading order diagrams for B_s^0–$\overline{B}_s^0$ mixing.

2.4.2 *Mass eigenstates*

The light and heavy mass eigenstates $B_s^{L,H}$ are given by the eigentates of the effective Hamiltonian H and are described by a superposition of flavor states as:

$$\begin{aligned} \left|B_s^L\right\rangle &= p\left|B_s^0\right\rangle + q\left|\overline{B}_s^0\right\rangle, \\ \left|B_s^H\right\rangle &= p\left|B_s^0\right\rangle - q\left|\overline{B}_s^0\right\rangle, \end{aligned} \tag{2.21}$$

where p and q are complex coefficients obeying the normalization condition $|p|^2 + |q|^2 = 1$. The mass and decay width of the light (heavy) mass eigenstate B_s^L (B_s^H) are indicated with M_L (M_H) and Γ_L (Γ_H), respectively. According to the Schrödinger equation, the time evolution of the mass eigenstates is described by:

$$i\frac{d}{dt}\begin{pmatrix}\left|B_s^L(t)\right\rangle \\ \left|B_s^H(t)\right\rangle\end{pmatrix} = \begin{pmatrix} M_L - \frac{i}{2}\Gamma_L & 0 \\ 0 & M_H - \frac{i}{2}\Gamma_H \end{pmatrix}\begin{pmatrix}\left|B_s^L(t)\right\rangle \\ \left|B_s^H(t)\right\rangle\end{pmatrix}, \tag{2.22}$$

where the 2×2 matrix has been obtained by diagonalizing H in the $\left(\left|B_s^L\right\rangle, \left|B_s^H\right\rangle\right)$ basis:

$$Q^{-1}HQ = \begin{pmatrix} M_L - \frac{i}{2}\Gamma_L & 0 \\ 0 & M_H - \frac{i}{2}\Gamma_H \end{pmatrix}, \quad \text{with } Q = \begin{pmatrix} q & p \\ p & -q \end{pmatrix}. \tag{2.23}$$

From Eq. 2.23, the ratio of the complex coefficients p and q is given by:

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}}. \tag{2.24}$$

The time evolution of the mass eigenstates is finally given by:

$$\left|B_s^{H,L}(t)\right\rangle = e^{-iM_{H,L}t - \frac{1}{2}\Gamma_{H,L}t}\left|B_s^{H,L}(t)\right\rangle. \tag{2.25}$$

MASS EIGENSTATES PROPERTIES The average mass (width) m_s (Γ_s) and the mass (width) differences Δm_s ($\Delta\Gamma_s$) of the two mass eigenstate can be defined as:

$$m_s \equiv \frac{M_H + M_L}{2} = M_{11} = M_{22}\,, \tag{2.26}$$

$$\Gamma_s \equiv \frac{\Gamma_H + \Gamma_L}{2} = \Gamma_{11} = \Gamma_{22}\,, \tag{2.27}$$

$$\Delta m_s \equiv M_H - M_L\,, \tag{2.28}$$

$$\Delta\Gamma_s \equiv \Gamma_L - \Gamma_H\,. \tag{2.29}$$

For B_s^0 mesons the approximation $\Gamma_{12} \ll M_{12}$ holds, which leads to

$$\Delta m_s \simeq 2|M_{12}|\,, \tag{2.30}$$

$$\Delta\Gamma_s \simeq 2|\Gamma_{12}|\cos\phi\,, \tag{2.31}$$

$$\frac{q}{p} \simeq \sqrt{\frac{M_{12}^*}{M_{12}}}\,, \tag{2.32}$$

where $\phi = \arg(-M_{12}/\Gamma_{12})$ is the relative phase between the off diagonal terms M_{12} and Γ_{12}. An important consequence of this approximation is that

$$\left|\frac{q}{p}\right| \simeq 1\,, \tag{2.33}$$

which is supported by both experimental results and theoretical predictions [34][9]. The ratio q/p, can then be expressed in terms of a complex phase as:

$$\frac{q}{p} \simeq e^{-i\phi_M}\,, \tag{2.34}$$

where $\phi_M = \arg(M_{12})$, which can be further expressed in terms of CKM elements as:

$$\frac{q}{p} \simeq \frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}} = e^{-i\phi_M}\,, \tag{2.35}$$

where $M_{12} \propto \left(V_{ts}^*V_{tb}\right)^2$ has been used, since the top quark is the dominant contribution in the mixing diagrams (Fig. 2.3) [45].

9 $|q/p|_{B_s^0} = 1.0003 \pm 0.0014$ (experiment), $|q/p|_{B_s^0} = 0.999989 \pm 0.000001$ (theory) [34].

2.4.3 *Time evolution of the flavor eigenstates*

Using Eqs. 2.21 and 2.25, the time evolution of the flavor eigenstates can be described as:

$$\begin{aligned} \left|B_s^0(t)\right\rangle &= g_+(t)\left|B_s^0\right\rangle + \frac{q}{p}g_-(t)\left|\overline{B}_s^0\right\rangle, \\ \left|\overline{B}_s^0(t)\right\rangle &= g_+(t)\left|\overline{B}_s^0\right\rangle + \frac{p}{q}g_-(t)\left|B_s^0\right\rangle, \end{aligned} \tag{2.36}$$

where $g_+(t)$ and $g_-(t)$ are time-dependent coefficients defined in terms of the light and heavy time-dependent mass eigenvalues:

$$\begin{aligned} g_+(t) &= e^{-im_s t}e^{-\frac{\Gamma_s}{2}t}\left[\quad\cosh\left(\tfrac{\Delta\Gamma_s t}{4}\right)\cos\left(\tfrac{\Delta m_s t}{2}\right) - i\sinh\left(\tfrac{\Delta\Gamma_s t}{4}\right)\sin\left(\tfrac{\Delta m_s t}{2}\right)\right], \\ g_-(t) &= e^{-im_s t}e^{-\frac{\Gamma_s}{2}t}\left[-\sinh\left(\tfrac{\Delta\Gamma_s t}{4}\right)\sin\left(\tfrac{\Delta m_s t}{2}\right) + i\cosh\left(\tfrac{\Delta\Gamma_s t}{4}\right)\cos\left(\tfrac{\Delta m_s t}{2}\right)\right]. \end{aligned} \tag{2.37}$$

From Eq. 2.36 the probability that a particle produced as B_s^0 ($\overline{B}_s^0$) at $t = 0$ has oscillated to a $\overline{B}_s^0$ (B_s^0) after a time t can be calculated as:

$$\begin{aligned} \left|\left\langle B_s^0(0)\middle|\overline{B}_s^0(t)\right\rangle\right|^2 &= \left|\frac{p}{q}\right|^2 |g_-(t)|^2, \\ \left|\left\langle \overline{B}_s^0(0)\middle|B_s^0(t)\right\rangle\right|^2 &= \left|\frac{q}{p}\right|^2 |g_-(t)|^2, \end{aligned} \tag{2.38}$$

with

$$|g_-(t)|^2 = \frac{e^{-\Gamma_s t}}{2}\left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) - \cos(\Delta m_s t)\right]. \tag{2.39}$$

It is worth to notice that the mixing probabilities for B_s^0 and $\overline{B}_s^0$ in Eq. 2.38 do not have to be equal, as the ratio $|q/p|$ does not have to be equal to one. This is one of the sources of CP violation (indirect CPV) that is discussed in Section 2.5.

In summary, the B_s^0 meson flavor eigenstates oscillate with frequency Δm_s, modulated by an exponential decay of time constant $1/\Gamma_s$, modified by the decay width $\Delta\Gamma_s$. The current world-average values for m_s, Δm_s, Γ_s and $\Delta\Gamma_s$ are reported in Tab. 2.1.

Table 2.1: World-average values for variables describing the B_s^0 system.

Variable	World-average [34]	Unit
m_s	5366.88 ± 0.14	MeV
Δm_s	17.749 ± 0.020	$\hbar\text{ps}^{-1}$
Γ_s	0.6600 ± 0.0016	ps^{-1}
$\Delta\Gamma_s$	0.085 ± 0.004	ps^{-1}

2.5 CP VIOLATION IN B MESONS

To better describe CPV, the following amplitudes are defined for a given $B^0_{(s)} \to f$ transition where a neutral B^0_s or B^0 meson decays into a final state f:

$$\begin{aligned} A_f &= \left\langle f \middle| H \middle| B^0_{(s)} \right\rangle, & \overline{A}_f &= \left\langle f \middle| H \middle| \overline{B}^0_{(s)} \right\rangle, \\ A_{\overline{f}} &= \left\langle \overline{f} \middle| H \middle| B^0_{(s)} \right\rangle, & \overline{A}_{\overline{f}} &= \left\langle \overline{f} \middle| H \middle| \overline{B}^0_{(s)} \right\rangle. \end{aligned} \tag{2.40}$$

Since a generic final state can always be redefined by a phase transformation $|f\rangle \to e^{i\alpha}|f\rangle$ without changing any physical observables, in describing CPV it is useful to use rephasing-invariant quantities, such as $|A_f|$, $|\overline{A}_f|$ and $|q/p|$. Another quantity invariant under phase redefinition and useful to parametrize CPV is the complex parameter λ_f, defined as:

$$\lambda_f \equiv \frac{q}{p}\frac{\overline{A}_f}{A_f}. \tag{2.41}$$

Using the mixing formalism presented in Section 2.4.3, the time-dependent decay rate $\Gamma_{B^0_{(s)} \to f}(t) \equiv \left| \left\langle f \middle| H \middle| B^0_{(s)}(t) \right\rangle \right|^2$ of an initial state $B^0_{(s)}$ to decay into f after a time t, and the corresponding $\Gamma_{\overline{B}^0_{(s)} \to f}(t)$ rate, can be expressed as[10]:

$$\begin{aligned} \Gamma_{B^0_{(s)} \to f}(t) = |A_f|^2 \quad \frac{1}{1+C_f} e^{-\Gamma_{(s)} t} \Bigg[&\cosh\left(\frac{\Delta\Gamma_{(s)} t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma_{(s)} t}{2}\right) \\ &+ C_f \cos\left(\Delta m_{(s)} t\right) - S_f \sin\left(\Delta m_{(s)} t\right) \Bigg], \end{aligned} \tag{2.42}$$

$$\begin{aligned} \Gamma_{\overline{B}^0_{(s)} \to f}(t) = |A_f|^2 \left|\frac{p}{q}\right|^2 \frac{1}{1+C_f} e^{-\Gamma_{(s)} t} \Bigg[&\cosh\left(\frac{\Delta\Gamma_{(s)} t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma_{(s)} t}{2}\right) \\ &- C_f \cos\left(\Delta m_{(s)} t\right) + S_f \sin\left(\Delta m_{(s)} t\right) \Bigg], \end{aligned} \tag{2.43}$$

with

$$C_f \equiv \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \quad S_f \equiv \frac{2\,\mathrm{Im}(\lambda_f)}{1+|\lambda_f|^2} \quad \text{and} \quad D_f \equiv -\frac{2\,\mathrm{Re}(\lambda_f)}{1+|\lambda_f|^2}. \tag{2.44}$$

CP is violated if Eq. 2.42 and 2.43 are different, which is the case if either λ_f or $|q/p|$ are different from one. CPV effects can be classified in various categories, depending on how they affect these variables.

10 As the discussion is identical for B^0 and B^0_s mesons, the decay width and decay width (mass) difference are indicated with $\Gamma_{(s)}$ and $\Delta\Gamma_{(s)}$ ($\Delta m_{(s)}$), respectively.

2.5.1 *CP violation in decay*

Direct CP violation occurs when the decay rate of a $B^0_{(s)}$ to a final state f and the decay rate of the CP-conjugated process $\overline{B}^0_{(s)} \to \bar{f}$ are not equal, i. e.:

$$\left| \frac{\overline{A}_{\bar{f}}}{A_f} \right| \neq 1 \,. \tag{2.45}$$

If the final state is a CP eigenstate, λ_f can be expressed in terms of A_f and $\overline{A}_{\bar{f}}$ as:

$$\lambda_f = \eta_{CP} \frac{q}{p} \frac{\overline{A}_{\bar{f}}}{A_f} \,, \tag{2.46}$$

with

$$|\bar{f}\rangle = CP|f\rangle = \eta_{CP}|f\rangle \,, \tag{2.47}$$

where $\eta_{CP} = \pm 1$ is the CP eigenvalue of f. Eq. 2.46 shows that, assuming $|q/p| = 1$, CP violation in decay is equivalent with a deviation from unity of $|\lambda_f|$.

The amplitudes can be factorized in their magnitude $|A_f|$, their strong phase term $e^{i\theta}$, and their weak phase term $e^{i\phi}$, as:

$$\begin{aligned} A_f &= |A_f| e^{i(\theta+\phi)} \\ \overline{A}_{\bar{f}} &= |A_f| e^{i(\theta-\phi)} \,, \end{aligned} \tag{2.48}$$

where the amplitudes and strong phases are invariant under CP[11]. Since physical observables are proportional to $|A_f|^2$, to produce measurable CPV at least two interfering amplitudes are needed[12]:

$$\begin{aligned} A_f &= |A_1| e^{i(\theta_1+\phi_1)} + |A_2| e^{i(\theta_2+\phi_2)} \,, \\ \overline{A}_{\bar{f}} &= |A_1| e^{i(\theta_1-\phi_1)} + |A_2| e^{i(\theta_2-\phi_2)} \,. \end{aligned} \tag{2.49}$$

CP violation can be then measured from the difference in the rates as:

$$\left|A_f\right|^2 - \left|\overline{A}_{\bar{f}}\right|^2 = -4|A_1||A_2| \sin(\theta_1 - \theta_2) \sin(\phi_1 - \phi_2) \,. \tag{2.50}$$

Equation 2.50 shows that in order to generate observable direct CPV the following requirements are needed:

1. at least two interfering amplitudes,
2. at least two different weak phases, so that $\sin(\phi_1 - \phi_2) \neq 0$,
3. at least two different strong phases, so that $\sin(\theta_1 - \theta_2) \neq 0$.

---

11 In absence of a weak phase from the CKM matrix, A_f and $\overline{A}_{\bar{f}}$ are the same complex number in the Standard Model.

12 With only one amplitude $|A_f|^2$ is identical to $|\overline{A}_{\bar{f}}|^2$.

The largest possible direct CPV occurs when two amplitudes of equal magnitude have a strong and weak phase difference of $\pi/2$. Direct CPV is the only type allowed for charged mesons, as they are not subject to mixing.

Direct CP violation has been discovered in kaons decay [46–48], and later observed in B [31, 32] and D [33] mesons decays.

2.5.2 *CP violation in mixing*

CP can be violated in mixing when the probability of the transition $B^0_{(s)} \to \overline{B}^0_{(s)}$ is different from the one of $\overline{B}^0_{(s)} \to B^0_{(s)}$:

$$\left|\left\langle B^0_{(s)}(0)\middle|\overline{B}^0_{(s)}(t)\right\rangle\right|^2 \neq \left|\left\langle \overline{B}^0_{(s)}(0)\middle|B^0_{(s)}(t)\right\rangle\right|^2 . \tag{2.51}$$

By using Eq. 2.38, the previous inequality can be rewritten as:

$$\left|\frac{p}{q}\right|^2 |g_-(t)|^2 \neq \left|\frac{q}{p}\right|^2 |g_-(t)|^2 , \tag{2.52}$$

showing that CP is violated in mixing if:

$$\left|\frac{q}{p}\right| \neq 1 \quad \leftrightarrow \quad |q| \neq |p| . \tag{2.53}$$

Assuming no direct CPV, this leads to a deviation from unity of $|\lambda_f|$ in the Standard Model. CP is conserved in B mesons mixing to a high degree, since $|q/p| \approx 1$ [34], but it has been observed to be violated in K^0 oscillations [30].

2.5.3 *CP violation in the interference between decay and mixing*

If the final state is accessible by both $B^0_{(s)}$ and $\overline{B}^0_{(s)}$, CP can also be violated as a result of the interference between the decay of mixed and unmixed B mesons in the same final state:

$$\Gamma\left(B^0_{(s)(\rightsquigarrow \overline{B}^0_{(s)})} \to f\right)(t) \neq \Gamma\left(\overline{B}^0_{(s)(\rightsquigarrow B^0_{(s)})} \to f\right)(t) . \tag{2.54}$$

Notable examples of such final states are CP eigenstates. This type of CPV is generated by a non-zero complex phase of λ_f:

$$\mathrm{Im}(\lambda_f) \neq 0 . \tag{2.55}$$

Using Eq. 2.42 and 2.43, and assuming $|q/p| \approx 1$[XIII], the time dependent CP asymmetry can be expressed as:

$$a_{CP}(t) = \frac{\Gamma_{\overline{B}^0_{(s)} \to f}(t) - \Gamma_{B^0_{(s)} \to f}(t)}{\Gamma_{\overline{B}^0_{(s)} \to f}(t) + \Gamma_{B^0_{(s)} \to f}(t)} = \frac{S_f \sin(\Delta m t) - C_f \cos(\Delta m t)}{\cosh(\frac{\Delta\Gamma t}{2}) - D_f \sinh(\frac{\Delta\Gamma t}{2})} \,. \tag{2.56}$$

Here, $S_f \propto 2\,\mathrm{Im}(\lambda_f)$ and $D_f \propto -\mathrm{Re}(\lambda_f)$ probe CPV in decay/mixing interference, while $C_f \propto 1 - |\lambda_f|^2$ is sensitive to direct CPV.

CP violation in the interference of mixing and decay has been observed in K^0 and B^0 mesons [35, 49].

GOLDEN MODES Decays where all leading diagrams carry the same CP-violating phase are called *golden modes* and satisfy $|A_f| = |\overline{A}_f|$, so that there is no direct CP violation. Mathematically this lead to $|\lambda_f| = 1$, simplifying Eq. 2.56 to:

$$a_{CP}(t) = \frac{\mathrm{Im}(\lambda_f) \sin(\Delta m t)}{\cosh(\frac{\Delta\Gamma t}{2}) + \mathrm{Re}(\lambda_f) \sinh(\frac{\Delta\Gamma t}{2})} \,. \tag{2.57}$$

In golden modes involving B mesons, the CP-violating phase can be easily computed, as the phase of $\overline{A}_f/A_f$ is trivially extracted from the phase of the CKM elements and the phase of q/p can be computed from Eq. 2.35.

XIII This is the case for B^0 and B^0_s mesons.

3 CP VIOLATION IN $B_s^0 \to J/\psi\,\phi$

The $B_s^0 \to J/\psi\,\phi$ decay is of particular interest to study CP violation in the interference of mixing and decay since the predicted complex phase $\phi_s^{c\bar{c}s}$ (hereafter ϕ_s) is one of the most theoretically clean CPV observables in the Standard Model. This process is characterized by a CP eigenstate of $J/\psi\,\phi$ in the final state, where ϕ refers to the $\phi(1020)$ strangeonium resonance, accessible to both B_s^0 and $\overline{B}_s^0$, with and without mixing, as schematically represented in Fig. 3.1. The Feynman diagrams of the dominant tree-level amplitude and the higher-order penguin contribution are shown in Fig. 3.2. This process is relatively rare, with a branching ratio of $\mathcal{B}(B_s^0 \to J/\psi\,\phi) \approx 10^{-3}$ [34]. In the measurement presented in this dissertation, the decay is reconstructed in the final state where the J/ψ and the ϕ mesons decay to two muons and two charged kaons, respectively. The branching ratio for this final states is $\mathcal{B}(J/\psi \to \mu^+\mu^-) \cdot \mathcal{B}(\phi \to K^+K^-) \approx 0.06 \cdot 0.5 \approx 0.03$ [34], leading to a total branching ratio of:

$$\mathcal{B}(B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-) \approx 3 \cdot 10^{-5}\,. \tag{3.1}$$

This chapter describes the CP violation phenomenology of this decay and the mathematical tools useful to probe it experimentally.

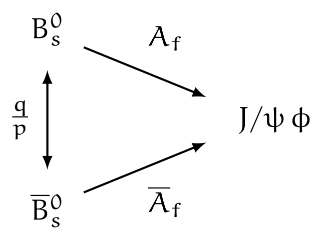
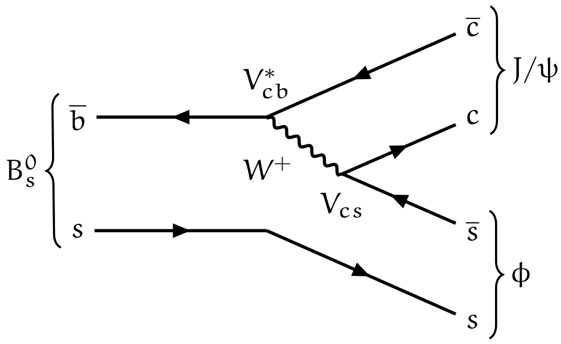
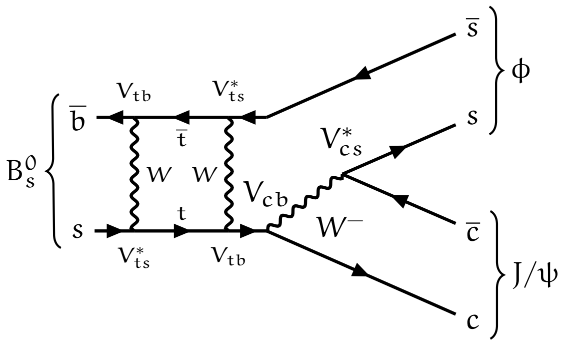


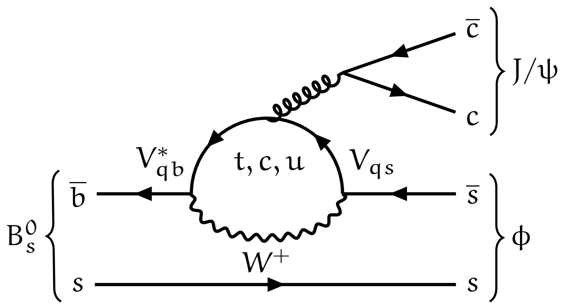
Figure 3.1: Schematic representation of the possible paths of B_s^0 mesons to the $J/\psi\,\phi$ final state.



(a) Tree-level amplitude.



(b) Tree-level amplitude with mixing. In the box diagram only the leading terms, related to the top quark, are shown. An alternative mixing diagram is shown in Fig. 2.3.



(c) Higher-order loop amplitude ("penguin"). This contribution will be neglected in this dissertation.

Figure 3.2: Relevant Feynman diagrams for the $B_s^0 \to J/\psi\,\phi$ decay. The process is dominated by the tree amplitude with or without mixing.

3.1 THE CP-VIOLATING PHASE ϕ_s

The decay amplitude $A_{J/\psi\phi}$ can be written in terms of the tree (t) and penguin (p_q) contributions as [34]:

$$\begin{aligned} A_{f_{CP}} &= (V^*_{cb}V_{cs})t + \sum_{q=u,c,t} (V^*_{qb}V_{qs})p_q \\ &= (V^*_{cb}V_{cs})(t - p_c - p_t) + (V^*_{ub}V_{us})(p_u - p_t) \\ &= (V^*_{cb}V_{cs})T + (V^*_{ub}V_{us})P\,, \end{aligned} \tag{3.2}$$

where $T \equiv t - p_c - p_t$, $P \equiv p_u - p_t$, and the following unitary relation has been used to reduce the number of CKM combinations: $V^*_{tb}V_{ts} = -V^*_{cb}V_{cs} - V^*_{ub}V_{us}$. In Eq. 3.2, the T and P contributions carry different weak phases, with the latter being suppressed by a factor "*loop* $\times \lambda^2$", where "loop" refers to a $\mathcal{O}(0.2\text{–}0.3)$ penguin versus tree-suppression factor and $\lambda \simeq 0.23$ is the Wolfenstein expansion parameter. The penguin transitions that carry different weak phase, with respect to the tree-level one, are therefore suppressed by a factor $\mathcal{O}(0.01)$.

By neglecting the penguin contributions, $B^0_s \to J/\psi\,\phi$ can be treated as a *golden mode* decay. In fact, penguin transitions are predicted to change the value of ϕ_s only by about ~0.05° [50], a shift orders of magnitude smaller than the current experimental precision on ϕ_s (~2° [35]). Under this assumption there is no direct CP violation, as only one amplitude contribute to the decay:

$$\left|A_{f_{CP}}\right| = \left|\overline{A}_{\overline{f}_{CP}}\right|, \quad \text{with } f_{CP} \equiv J/\psi\,\phi\,. \tag{3.3}$$

According to Eq. 2.57, the CP asymmetry in golden modes is completely characterized by the real and imaginary components of $\lambda_{f_{CP}}$. It is worth to notice that, for neutral B meson decays, the golden mode assumption is equivalent to $\left|\lambda_{f_{CP}}\right| = 1$ since $|q/p| \approx 1$ in very good approximation[1]. To ease the reading, the following definitions will be used hereafter:

$$\lambda_{J/\psi\phi} \equiv \lambda\,, \quad f_{J/\psi\,\phi} \equiv f\,, \tag{3.4}$$

unless otherwise specified.

Using Equations 2.47 and 2.49, the tree-amplitudes A_f and $\overline{A}_f$ can be expressed as:

$$\begin{aligned} A_f &= \quad |A_f|e^{i\theta_D}e^{i\phi_D}\,, \\ \overline{A}_f &= \eta_f|A_f|e^{i\theta_D}e^{-i\phi_D}\,. \end{aligned} \tag{3.5}$$

where θ_D and ϕ_D are the decay strong and weak phases, respectively, and η_f is the CP eigenvalue of the final state.

In the $B^0_s \to J/\psi\,\phi$ decay, the CP eigenvalue η_f is not uniquely determined, as the final state is a mixture of CP-even and CP-odd eigenstates. This is a

1 See Section 2.5.2.

consequence of the process being a decay of a pseudo-scalar meson with spin 0 to two vector mesons with spin 1, allowing three values of the final-state orbital momentum. Indeed, the spins in the final state can be pointing parallel, orthogonal, or opposite, leading to an orbital momentum of $\ell = 2$, 1 and 0, respectively, to conserve the spin 0 of the initial state. Knowing this, the CP eigenvalue of the final state can be calculated from the value of the orbital momentum ℓ as:

$$CP|J/\psi\,\phi\rangle_\ell = \eta_f|J/\psi\,\phi\rangle_\ell = (-1)^\ell|J/\psi\,\phi\rangle_\ell\,. \tag{3.6}$$

The $\overline{A}_f/A_f$ ratio is then equal to:

$$\frac{\overline{A}_f}{A_f} = \eta_f e^{-2i\phi_D} = \eta_f \frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*}\,, \tag{3.7}$$

where the amplitudes have been expressed in terms of the associated CKM contributions to the tree-level transition:

$$\begin{aligned} A_f &= \quad V_{cs} V_{cb}^*\,, \\ \overline{A}_f &= \eta_f V_{cs}^* V_{cb}\,. \end{aligned} \tag{3.8}$$

Using Eq. 2.35 and 3.7, λ can be written as:

$$\lambda = \frac{q}{p}\frac{\overline{A}_f}{A_f} = \eta_f \left(\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}\right)\left(\frac{V_{cs}^*V_{cb}}{V_{cs}V_{cb}^*}\right) = \eta_f e^{-i\phi_s}\,, \tag{3.9}$$

where $\phi_s = \phi_M + 2\phi_D$ is the phase associated to the CP violation in the interference between decay and mixing of $B_s^0 \to J/\psi\,\phi$. Using the properties of the complex argument, the CP-violating phase ϕ_s can be computed as[2]:

$$\begin{aligned} \phi_s &= -\arg\left(\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}\frac{V_{cs}^*V_{cb}}{V_{cs}V_{cb}^*}\right) \\ &= -2\arg\left(\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) \\ &= -2\left[\arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) - \pi\right] \\ &= -2\beta_s - 2\pi \\ &= -2\beta_s\,, \end{aligned} \tag{3.10}$$

where β_s is the B_s^0 unitary triangle angle defined in Eq. 2.17.

Equation 3.10 is very important, as it connects the CP-violating phase to one of the angles of the B_s^0 unitary triangle, which can be determined very precisely

2 The periodicity in 2π of complex phases has been used in the last passage.

by a CKM global fit to experimental data. The current best determinations of $-2\beta_s$ are from the CKMfitter Group and UTfit Collaboration:

$$\phi_s = -2\beta_s = \begin{cases} -36.96^{+0.72}_{-0.84}\ \text{mrad} & \text{(CKMfitter [44])} \qquad (3.11) \\ -37 \pm 2 \quad \text{mrad} & \text{(UTfit [51])} \qquad (3.12) \end{cases}$$

which provide extremely precise predictions for ϕ_s. Through this dissertation, the CKMfitter result will be used as SM prediction for comparisons, although using the UTfit result would lead to identical conclusions.

Using Eq. 2.57, the time-dependent CP asymmetry for the decay of a B_s^0 meson to a $J/\psi\,\phi$ final state is given by :

$$\begin{aligned} a_{CP}(t) &= \frac{\Gamma_{\overline{B}_s^0 \to J/\psi\,\phi}(t) - \Gamma_{B_s^0 \to J/\psi\,\phi}(t)}{\Gamma_{\overline{B}_s^0 \to J/\psi\,\phi}(t) + \Gamma_{B_s^0 \to J/\psi\,\phi}(t)} \\ &= \frac{-\eta_f \sin\phi_s \sin(\Delta m_s t)}{\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f \cos\phi_s \sinh\left(\frac{\Delta\Gamma_s t}{2}\right)} \,. \end{aligned} \tag{3.13}$$

3.2 THE $B_s^0 \to J/\psi\,\phi$ DECAY RATE

In this section, the differential decay rate used in the analysis of $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ events is described.

From Eq. 2.42 the time dependent decay rates are:

$$\begin{aligned} \frac{d\Gamma(B_s^0 \to J/\psi\,\phi)}{dt} &\propto |A_f|^2 \frac{1}{1+C} e^{-\Gamma_s t} \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f D \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\qquad \left. + C\cos(\Delta m_s t) - \eta_f S \sin(\Delta m_s t) \right], \\ \frac{d\Gamma(\overline{B}_s^0 \to J/\psi\,\phi)}{dt} &\propto |A_f|^2 \frac{1}{1+C} e^{-\Gamma_s t} \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f D \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\qquad \left. - C\cos(\Delta m_s t) + \eta_f S \sin(\Delta m_s t) \right], \end{aligned} \tag{3.14}$$

where the coefficients C, S and D here are defined using Eq. 3.9 as:

$$C = \frac{1-|\lambda|^2}{1+|\lambda|^2}, \quad S = -\frac{2|\lambda|\sin\phi_s}{1+|\lambda|^2} \quad \text{and} \quad D = -\frac{2|\lambda|\cos\phi_s}{1+|\lambda|^2}, \tag{3.15}$$

with the $|\lambda| = 1$ assumption dropped[3] and η_f factorized out. It is worth to notice that the decay rates in Eq. 3.14 are invariant under the following simultaneous transformations:

$$\begin{cases} \phi_s & \leftrightarrow & \pi - \phi_s\,, \\ \Delta\Gamma_s & \leftrightarrow & -\Delta\Gamma_s\,. \end{cases} \tag{3.16}$$

In this dissertation $\Delta\Gamma_s$ is assumed to be positive, as measured by the LHCb Collaboration in Ref. [52].

Equation 3.14, depends on the CP eigenvalue of the final state η_f, which in turn depends on the final-state orbital momentum as:

$$\eta_f = (-1)^{\ell} \quad \text{for} \quad \ell = 0,\, 1,\, 2\,. \tag{3.17}$$

For a given event the orbital momentum of the final state is unknown, however the different CP eigenstates can be separated statistically, by performing an angular analysis on the final-state particle system.

3.2.1 *The transversity basis*

The different CP eigenstates of the final states can be separated by exploiting the different angular distributions of the final-state particles. In this dissertation, the $B_s^0 \to J/\psi\,\phi$ decay amplitude is decomposed using the transversity basis [53]. This basis is defined by three different polarization states, each related to the spin projections alignment. The three polarization amplitudes, shown in Fig. 3.3, are:

- the CP-even amplitude A_0, corresponding to the longitudinal polarization, where the spins are both aligned to the vector mesons momenta;
- the CP-even amplitude $A_{\parallel}$, corresponding to the parallel transverse polarization, where the spins are parallel to each other;
- the CP-odd amplitude $A_{\perp}$, corresponding to the perpendicular transverse polarization, where the spins are perpendicular to each other.

Each amplitude of the transversity basis is associated with a corresponding strong phase, identified as δ_0, $\delta_{\perp}$, and $\delta_{\parallel}$, which does not change sign under CP. The three amplitudes obey the following normalization condition:

$$|A_0|^2 + |A_{\perp}|^2 + \left|A_{\parallel}\right|^2 = 1\,. \tag{3.18}$$

The three amplitudes of the transversity basis model the resonant $B_s^0 \to J/\psi\,\phi$ decays. However, in any realistic data sample, a small contribution from non-resonant $B_s^0 \to J/\psi\, K^+K^-$ and $B_s^0 \to J/\psi\, f_0(980) \to \mu^+\mu^-\, K^+K^-$ decays is

3 The assumption $|\lambda| = 1$ is directly tested in the measurement.

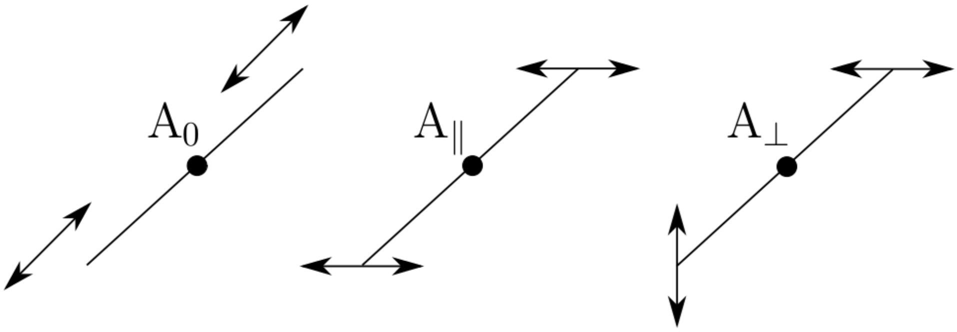


Figure 3.3: Schematic representation of the polarization amplitudes of the $J/\psi\,\phi$ system in the transversity basis. The black circle represents the B_s^0 meson, while the straight lines and the short arrows represent the final-state vector meson momenta and spin orientations, respectively.

always present. These scalar processes are taken into account by including an additional S-wave amplitude A_S, with an associated strong phase δ_S, in the decay model.

In the transversity basis angular distributions of the decay products are described as a function of three decay angles $\Theta = (\theta_T, \psi_T, \varphi_T)$ [53]. The angles θ_T and φ_T are, respectively, the polar and azimuthal angle of the μ^+ in the rest frame of the J/ψ meson, where the x-axis is defined by the B_s^0 direction and the xy-plane is defined by the plane of the $\phi \to K^+K^-$ decay. The angle ψ_T is computed in the ϕ rest frame between the K^+ momentum and the negative J/ψ momentum direction. The three angles are shown in Fig. 3.4.

3.2.2 *The $B_s^0 \to J/\psi\,\phi$ differential decay rate*

By rewriting Eq. 3.14 in the transversity basis one can describe the differential decay rate of $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ in terms of proper decay time and angles of the transversity basis as [54]:

$$\frac{d^4\Gamma(B_s^0)}{d\Theta\, dt} = f(\Theta, t\,|\,\alpha) \propto \sum_{i=1}^{10} O_i(\alpha, t) \cdot g_i(\Theta)\,, \tag{3.19}$$

where:

- O_i are time-dependent functions:

$$O_i(\alpha, t) = N_i e^{-\Gamma_s t}\left[a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ \left. + c_i \cos(\Delta m_s t) + d_i \sin(\Delta m_s t)\right], \tag{3.20}$$

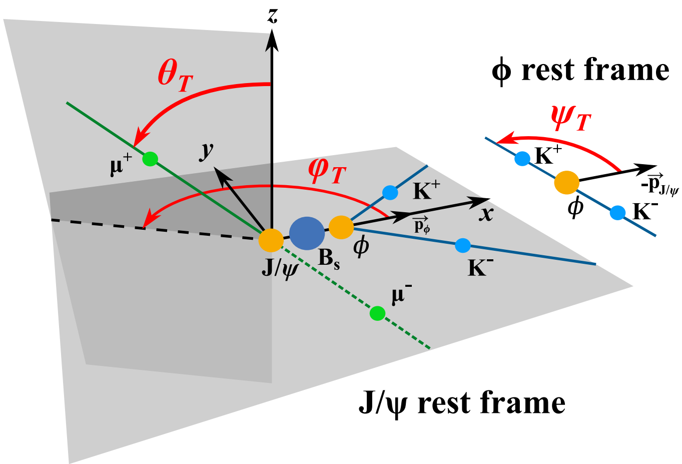


Figure 3.4: Definition of the three angles of the transversity basis θ_T, ψ_T, and φ_T describing the topology of the $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ decay. Figure from Ref. [1].

with N_i, a_i, b_i, c_i, and d_i defined in Tab. 3.1,

- g_i are the angular functions defined in Tab. 3.2,
- $\alpha = (\phi_s, \Delta\Gamma_s, \Gamma_s, \Delta m_s, |\lambda|, |A_0|, |A_\perp|, \left|A_\parallel\right|, |A_S|, \delta_0, \delta_\parallel, \delta_\perp, \delta_S)$ is the set of physics parameters,
- the coefficients C, S, and D in Tab. 3.1 are defined as in Eq. 3.15:

$$C = \frac{1-|\lambda|^2}{1+|\lambda|^2}, \quad S = -\frac{2|\lambda|\sin\phi_s}{1+|\lambda|^2} \quad \text{and} \quad D = -\frac{2|\lambda|\cos\phi_s}{1+|\lambda|^2}. \tag{3.21}$$

The coefficient C measures the amount of direct CP violation, and it is equal to zero if $|\lambda| = 1$. The coefficients S and D are sensitive to CP violation in the interference between decay and mixing, with S being more sensitive to small ϕ_s, which is the case in the Standard Model. The S coefficient is mostly present in terms proportional to d_i, which are dependent on the flavor of the B_s^0 meson.

Equation 3.19 describes the model for the $B_s^0 \to J/\psi\,\phi$ meson decay, the model for $\overline{B}_s^0 \to J/\psi\,\phi$ is obtained by changing the sign of the c_i and d_i terms in Eq. 3.20.

Table 3.1: Time-dependent terms of the $B_s^0 \to J/\psi\,\phi$ differential decay rate model. The terms with $i = 1, 2, 3, 7$ are each related to one of the amplitudes, while the others describe interference terms. All amplitudes included in the N_i coefficients definition are evaluated at $t = 0$. The coefficients C, S and D are defined in Eq. 3.21.

i	N_i	a_i	b_i	c_i	d_i
1	$\lvert A_0\rvert^2$	1	D	C	$-S$
2	$\lvert A_\parallel\rvert^2$	1	D	C	$-S$
3	$\lvert A_\perp\rvert^2$	1	$-D$	C	S
4	$\lvert A_\parallel\rvert\lvert A_\perp\rvert$	$C\sin(\delta_\perp-\delta_\parallel)$	$S\cos(\delta_\perp-\delta_\parallel)$	$\sin(\delta_\perp-\delta_\parallel)$	$D\cos(\delta_\perp-\delta_\parallel)$
5	$\lvert A_0\rvert\lvert A_\parallel\rvert$	$\cos(\delta_\parallel-\delta_0)$	$D\cos(\delta_\parallel-\delta_0)$	$C\cos(\delta_\parallel-\delta_0)$	$-S\cos(\delta_\parallel-\delta_0)$
6	$\lvert A_0\rvert\lvert A_\perp\rvert$	$C\sin(\delta_\perp-\delta_0)$	$S\cos(\delta_\perp-\delta_0)$	$\sin(\delta_\perp-\delta_0)$	$D\cos(\delta_\perp-\delta_0)$
7	$\lvert A_S\rvert^2$	1	$-D$	C	S
8	$\lvert A_S\rvert\lvert A_\parallel\rvert$	$C\cos(\delta_\parallel-\delta_S)$	$S\sin(\delta_\parallel-\delta_S)$	$\cos(\delta_\parallel-\delta_S)$	$D\sin(\delta_\parallel-\delta_S)$
9	$\lvert A_S\rvert\lvert A_\perp\rvert$	$\sin(\delta_\perp-\delta_S)$	$-D\sin(\delta_\perp-\delta_S)$	$C\sin(\delta_\perp-\delta_S)$	$S\sin(\delta_\perp-\delta_S)$
10	$\lvert A_S\rvert\lvert A_0\rvert$	$C\cos(\delta_0-\delta_S)$	$S\sin(\delta_0-\delta_S)$	$\cos(\delta_0-\delta_S)$	$D\sin(\delta_0-\delta_S)$

Table 3.2: Angular terms of the $B_s^0 \to J/\psi\,\phi$ differential decay rate model. The transversity basis angles are defined in Fig. 3.4.

i	$g_i(\theta_T, \psi_T, \varphi_T)$
1	$2\cos^2\psi_T(1-\sin^2\theta_T\cos^2\varphi_T)$
2	$\sin^2\psi_T(1-\sin^2\theta_T\sin^2\varphi_T)$
3	$\sin^2\psi_T\sin^2\theta_T$
4	$-\sin^2\psi_T\sin 2\theta_T\sin\varphi_T$
5	$\frac{1}{\sqrt{2}}\sin 2\psi_T\sin^2\theta_T\sin 2\varphi_T$
6	$\frac{1}{\sqrt{2}}\sin 2\psi_T\sin 2\theta_T\cos\varphi_T$
7	$\frac{2}{3}(1-\sin^2\theta_T\cos^2\varphi_T)$
8	$\frac{1}{3}\sqrt{6}\sin\psi_T\sin^2\theta_T\sin 2\varphi_T$
9	$\frac{1}{3}\sqrt{6}\sin\psi_T\sin 2\theta_T\cos\varphi_T$
10	$\frac{4}{3}\sqrt{3}\cos\psi_T(1-\sin^2\theta_T\cos^2\varphi_T)$

3.3 STATE-OF-THE-ART

The CP-violating phase ϕ_s was first measured by the Fermilab Tevatron experiments [55–59], and then at the LHC by the ATLAS, CMS, and LHCb experiments [60–69], using $B_s^0 \to J/\psi\,\phi$, $B_s^0 \to J/\psi\,f_0(980)$, and $B_s^0 \to J/\psi\,h^+h^-$ decays, where $h = \pi$, K. Measurements of ϕ_s in B_s^0 decays to $\psi(2S)\phi$ and $D_s^+D_s^-$ were performed by the LHCb Collaboration [70, 71].

With the same analysis used to extract ϕ_s, several other parameters of the differential decay rate model can be measured, with Γ_s, $\Delta\Gamma_s$ being the most common choices. The available theoretical predictions for the physics parameters of Eq. 3.19 are reported in Tab. 3.3, while the most recent world-average values are reported in Tab. 3.4. The strong phase δ_0 is not reported, as the general convention [56, 58] is to set it at zero, since only phase differences are physically meaningful (see Tab. 3.1). The two-dimensional world average contours for ϕ_s and $\Delta\Gamma_s$ are shown in Fig. 3.5, from which it is clear that the experimental uncertainties on ϕ_s are significantly larger than the theoretical ones. A more updated picture of the two-dimensional contours for ϕ_s and $\Delta\Gamma_s$, including the latest preliminary results from ATLAS [63] and CMS [1] (this dissertation) and the latest $\Delta\Gamma_s$ theoretical prediction, is presented in Fig. 3.6.

3.4 NEW-PHYSICS CONTRIBUTIONS

New physics (NP) can modify ϕ_s via the contribution of beyond Standard Model (BSM) particles to the B_s^0–$\overline{B}_s^0$ mixing diagrams shown in Fig. 2.3 [72, 73]. These effects can be parametrized by a general modification of the off-diagonal matrix element of the mass matrix M_{12} in terms of an arbitrary complex parameter Δ_s:

$$M_{12} = M_{12}^{SM} |\Delta_s| e^{i\phi_s^{\Delta}} . \tag{3.22}$$

In this formalism, any $\Delta_s \neq 1$ corresponds to new-physics effects, with the absolute value affecting Δm_s and the complex phase modifying ϕ_s, as:

$$\Delta m_s = 2\left|M_{12}^{SM}\right| |\Delta_s| , \tag{3.23}$$

$$\phi_s = \phi_s^{SM} + \phi_s^{\Delta} . \tag{3.24}$$

Experimental results can be used in conjunction with SM predictions to obtain bounds on the complex parameter Δ_s. The most recent result of such an investigation is shown in Fig. 3.7, which shows good agreement between the SM global fit and $\Delta_s = 1 + 0i$. The constraint due to the mass difference Δm_s is denoted by the orange ring and its size is mostly due to theory uncertainties, while the purple region represents the constraint on the imaginary part of Δ_s stemming from ϕ_s measurements. Since ϕ_s in the SM is predicted very precisely, the bounds on ϕ_s^{Δ} are mostly limited by experimental uncertainties.

In this formalism, new-physics effects are still allowed to modify ϕ_s by several degrees [73, 74][4], providing strong motivations to improve the measurement of ϕ_s, as any reduction of the experimental uncertainties directly leads to better sensitivity for BSM effects.

Table 3.3: Available theoretical predictions for the physics parameters describing the $B_s^0 \to J/\psi\, \phi$ decay rate.

Parameter	Prediction	Ref.
ϕ_s	$-36.96\,^{+0.72}_{-0.84}$ mrad	[44]
$\Delta\Gamma_s$	$0.091 \pm 0.013\,\mathrm{ps}^{-1}$	[75]
Δm_s	$18.77 \pm 0.86\,\hbar\mathrm{ps}^{-1}$	[75]
$\lvert\lambda\rvert$	≈ 1	SM

Table 3.4: World-averages for the physics parameters describing the $B_s^0 \to J/\psi\, \phi$ decay rate. The strong phase δ_0 is not reported, as the general convention [56, 58] is to set it at zero. The S-wave parameters A_S and δ_S are not reported, as no world-averages are available for these observables (at the time of writing).

Parameter	World-average	Ref.
ϕ_s	-51 ± 23 mrad	[35]
$\Delta\Gamma_s$	0.085 ± 0.004 ps^{-1}	[35]
Γ_s	0.6600 ± 0.0016 ps^{-1}	[34]
Δm_s	$17.749 \pm 0.020\ \hbar\mathrm{ps}^{-1}$	[35]
$\lvert\lambda\rvert$	1.012 ± 0.017	[34]
$\lvert A_0\rvert^2$	0.5199 ± 0.0035	[34]
$\lvert A_\parallel\rvert^2$	0.228 ± 0.007	[34]
$\lvert A_\perp\rvert^2$	0.245 ± 0.004	[34]
$\delta_\parallel$	3.10 ± 0.06 rad	[34]
$\delta_\perp$	2.9 ± 0.4 rad	[34]

4 The weak phase ϕ_s is predicted to be $\approx 2°$

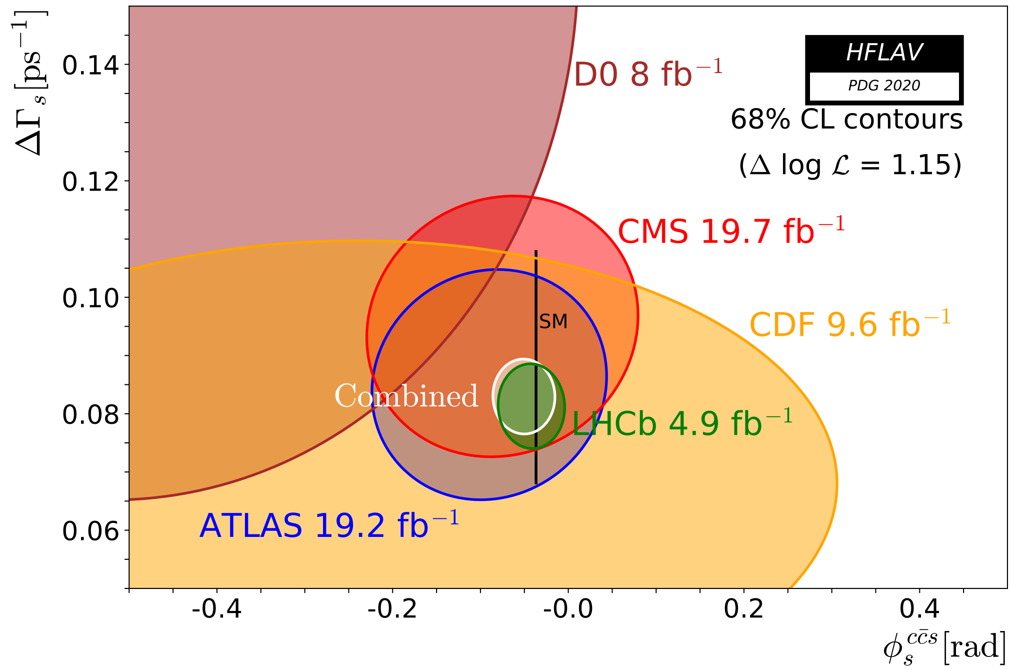


Figure 3.5: Two-dimensional $(\phi_s, \Delta\Gamma_s)$ contours at 68% confidence level of the ATLAS, CMS, CDF, and LHCb results, together with their combined contour, and the Standard Model predictions from Ref. [44, 73]. The plotted $\Delta\Gamma_s$ prediction is not up to date with the most recent (and precise) result (Ref. [75]). Note that the LCHb result here presented includes final state different from $J/\psi K^+K^-$. This figure does not include the results presented in this dissertation [1] and the latest ATLAS preliminary result [63]. Figure from Ref. [35]. A more updated, yet preliminary, picture of the two-dimensional $(\phi_s, \Delta\Gamma_s)$ contours is presented in Fig. 3.6.

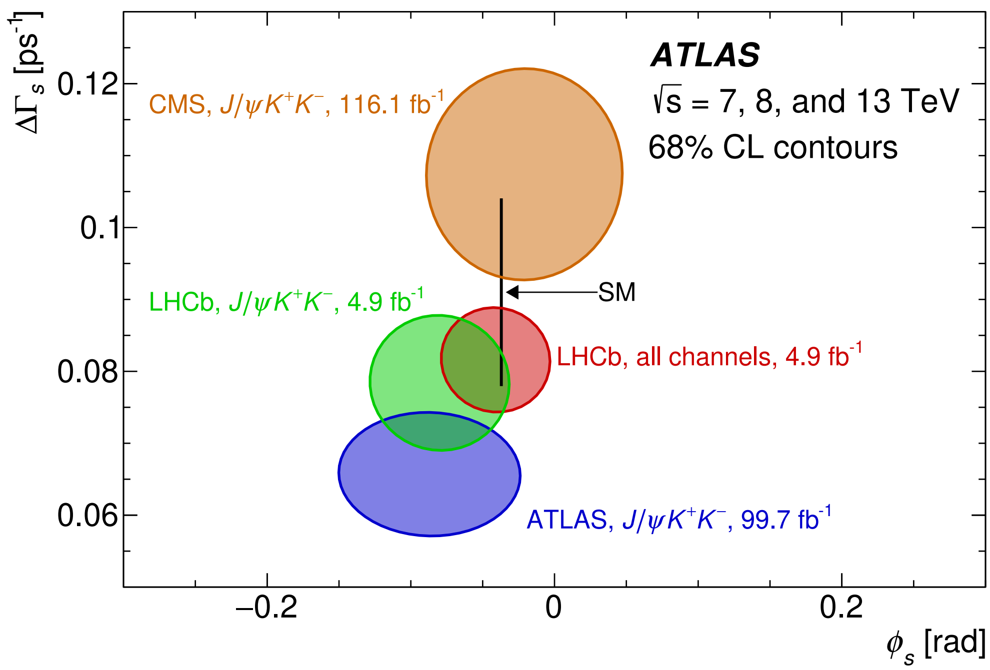


Figure 3.6: Two-dimensional $(\phi_s, \Delta\Gamma_s)$ contours at 68% confidence level of the latest results from ATLAS, CMS, and LHCb, together with the Standard Model predictions from Ref. [74, 75]. Figure from Ref. [63].

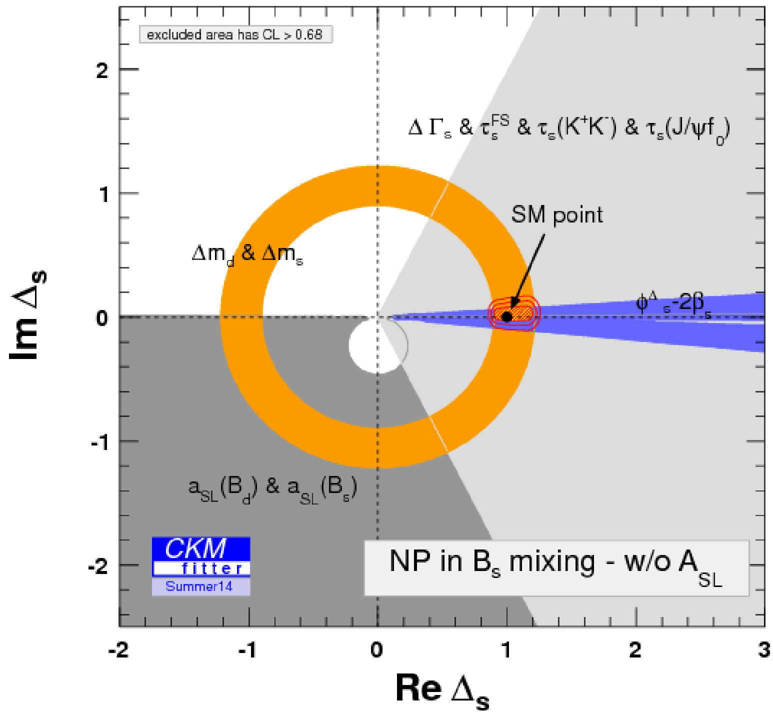


Figure 3.7: Constraint of the NP parameter Δ_s. The constrain due to the mass difference Δm_s is denoted by the orange ring, while the constrain due to ϕ_s by the purple region. The red area shows the combined fit to experimental data. Figure from Ref. [74].

Part II

EXPERIMENTAL SETUP

In order to characterize the $B^0_s \rightarrow J/\psi\,\phi$ decay rate, a complex experimental setup is required. B^0_s mesons need to be produced in large quantities and their decay products reconstructed with high precision. The Large Hadron Collider is the most powerful particle accelerator ever built and is able to collide protons at a center-of-mass energy of 13 TeV at a rate of over one billion collisions per second. It is located at CERN, the largest particle physics laboratory in the world. During the data-taking periods of interest for this work, the LHC has produced approximately $4 \cdot 10^{12}$ B^0_s mesons at the interaction point around which the Compact Muon Solenoid apparatus is located. The CMS detector is a large general-purpose detector, consisting of several sub-systems to measure different types of particles and their properties, for a total of over 124 million readout channel. This huge stream of information is combined in sophisticated reconstruction algorithms to obtain a full event description that can be analyzed to perform physics measurements. The following chapters provide a brief review of the experimental environment of the LHC and the CMS detector, as well as the techniques and algorithms used at CMS to reconstruct the physics objects of interest for this work.

4 THE CMS DETECTOR

The analysis presented in this dissertation is based on a data sample of $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ candidates collected in proton-proton (pp) collisions provided by the Large Hadron Collider (LHC) and collected by the Compact Muon Solenoid (CMS) experiment.

At the time of writing, the LHC is the largest and most powerful particle accelerator ever built, while CMS is a general-purpose detector made up of several sub-systems for measuring different types of particles and their properties. The LHC is located at the European Organization for Nuclear Research (CERN), near Geneva, whose main function is to provide particle accelerators and other infrastructures needed for particle physics research. Several important achievements in the field of particle physics have been made by experiments at CERN, such as the discovery of weak neutral currents in the Gargamelle bubble chamber [76], the discovery of the *W* and *Z* bosons by the UA1 and UA2 experiments [77, 78], the discovery of direct CP violation in the NA48 experiment [48], the determination of the number of light neutrino families by the LEP experiments [79–82], and the discovery of the Higgs boson by the ATLAS and CMS experiments [83, 84].

This section describes the experimental setup of the LHC and CMS experiment, with particular attention to the sub-systems of most interest for this analysis.

4.1 THE LARGE HADRON COLLIDER

The LHC is a dual-ring hadron accelerator, storage ring, and collider operated by CERN and situated in the 26.7 km tunnel formerly built for the Large Electron-Positron Collider (LEP)[1], at a depth between 45 m and 170 m underground on the French-Swiss border near Geneva. The LHC is designed to accelerate protons and heavy-ions[2] to energies up to 7 TeV and 2.76 TeV per nucleon, respectively. The machine was commissioned in 2008 and the first particle collisions were achieved in 2010. Since only pp collisions are used in this analysis, the following description only considers protons.

1 The LEP was an e^+e^- accelerator and collider operated by CERN from 1989 and 2000. The machine serviced four detectors: ALEPH, DELPHI, L3, and OPAL.
2 Predominantly lead nuclei.

4.1.1 *Acceleration chain*

Before being injected into the LHC, protons are pre-accelerated by the accelerator chain shown in Fig. 4.1. The proton source consists of a bottle of molecular hydrogen gas, where an electric field is used to strip hydrogen atoms of their electrons and a duoplasmatron creates the first beam with an energy of 90 keV [85]. Subsequently, the protons are accelerated in four stages by the LINAC2, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), to energies of 50 MeV, 1.4 GeV, 25 GeV and 450 GeV, respectively. The PS also arranges the beam in groups of protons called "*bunches*", each containing $\approx 1.15 \cdot 10^{11}$ protons. Each bunch is injected in the LHC with a frequency of 40 MHz, corresponding to a time separation of 25 ns. During the data-taking periods of interest for this analysis a maximum number of bunches of 2556 in each direction was achieved.

The beams are accelerated in the LHC from 450 GeV up to 6.5 TeV[III] by resonant waves of EM fields generated in eight radio-frequency (RF) cavities per beam, at the rate of 16 MeV per turn. The acceleration process takes around 20 to 25 minutes and the time period for which the LHC has bunches circulating is called "*fill*". A fill is typically maintained for several hours, up to thirty. Around the beam pipes, 1232 dipole superconducting magnets are installed to bend the beams into a circular trajectory. Additionally, quadrupole and higher multipole[4] magnets are used to focus the beam despite the electromagnetic repulsion of the protons. At the interaction points of the LHC, the beams can be confined to a diameter in the order of ≈ 10 µm. A cross section of an LHC dipole magnet is shown in Fig. 4.2.

The center-of-mass energy of a collision of the two proton beams is equal to

$$\sqrt{s} = 2E_{\text{beam}} \,, \tag{4.1}$$

though the energy available to the hard parton scattering is smaller and dependent on the amount of energy carried by the involved parton. The beams are made to collide at four interaction points (IP) around the LHC ring, where the following main detectors are located:

- *ALICE*, a dedicated detector to study heavy-ion physics [88],
- *ATLAS*, a general-purpose detector [89],
- *CMS*, a general-purpose detector (thoroughly described in Section 4.2),
- *LHCb*, a low luminosity detector dedicated to heavy flavor physics [90].

The ATLAS and CMS experiments have similar physics programs, which are achieved with different strategies in the detector layout. The main difference

III This is slightly less than the final design energy of 7 TeV per beam.
4 Sextupoles and octupoles.

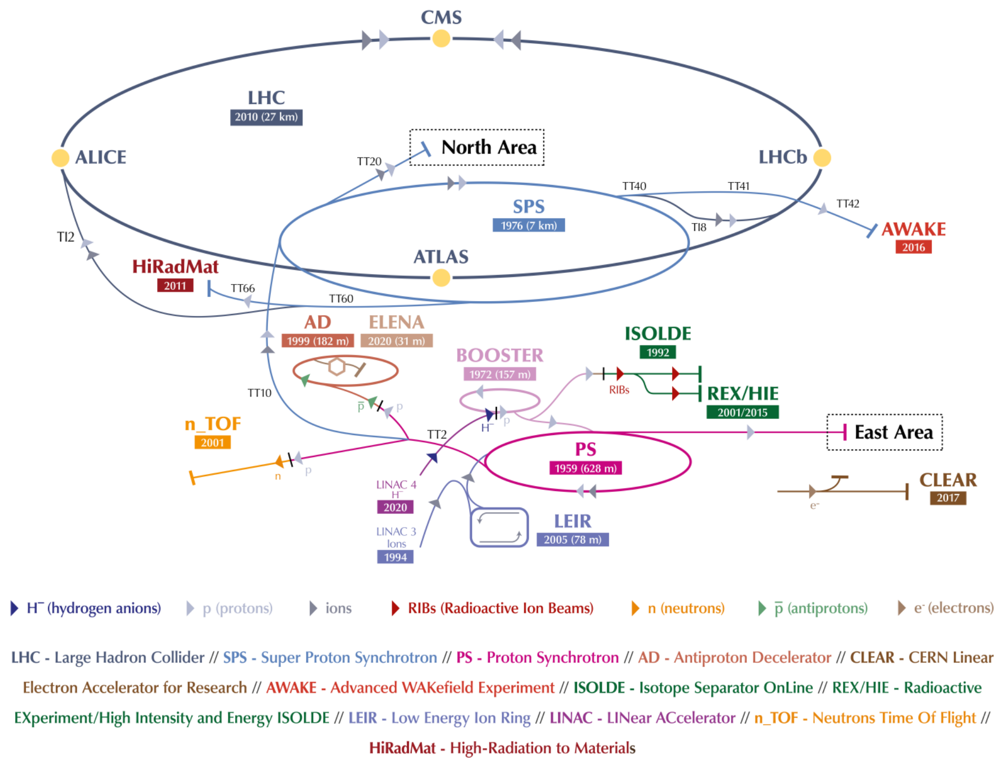


Figure 4.1: The CERN accelerator complex. Figure from Ref. [86].

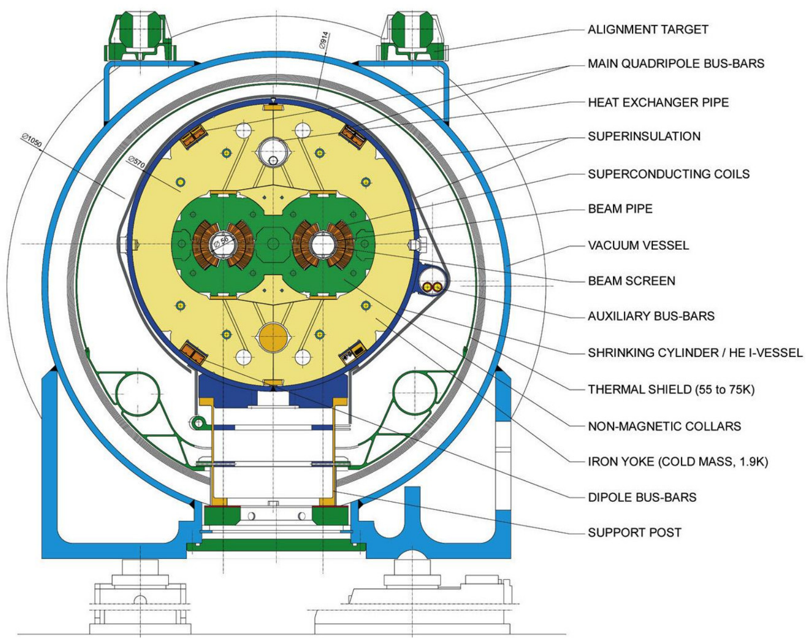


Figure 4.2: Cross section of an LHC dipole magnet. Figure adapted from Ref. [87].

is that the ATLAS superconducting solenoid houses only the tracking system, while in CMS also most of the calorimeter system is present inside the solenoid volume. The overall performances of the two detectors are similar when averaged over a large amount of different analysis scenario[5]. Since this analysis concerns heavy flavor physics phenomenology the LHCb experiment is a prime candidate for comparisons, however, the ATLAS detector is more similar in terms of phase space and absence of a dedicated hadronic particle identification (PID) system.[6]

4.1.2 *Particle production*

At every bunch crossing, only few of the protons actually collides. In the majority of the cases, particles merely graze each other in what is called an *elastic* collision, in which the proton structure is unaltered. Most of the physics of interest occurs only when a proton undergoes an *inelastic* (hard) collision in which its fundamental constituents (i. e., quarks and gluons) interact with the ones of another proton. The rate of events for a given process is given by:

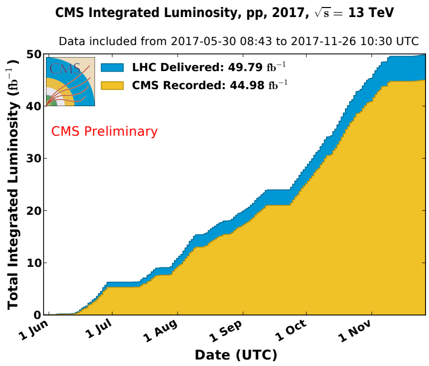
$$\frac{dN}{dt} = L\,\sigma\,, \tag{4.2}$$

where σ is the process cross section and L is the LHC instantaneous luminosity. The instantaneous luminosity measures the ability of a particle accelerator to produce interactions and can be calculated from the beam parameters as [91]:

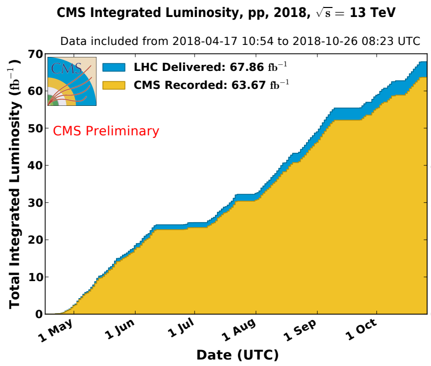
$$L = \frac{N_b^2 n_b f_{rev} \gamma_r F}{4\pi\epsilon_n \beta^*}\,, \tag{4.3}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per bean, f_{rev} is the revolution frequency, γ_r is the relativistic Lorentz factor, F is a reduction factor due to the inclined beam profile[7], ϵ_n is the normalized transverse beam emittance, and β^* is the beta function at the interaction point[8]. The design instantaneous luminosity of the LHC is 10^{34} $\text{cm}^{-2}\text{s}^{-1}$, which has been exceeded in the 2017 and 2018 runs by a factor of ≈ 2 [92]. The nominal luminosity described by Eq. 4.3 is typically achieved only at the start of a fill and decays over time due to collisions and beam losses. At the LHC the luminosity

5 Performance on individual analyses may differ.
6 At the LHC the CP-violating phase ϕ_s is studied by ATLAS, CMS, and LHCb, with comparable precision on single-channel measurements. However, the LHCb experiment is able to exploit different decay channels.
7 F is defined as $F = 2\left[1 + (\theta_c \sigma_z / 2\sigma^*)^2\right]^{-\frac{1}{2}}$, where θ_c is the crossing angle at the interaction point, σ_z is the RMS bunch length and σ^* is the transverse RMS beam size at the interaction point.
8 The transverse emittance describes the spread of the particles in the position-momentum phase space, while the beta function characterize the size of the beam at the interaction point. The transverse beam size at any location s is given by $\sigma(s) = \sqrt{\epsilon_n \beta(s)}$, where $\beta(s) = \beta^*[1 + (s/\beta^*)^2]$.



(a) 2017 data-taking period.



(b) 2018 data-taking period.

Figure 4.3: Cumulative measured luminosity versus day delivered by the LHC to CMS during the 2017 (left) and 2018 (right) pp run at 13 TeV. Figure from Ref. [93].

decrease by a factor of $1/e$ over a time interval called "beam lifetime", which in the 2017 and 2018 runs was equal approximately to 15 hours.

Following Eq. 4.2 total number of events for a given process is given by:

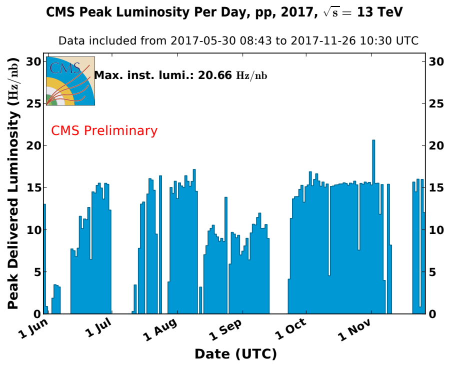
$$N = \int \frac{dN}{dt}\, dt = \sigma \int L\, dt = \sigma \mathcal{L}_{\text{int}}\,, \tag{4.4}$$

where $\mathcal{L}_{\text{int}}$ is the so-called "integrated luminosity" that describes the total number of collisions in a given time interval. The total inelastic pp cross section at $\sqrt{s} = 13$ TeV is estimated to be ≈70 mb[IX], which corresponds to an average of ≈30 collisions per bunch crossing in addition to any event of interest [93][10]. These *uninteresting* inelastic pp collisions are referred to as "pileup". The measured luminosity values delivered by the LHC to CMS during stable beams and for proton-proton collisions in the data-taking periods considered in this work are shown in Figs. 4.3 and 4.4, while the mean number of interactions per bunch crossing is shown in Fig. 4.5.

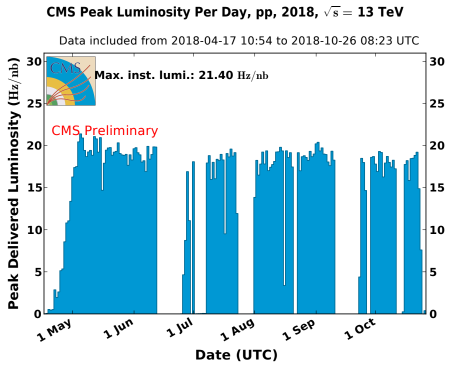
BOTTOM QUARK PRODUCTION AT THE LHC At the LHC bottom quarks are predominantly produced in $pp \to b\bar{b}X$ interaction through *gluon-gluon fusion*, *quark annihilation*, *flavor excitation* and *gluon splitting*. The leading order and next-to-leading order Feynman diagrams for these processes are shown in Figs. 4.6 and 4.7.

The predicted NNLO $b\bar{b}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV is $\sigma(pp \to b\bar{b}X) \approx 500$ μb [94], while its estimated value as a function of the center-of-mass energy is shown in Fig. 4.8. The total number of

IX The value of $\sigma^{pp}_{\text{inelastic}} \approx 70$ mb is the "CMS recommended" value for CMS analyses [93].
10 Processes of interest typically have a very small cross section, as such it is extremely rare than more than one occurs in the same bunch crossing.

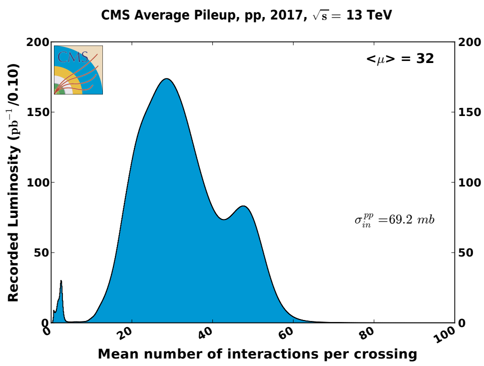


(a) 2017 data-taking period.

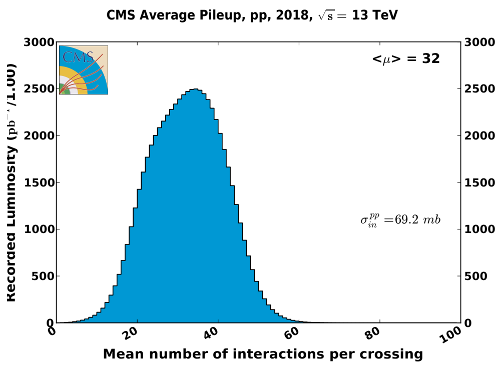


(b) 2018 data-taking period.

Figure 4.4: Peak day-by-day instantaneous luminosity delivered by the LHC to CMS during the 2017 (left) and 2018 (right) pp run at 13 TeV. Figure from Ref. [93].



(a) 2017 data-taking period.



(b) 2018 data-taking period.

Figure 4.5: Mean number of interactions per bunch crossing for the 2017 (left) and 2018 (right) pp run at 13 TeV, using the online luminosity values. The 2017 distribution has multiple peaks due to data taking-periods with different peak instantaneous luminosity, as can be seen from Fig. 4.4. Figure from Ref. [93].

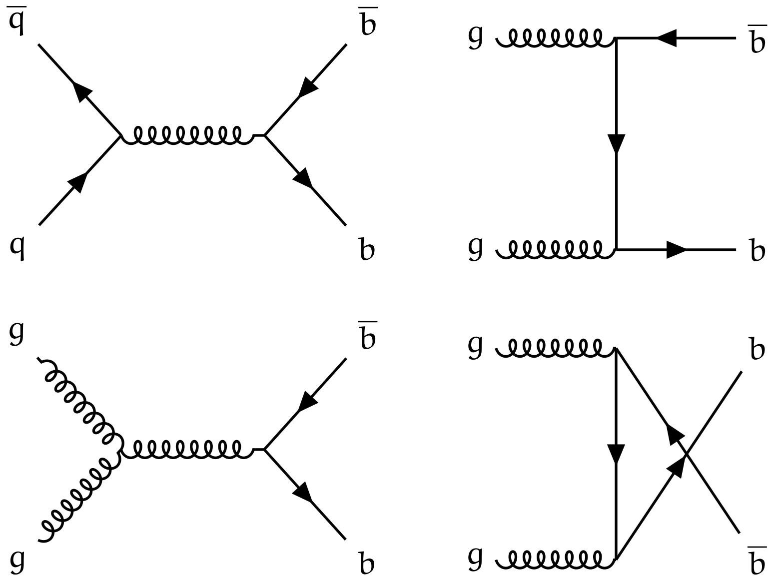


Figure 4.6: Leading order Feynman diagrams for b-quark pair production: $q\bar{q}$ annihilation (top left) and gluon-gluon fusion (others).

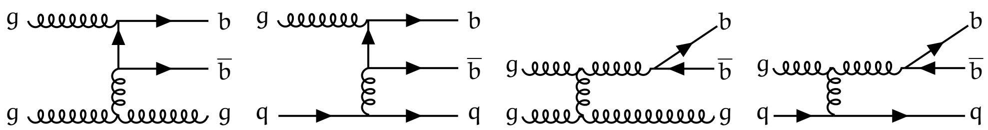


Figure 4.7: Next-to-leading order Feynman diagrams for $b\bar{b}$ pair production: flavor excitation (left and middle-left) and gluon splitting (middle-right and right).

$b\bar{b}$ pairs produced in the 2017 and 2018 data-taking periods can be calculated with Eq. 4.4, as:

$$N_{b\bar{b}} = \sigma(pp \to b\bar{b}X) \cdot \mathcal{L}_{int} \approx 5.5 \cdot 10^{13}, \tag{4.5}$$

where the integrated luminosity values reported in Fig. 4.3 are used[11]. Mind that this is the total number of $b\bar{b}$ pairs produced in pp collisions and not the number of those produced in the CMS geometrical acceptance.

11 For a description of the data samples used for this analysis, and respective luminosity values see Tab. 7.4 in Section 7.2.2.

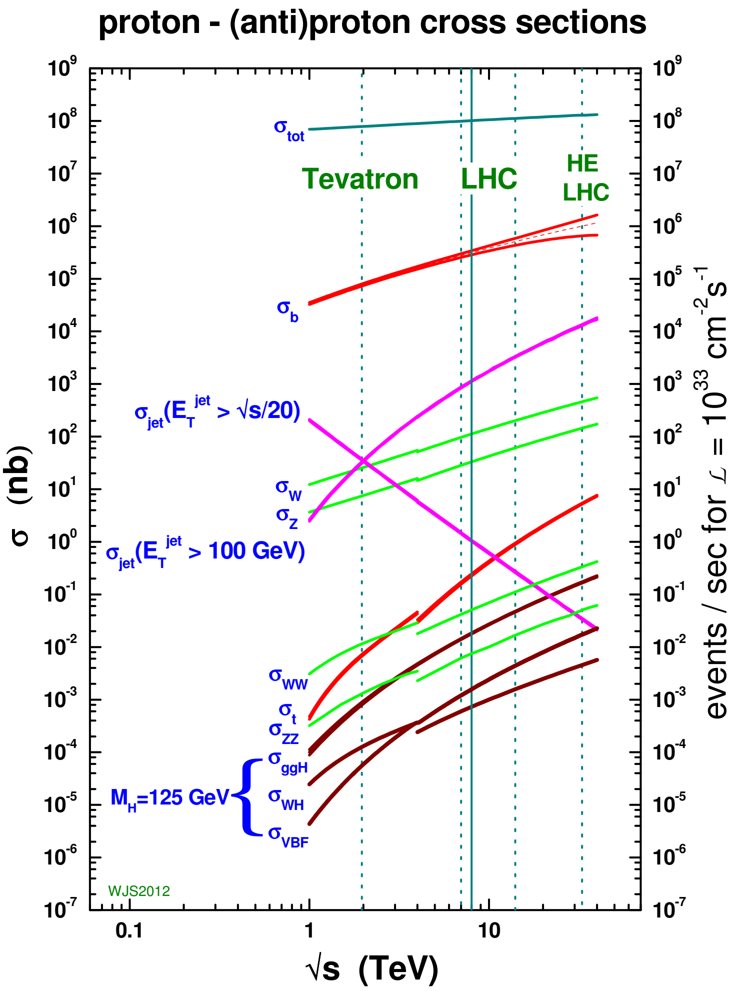


Figure 4.8: Production cross section at NLO in pp collisions of several process of interests as a function of the center-of-mass energy. Figure from Ref. [95].

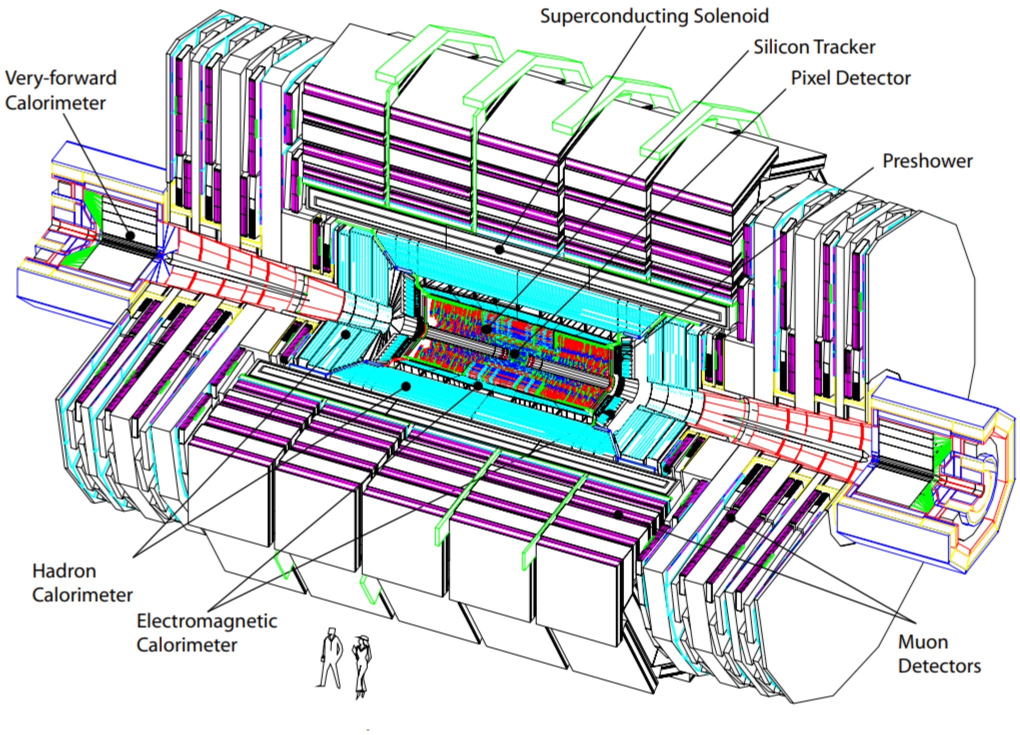


Figure 4.9: Overview of the CMS detector and its sub-systems. Figure from Ref. [96].

4.2 THE CMS DETECTOR

The Compact Muon Solenoid is a large general-purpose apparatus for investigating proton and heavy-ion collisions provided by the LHC at high energies and high instantaneous luminosity [96]. It is located at the opposite end of the LHC ring from the main CERN site, near the city of Cessy, at about 100 m underground. CMS is constructed in 15 circular *slices* that surround the LHC beam pipe in a cylindrical form of 21 m overall length and 15 m diameter. With a total weight of about 14 000 t, it is the heaviest LHC detector. An overview of the CMS detector and its sub-systems is shown in Fig. 4.9. At the time of writing, the CMS Collaboration is composed of over 5000 scientists, engineers, technicians, and students from around 229 institutions from more than 50 countries.

COORDINATE SYSTEM To describe the positions and directions, CMS adopts a coordinate system centered at the nominal LHC collision point, with the x-axis pointing inward to the center of the LHC ring and the y-axis pointing upward. The z-axis is tangent to the beam with a direction that completes a right-handed coordinate system. A mixture of cartesian, cylindrical, and spherical coordinates is used in this dissertation depending on the cases. The azimuthal angle[12] ϕ is measured in the x-y plane (often called "transverse plane") relative to the x-axis.

12 Not to be confused with the transversity angle φ_T.

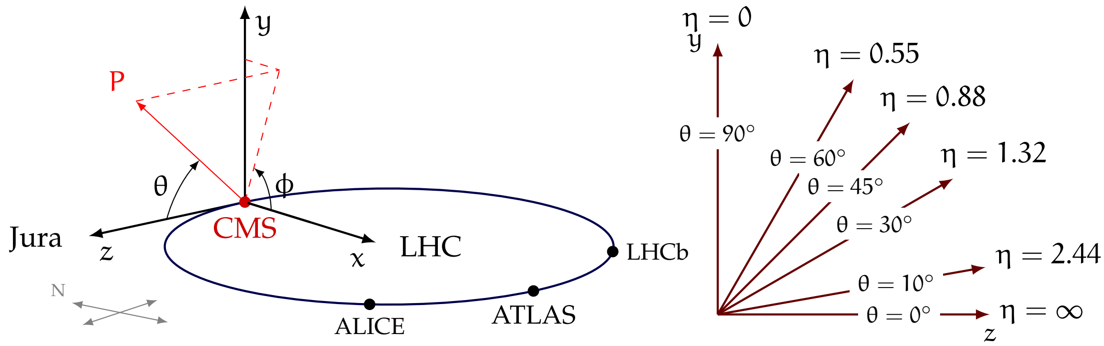


Figure 4.10: Illustration of the CMS coordinate system (left) and the relation between the pseudorapidity η and the polar angle θ (right). "Jura" refers to the mountain range near CERN. Figures from Ref. [97].

The transverse plane is used to define the transverse momentum p_T and energy E_T, which are extremely convenient in a detector with cylindrical symmetry such as CMS. The polar angle θ is measured from the z-axis and it used to define the pseudorapidity, as:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{4.6}$$

The pseudorapidity is typically preferred over the polar angle θ as the rate of particle production at hadron colliders is approximately constant as a function of η since pseudorapidity differences are Lorentz invariant under boosts along the longitudinal axis. Illustrations of the CMS coordinate system and the pseudorapidity relationship with the polar angle are shown in Fig. 4.10.

4.2.1 *Overall design*

The CMS sub-detectors are arranged radially around the beam axis, enveloping the interaction point completely in the azimuthal direction and providing good coverage in the polar one[13]. The detector is divided into a *barrel* component and two *endcaps*, which typically correspond to different type of detector technologies.

SUPERCONDUCTING SOLENOID The core feature of CMS is a large superconducting solenoid of 12.5 m length and 6.3 m internal diameter [98], which is used to generate a uniform 3.8 T magnetic field inside the solenoid. The magnetic field is used to bend the trajectory of charged particles, enabling the measurement of their momenta and charge sign. The solenoid is composed of four layers of coiled superconducting NbTi[XIV] cooled down to 4.6 K with

13 The exact pseudorapidity coverage depends on the individual sub-detectors, ranging from $|\eta| < 2.4$ for the muon system up to $|\eta| < 5.2$ for the forward hadronic calorimeter.

XIV The same technology is used in the LHC magnets.

liquid helium, conducting a nominal current of 19.41 kA. At the time of writing, the CMS magnet is the most powerful single magnet ever built, storing a total energy of 2.6 GJ in its field. The solenoid surrounds the tracking and calorimetry systems in order to not add *dead* material in front of them. It is complemented with a 12 500 t iron return yoke in which the magnetic field saturates to a value of 1.8 T. The return yoke has the dual purpose of reducing the intensity of the magnet field on components far from the magnet and guiding the magnetic field outside the bore of the magnet. A mapping of the magnetic field generated by the superconducting solenoid is shown in Fig. 4.11.

SUB-SYSTEMS LAYOUT The tracking system is placed in the most inner part of the detector to achieve the best possible spatial resolution and provides an accurate measurement of the charged particles momenta, charge, and trajectories [100–103]. The tracker is surrounded by the calorimetry system which comprises a lead tungstate crystal electromagnetic calorimeter (ECAL) [104] and a brass and scintillator hadron calorimeter (HCAL) [105]. The ECAL detects particles that interact electromagnetically, such as electrons, positrons, and photons, whereas the HCAL is designed for detecting charged and neutral hadrons, i. e., protons, neutrons, pions, and kaons. Finally, outside the solenoid, gas detectors that make up the muon system are installed between the layers of the return yoke[15] [106]. A transverse slice of the CMS detector and its sub-system is shown in Fig. 4.12. The following sections present further details of each sub-detector, while the trigger system and physics objects reconstruction are described in Chapter 5.

4.2.2 *Tracking system*

The CMS tracking system ("tracker") is installed close to the interaction point and it is entirely based on silicon semiconductor technology [100, 101]. It is designed to provide precise and efficient measurements of the momenta, charge, and trajectories of charged particles, as well as an accurate reconstruction of primary and secondary *vertices*[16]. Since this analysis makes use only of data collected between 2017 and 2018, this section describes only the CMS *Phase-1* tracker, comprising the upgraded inner pixel detector installed in March 2017 [102, 103].

As the innermost detector, the tracker faces a highly challenging environment. At the design luminosity of the LHC, on average up to 10^3 particles hit the tracker every 25 ns bunch crossing, which implies a rate density of 1 MHz/mm^2 at $r = 4$ cm, 60 kHz/mm^2 at $r = 22$ cm and 3 kHz/mm^2 at $r = 115$ cm.

15 The iron return yoke has the additional purpose of absorbing most of the particles different from muons.

16 The vertex reconstruction procedure is described in Section 5.2.1.

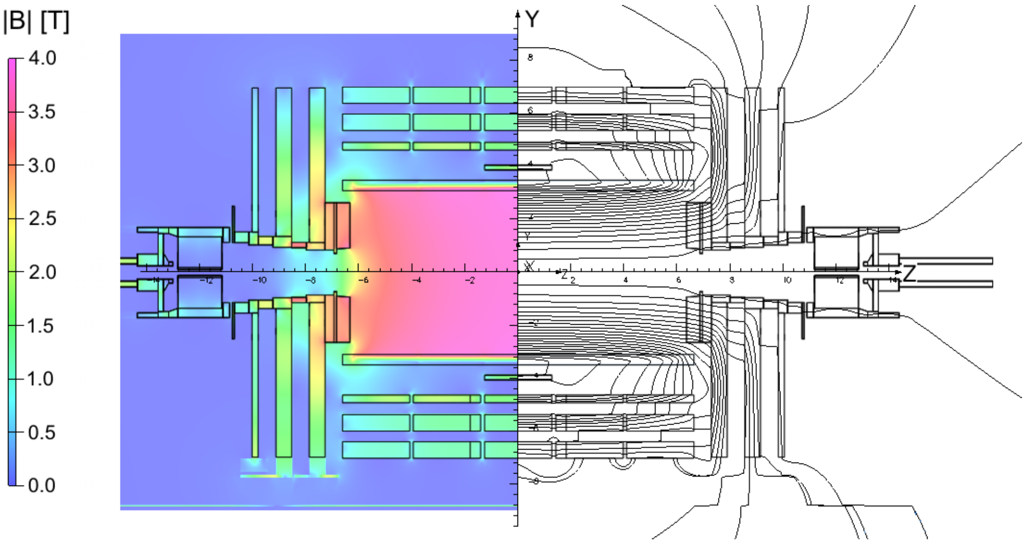


Figure 4.11: Value of |B| (left) and field lines (right) predicted on a longitudinal section of the CMS detector at a central magnetic field of 3.8 T. Figure from Ref. [99].

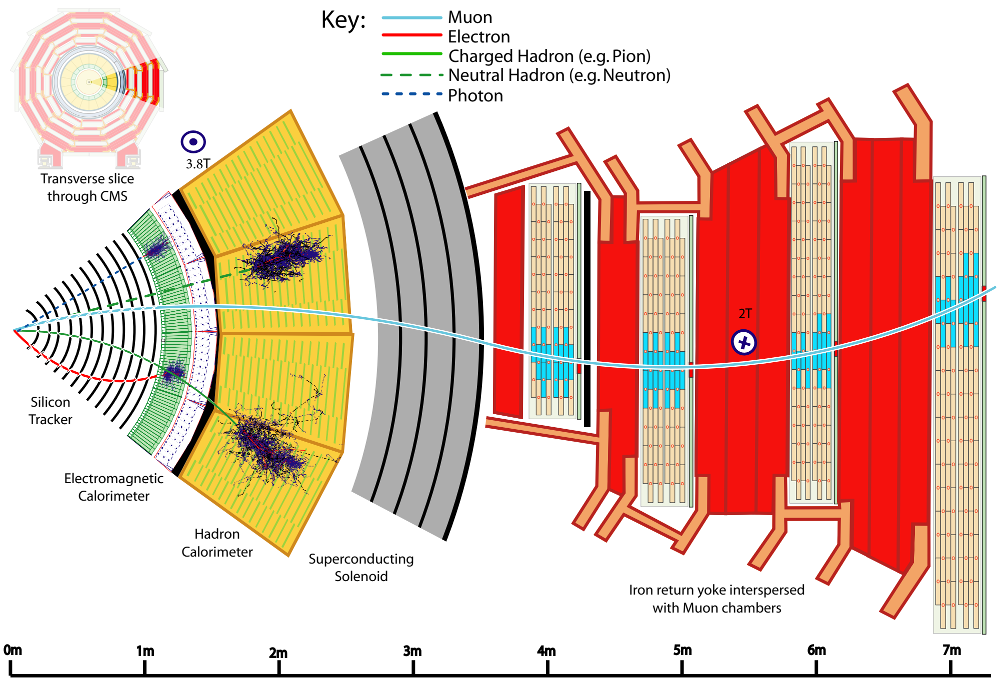


Figure 4.12: View of a transverse slice of the CMS detector with its sub-systems. A schematic view of the specific particle interactions, from the beam interaction region to the muon system, is also shown. The muon and the charged pion are positively charged, and the electron is negatively charged. Figure from Ref. [107].

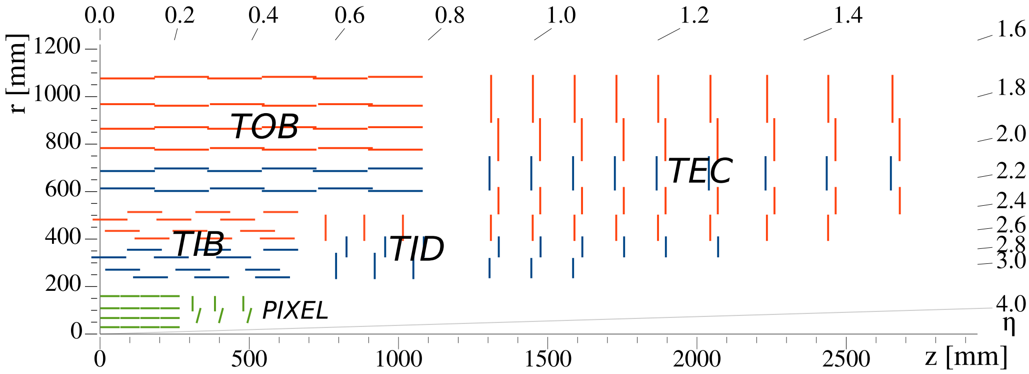


Figure 4.13: Sketch of one quarter of the CMS Phase-1 tracking system, in r-z view, showing the tracker pixel (PIXEL), inner barrel (TIB), inner disk (TID), outer barrel (TOB), and endcap (TEC) subsystems. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively. Figure from [109].

To keep the hit occupancy[17] below 1%, the tracking system consists of high-granularity pixel detectors up to 16 cm in radius, while microstrip detectors are used in the outer region with reduced flux. In total, the tracking system covers a pseudorapidity region up to $|\eta| < 2.5$, has a length of 5.4 m, a diameter of $\approx$2.4 m, and an active silicon area of about 200 m^2, making it the largest silicon tracker ever constructed[18]. All tracking components are cooled down to a temperature of $-10\,°$C to improve radiation hardness. The overall used material, comprising also support structures, cables, and cooling system, is optimized to keep the material budget to a minimum. Depending on the pseudorapidity region, the total material budget of the Phase-1 tracker is equivalent to 0.4–1.5 radiation lengths[19] [108]. A schematic view of the Phase-1 tracker is shown in Fig. 4.13.

PIXEL TRACKER The inner pixel detector covers a pseudorapidity region of $|\eta| < 2.5$ in the innermost region of the tracker, and by extension of the CMS apparatus, where the flux of particles produced is higher. It consists of four layers in the barrel region at radii of 2.9, 6.8, 10.9, and 16.0 cm, and three fan-shaped disks on each endcap at distances of 29.1, 39.6, and 51.6 mm from the center of the detector. The pixel detector is built from 1856 segmented silicon sensor modules, each consisting of a sensor with 160×416 pixels, for a total of 124 million readout channels. Each pixel has a size of 100×150 μm^2, which amounts to a total net active area of the pixel tracker of 1.9 m^2. This high granularity allows to resolve large amounts of particles in high pileup conditions and to precisely reconstruct tracks and vertices in three dimensions. The spatial

17 The occupancy is defined as the fraction of sensors hit at each bunch crossing.
18 At the time of writing
19 The radiation length X_0 is defined as the mean distance over which the energy of a high-energy particle is reduced by a factor of $1/e$: $-dE/dx = E/X_0$.

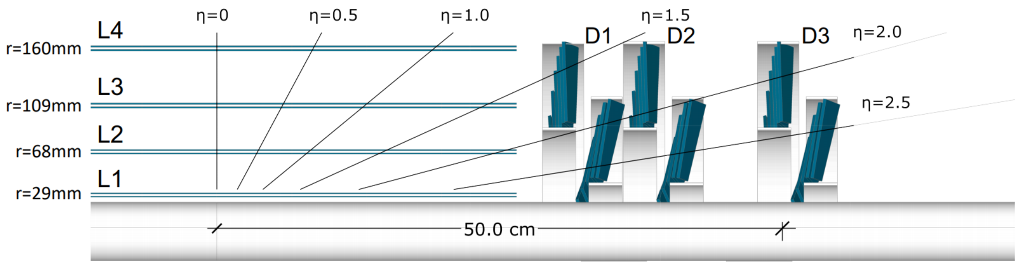


Figure 4.14: Layout of on quadrant of the CMS Phase-1 pixel detector. The four barrel layers are indicated with L1, L2, L3 and L4, whereas the three endcap disks with D1, D2, and D3. Figure from [103].

resolution of the inner tracker in the r-ϕ plane and z direction is 10 µm and 20 µm, respectively [103]. Within a given layer, the modules are overlapped to minimize holes in the acceptance[20]. The inner pixel detector layout is shown in Fig. 4.14.

STRIP TRACKER The outer part of the tracking system consists of silicon microstrips organized in four different sub-systems and covers the radial region between 20 and 116 cm where the occupancy decreases. It is composed of a total of 9.3×10^5 microstrip sensors, for a total active area of $\approx$198 m^2. The tracker inner barrel (TIB) has a length of 1.4 m and consists of four layers of 320 µm thick microstrips sensors, covering radii up to 50 cm The TIB sensors are installed parallel to the beam axis and are purposely misaligned by a tilt angle of 100 mrad to allow the measurement of the z coordinate. The TIB is complemented by the tracker inner disks (TID), consisting of three disks at each endcap, installed from 80 to 90 cm in the z direction. The TID uses 320 µm thick microstrip sensors mounted radially on the disks. The TIB and the TID provide up to four hits for a traversing charged track, each with a spatial resolution of 23–35 µm, up to $|\eta| < 2.5$.

The inner tracking system[21] is surrounded by the tracker outer barrel (TOB) in the central, and by the tracker endcaps (TEC) in the forward regions. The TOB covers the region $50 < r < 116$ cm and extends to $z = \pm 1.2$ m, corresponding to the pseudorapidity region $|\eta| < 1.6$. It is composed of 500 µm thick microstrip sensors, organized in six layers, and provides up to six r-ϕ hits with a single point resolution of 35–53 µm. The TECs consist of nine disks, each composed of four to seven rings of silicon microstrip detectors with a thickness ranging from 320 µm to 500 µm. The microstrips are placed radially, providing up to nine hits per charged track up to $|\eta| < 2.5$.

The tracking performance is discussed in Section 5.2.1, where also the track reconstruction and vertexing algorithms are described.

20 This applies also to the outer tracker layers.
21 The inner tracking system comprises the pixel tracker, the TIB and the TID.

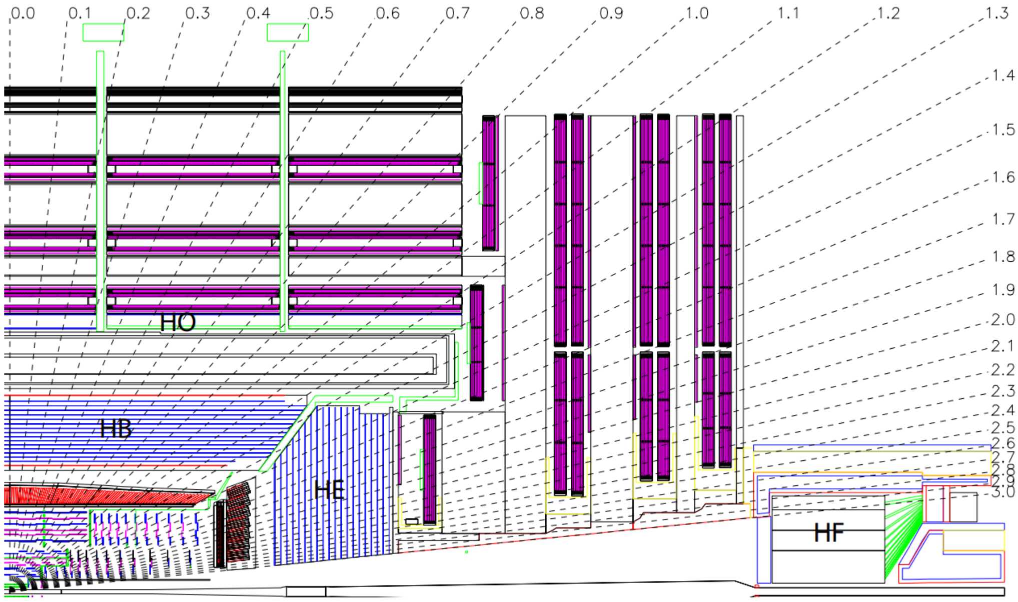


Figure 4.15: Cross section of one quadrant of the CMS calorimetry system showing the locations of the HCAL barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. The radial numbers indicate pseudorapidity values. Figure from Ref. [96].

4.2.3 *Calorimetry system*

The CMS calorimetry system surrounds the tracker and is mostly enveloped by the solenoid in order to minimize the *dead* material in front of it. Its main purpose is the precise energy measurement and full absorption of all particles except for muons and neutrinos. It is divided into two sub-detector with complementary detection strategies. The ECAL measures photons and charged particles, mostly electrons, and positrons, whereas the HCAL detects strongly interacting particles, such as protons, neutrons, pions, and kaons. In order to provide a reliable measurement of missing transverse energy associated with neutrinos, the calorimetry system is fully hermetic in the pseudorapidity region $|\eta| < 5$. An overview of the geometrical layout is shown in Fig. 4.15.

4.2.3.1 *Electromagnetic calorimeter*

The ECAL is a high granularity homogeneous and hermetic calorimeter consisting in 61 200 lead tungstate crystals ($PbWO_4$) mounted in the barrel part (EB), closed by 7324 crystals in each of the two endcaps (EE) [104]. The EB has an inner radius of 129 cm and covers the pseudorapidity region $|\eta| < 1.479$, while the two endcaps cover $1.479 < |\eta| < 3.0$ and are placed at a ± 314 cm distance from the interaction point. In addition preshower detectors (ES) are installed in front of the endcaps in the region $1.653 < |\eta| < 2.6$ to deal with their reduced

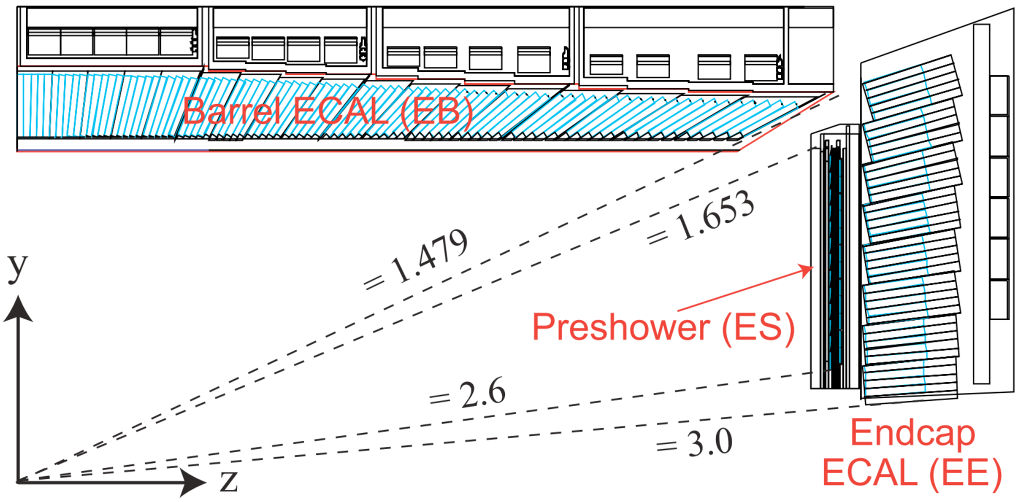


Figure 4.16: Transverse section through the ECAL, showing the geometrical configuration of the ECAL barrel (EB), endcap (EE), and preshower (ES) detectors in terms of the pseudorapidity. Figure from Ref. [110].

granularity. They consist of a 20 cm thick sampling calorimeter[22] composed of two alternating layers of lead absorber planes and silicon microstrip detectors. An overview of the ECAL geometrical configuration is shown in Fig. 4.16.

Lead tungstate is characterized by a high density ($\rho = 8.28$ g/cm^3), a short radiation length ($X_0 = 0.89$ cm) and a small Molière radius[23] ($R_M = 2.19$ cm). Moreover, the scintillation decay time is very short, such that 80% of the light being collected during a single 25 ns bunch spacing. The crystals in the EB have a front cross section of 2.2×2.2 cm^2 and a length of 23 cm, corresponding to $\approx$26 radiation lengths, and a granularity of $\Delta\eta \times \Delta\phi \approx 0.017 \times 0.017$. The EE crystals have slightly different dimensions with 2.86×2.86 cm^2 front cross section and a length of 22 cm. Scintillation light is produced when the crystal is traversed by charged particles or photons and, before readout, it is amplified by avalanche photodiodes (APD) in the EB and vacuum phototriodes in the EE. The average yield for a particle of $E = 1$ MeV is $\approx$4.5 photoelectrons. Since the number of scintillation photons emitted by the crystals and the amplification of the APDs are both temperature-dependent, a cooling system is installed to keep the crystal temperature stable at $18.00 \pm 0.05\,°$C.

The ECAL's main purpose is, in combination with the tracking detector, to identify electrons and photons and to facilitate high-resolution energy measure-

22 A sampling calorimeter is a detector where the absorber material and the scintillating material are interleaved. The absorber material induces a hadronic shower and the scintillating material samples the shower along its length.

23 The Molière radius of a material describes the scale of the transverse dimension of electromagnetic showers. It is defined as the radius of a cylinder containing on average 90% of the shower's energy deposition. Two Molière radii contain 95% of the shower's energy deposition.

ments. Its energy resolution has been measured in electron test beam and found to be [96]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%\,, \tag{4.7}$$

where E is in GeV and $\oplus$ is the squared sum operator. The three terms in Eq. 4.7 refer to the stochastic, noise, and irreducible components of the resolution, respectively.

4.2.3.2 *Hadronic calorimeter*

The HCAL is a hermetic sampling calorimeter that has the purpose of measuring the energy of charged and neutral hadrons, enabling a basic division of the particle types and the measurement of missing energy associated with neutrinos [105]. The HCAL consists of four different sub-detector systems: the HCAL barrel (HB), endcaps (HE), forward (HF), and outer (HO). The HCAL geometrical configuration has been already shown in Fig. 4.15.

The HB and HE consist of alternating layers of non-magnetic brass absorber[24] and plastic scintillators. Traversing particles create showers in the brass layers, which induce detectable light in the subsequent scintillators. The scintillation light is then guided by wavelength-shifting fibers to the readout system, where it is read by hybrid photodiodes. The HB has a length of 9 m, an inner diameter of 6 m and covers the pseudorapidity region $|\eta| < 1.4$. It is segmented in 2304 towers for a granularity of $\Delta\eta \times \Delta\phi \approx 0.087 \times 0.087$. The HE covers a range of $1.3 < |\eta| < 3.0$ and follows a similar structure as the barrel.

The HF is located at a ± 11 m distance from the interaction point, has a length of 165 cm and covers the range $3 < |\eta| < 5$. Its purpose is to improve the hermicity of the HCAL system in order to reinforce the measurement of the missing transverse energy. Due to the high radiation environment in this region of about 100 Mrad/year, the HF is made of more robust quartz fibers as active material and steel absorber plates and is read out by radiation-resistant photomultiplier tubes. The HCAL is completed by the HO, a single layer of 10 mm thick scintillators, corresponding to 1.4 interaction length, covering the barrel region $|\eta| < 1.26$. Its main purpose is to absorb the *tails* of the hadronic showers, leaking through the barrel calorimeters, using the solenoid as absorber. The total thickness of the ECAL+HCAL calorimeter system is about 12 and 10 interaction lengths in the barrel and endcaps, respectively.

The energy resolution of the HCAL has been measured to be [111]:

$$\frac{\sigma_E}{E} = \frac{110\%}{\sqrt{E}} \oplus 9\%\,, \tag{4.8}$$

where E is in GeV, with a typical readout noise of 200 MeV.

---

24 Non-magnetic materials are chosen as the HB and HE are immersed in the 3.8 T magnetic field.

4.2.4 *Muon system*

The CMS muon system constitutes the outermost component of the CMS apparatus and is designed to reconstruct the momenta and charge of muons over a wide angular range up to $|\eta| < 2.4$ [106]. Ideally, the only detectable particles surviving the calorimetry system are muons[25], due to their long lifetime ($\tau_\mu \approx 2.2$ µs) and scarce interaction with matter[26]. The muon system is interlaced with the magnet return yokes and is composed of three different types of gaseous detectors. An overview of its geometrical configuration is shown in Fig. 4.17.

In the barrel region ($|\eta| < 1.2$) a total of 250 *drift tube* (DT) chambers are organized into four layers, called "stations", housed in five cylindrical sections called "wheels". Each DT chamber consists of three "superlayers" (SL) of four stacked cells, filled with a mixture of Ar/Co_2. The central SL measures the z coordinate while the other two r-ϕ. The outermost station does not contain the z-measuring planes. The single point resolution of the DT system is measured to be 80–120 µm [112].

In part of the barrel ($0.9 < |\eta| < 1.2$) and endcap ($1.2 < |\eta| < 2.4$) regions, where the flux of particles is higher and the magnetic field is stronger and inhomogeneous, *cathode strip chambers* (CSC) are chosen for their radiation hardness and fast response. The CSC are multi-wire proportional chambers, which use gas under strong electric fields as active and amplification medium. Each endcap is equipped with four CSC stations, for a total of 540 chambers. The CSCs are positioned perpendicular to the beam axis and overlap in ϕ to avoid gaps in the muon acceptance. A CSC consists of six layers, each capable of measuring all three spatial coordinates (r-ϕ-z). The spatial resolution of a hit in the CSC system is 40-150 µm [112].

The muon system is completed by a trigger system based on *resistive plate chambers* (RPC) that covers the pseudorapidity range $0.0 < |\eta| < 1.9$. RPCs consists of two parallel gas chambers operated in avalanche mode and can achieve a time resolution of a few nanoseconds, which allows assigning the muon system readout to a bunch crossing with rates up to 1 kHz/cm^2. With respect to DTs and CSCs, RPCs have a coarser spatial resolution of ≈1 cm.

The efficiency for reconstructing hits and track segments originating from muons traversing the muon system is in the range ≈94–99%, with a timing resolution of ≈1.4 ns [112]. The momentum resolution for muons up to 100 GeV is ≈1% in the barrel and ≈3% in the endcaps [112]. The muon reconstruction is described in Section 5.3.

25 Neutrinos do not interact with CMS and are not measurable.
26 Muons do not annihilate with ordinary matter and do not decay via the strong force. They can only decay via the weak interaction into an electron and two neutrinos.

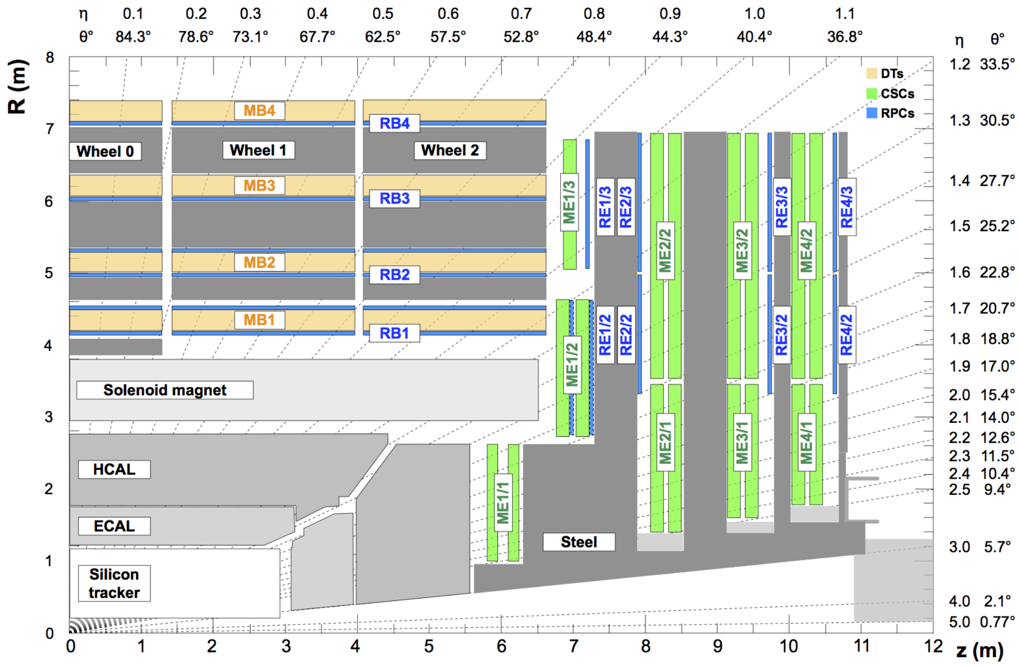


Figure 4.17: Longitudinal r-z cross section of one quadrant of the CMS detector with the muon system highlighted. The DT and CSC stations are labeled MB ("Muon Barrel") and ME ("Muon Endcap"), respectively. RPC stations are mounted in both the barrel and endcaps of CMS, and labeled RB and RE, respectively. Figure from Ref. [112].

5 TRIGGER AND OBJECT RECONSTRUCTION

At the design instantaneous luminosity of $L \approx 10^{34}$ cm^{-2}s^{-1}, the LHC provides approximately 10^9 pp collisions per second [110], which are typically grouped in time around the 40 MHz bunch-crossing rate[1]. The term *event* is used to refer to all the collisions, and corresponding physics processes, occurring in a given bunch crossing. Considering the full information from all sub-detectors, the storage size of an event in CMS is of the order of 1 MB [113], which leads to a data output for the CMS detector of ~40 TB/s, an amount impossible to store and process in a viable way. Moreover, only a few of the produced events contain *exciting* physics, with the majority of them being of scarce interest. The CMS data acquisition (DAQ) and trigger systems are designed to reduce the acquisition rate by a factor of $\approx 10^6$, by selecting only a small fraction of the events for storage and further analysis [114–118].

After an event is selected for permanent storage, the ≈135 million readout channels[2], of the CMS detector need to be combined into a finite number of "physics objects", which can be used to perform physics analyses. This procedure is called physics object reconstruction and, together with the trigger, is the cornerstone of every CMS physics results.

In this chapter, a brief description of the CMS trigger system and the procedures used to reconstruct the physics objects is given. A short summary of the Monte Carlo (MC) techniques used to generate the simulated samples used thorough this work is also presented.

5.1 TRIGGER AND DATA ACQUISITION

To select the fraction of events of interest, i. e., those which exhibit signatures compatible with the CMS physics program, a two-level trigger system is used, where the "Level-1" (L1) trigger operates on fast-response data from the calorimeters and muon detectors, whereas the "high-level" trigger (HLT) has access to a rough reconstruction of the complete event.

The "Level-1" (L1) trigger consists of custom hardware trigger processors and is used to reduce the rate down to 100 kHz. It has a maximum processing time of 4 µs, in which it needs to decide if an event should be passed to the HLT system or permanently discarded. Due to the extremely tight time constraint, the L1 trigger only exploits partial information from the calorimetry and muon

1 The number of collisions per bunch crossing is around 30, as shown in Fig. 4.5.
2 The number of readout channels has increased from ≈75 to ≈135 millions with the installation of the upgraded Phase-1 inner pixel tracker.

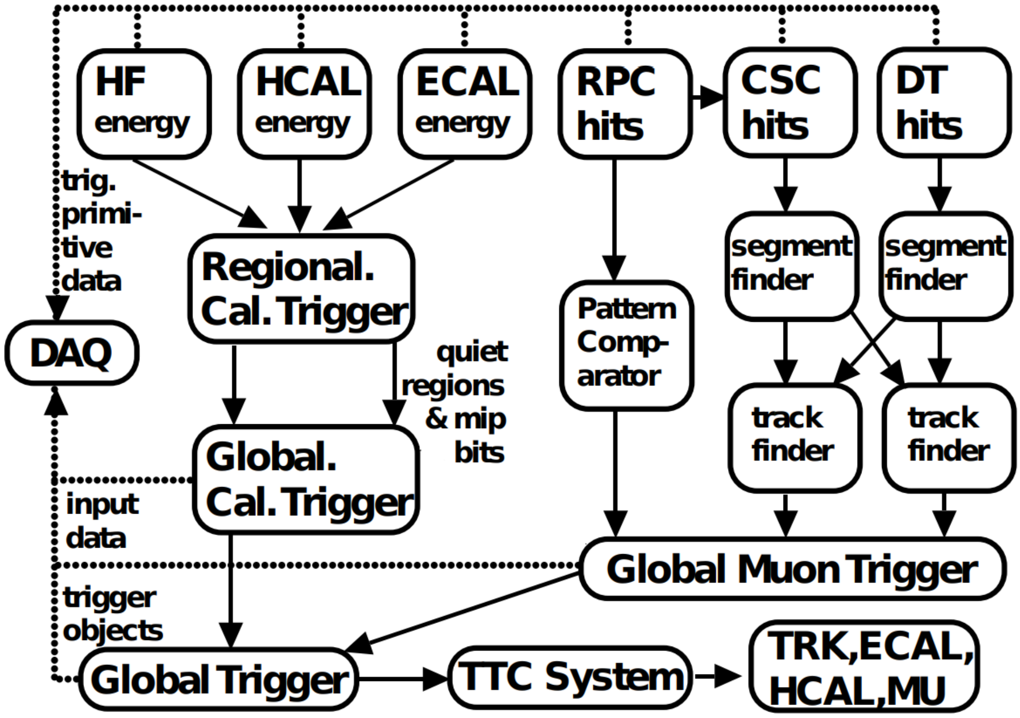


Figure 5.1: Overview of the CMS L1 trigger system, to be read from top to bottom. Data from the HCAL and ECAL calorimeters are processed first regionally and then globally. Hits from the RPC, CSC, and DT chambers are processed either via a pattern comparator or via a system of segment- and track-finders and then passed to a global muon trigger. The information from the calorimeter and muon triggers is combined in a global trigger, which makes the final trigger decision. The DAQ reads data from various sub-systems for offline storage. Figure from Ref. [119].

systems to construct "trigger primitive" objects consistent with muons, electrons, photons, jets, and global sums of transverse and missing transverse energy. An overview of the CMS L1 trigger system logic is shown in Fig. 5.1. Data from the calorimeters and muon detectors are processed first regionally and then in the global trigger, which makes the final trigger decision. Accepted events are passed on to the HLT, via the data acquisition system (DAQ).

The HLT is a software-based trigger system that runs on a processor farm consisting of around 22 000 CPU cores. The maximum processing time for an event in the HLT is constrained by the 100 kHz L1 rate to be around 220 ms. The trigger decision is based on a rough reconstruction of the full event that exploits also complex information from the tracking system. The particle candidates built by the HLT fast reconstruction algorithm are called "trigger objects". The data processing is structured in several so-called "HLT paths", each consisting of algorithms that run in parallel targeting different typologies. An event is immediately accepted for storage and further analysis if it is accepted by at least one path. On the other hand, events are rejected as soon as it is established that

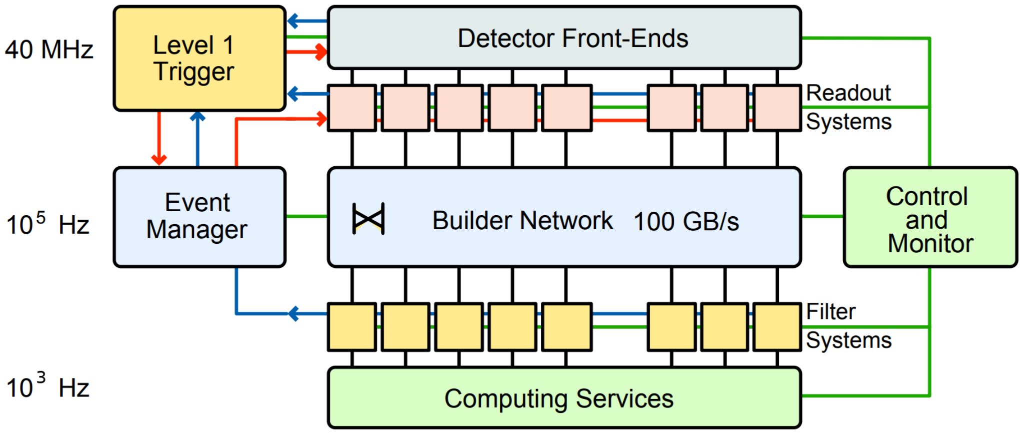


Figure 5.2: Schematic overview of the CMS DAQ and trigger systems. The trigger (left), data acquisition (center), and detector control systems (right) work in parallel to select and collect events. Figure from Ref. [96].

no path will accept them. The processing benefits from the sharing of objects, as an object only needs to be reconstructed once for use in several paths. During the LHC Run-2, the maximum sustainable HLT output rate was slightly above 1 kHz. The HLT used to collect the candidate events used in this analysis is described in Section 7.1.

The CMS DAQ system is closely integrated with the trigger system and is responsible for collecting and processing the data from all sub-systems [117, 118]. It is designed to read out a fraction of the sub-detectors at 40 MHz and the full CMS output at the L1 trigger rate of 100 kHz. Considering an event size of about 1–2 MB, the DAQ needs to be able to sustain an input rate of up to 200 GB/s. A schematic representation of the CMS DAQ system is presented in Fig. 5.2.

Events accepted by the trigger system are permanently stored for further analysis, which involves the full event reconstruction. A brief description of the physics object reconstruction of most interest for this work is presented in the next sections.

5.2 RECONSTRUCTION OF PHYSICS OBJECTS

The reconstruction of physics objects is the procedure in which the readout of all the CMS sub-detectors is combined to form hypotheses of stable particles in terms of momentum, energy, flight trajectory, and particle type. At CMS, this is performed with the so-called "Particle Flow" (PF) reconstruction algorithm [107], which relies on the high granularity of the CMS sub-detectors and allows signals from various sub-systems to be correlated. The PF algorithm is briefly described in the next sections, particular attention is given to the object of most interest for this work, that is interaction and decay vertices and muons[3].

5.2.1 *The Particle Flow algorithm*

The purpose of the PF algorithm is the reconstruction and identification of individual particles by combining measurements of all CMS sub-detectors in a global reconstruction of all physics objects in the event. The PF algorithm can be structured into three consecutive steps. First, information from the sub-system is collected to form so-called "PF objects". These comprise tracks of charged particles in the tracking system, clusters of energy deposited in the calorimeters, and tracks in the muon detectors. Next, the PF objects are spatially correlated with a *link algorithm*. Last, particle hypotheses are inferred and complex objects, such as vertices, jets, and missing energy, are computed.

TRACK AND CLUSTER FINDING Tracks are reconstructed iteratively starting from hits in the tracking system using the so-called "Combinatorial Track Finder" [120], an adaptation of the combinatorial Kalman filter [121–123]. Track candidates are extrapolated starting from seeds of pairs or triplets of hits in the outer layers of the tracker, with compatible inner hits assigned using the fit quality of the expected flight path. The iterative approach starts by reconstructing tracks with very tight criteria (e. g. large momentum), leading to a moderate tracking efficiency and a small fake rate. Next, the hits unambiguously assigned to the tracks found in the previous step are removed and new tracks are reconstructed with slightly looser criteria, increasing the efficiency. This procedure is repeated until the desired efficiency and fake rate is achieved.

The calorimeters energy deposits are clustered separately in the ECAL barrel and endcaps, in the HCAL barrel and endcaps, and in the two ECAL preshower detectors. Cells with deposited energy above a given threshold are used as clustering starting points, and neighboring cells are iteratively added.

Tracks in the muon system are obtained in two steps. First, hits in the DT and CSC sub-detectors are clustered in track segments. Next, these segments are used as seeds for pattern recognition techniques exploiting information from the full muon system readout. The resulting PF objects are called "standalone-muon track". More details on the reconstruction of muons are given in Section 5.3.

LINK ALGORITHM The link algorithm is the core of the PF reconstruction and connects the tracks and clusters reconstructed in the previous step by testing their compatibility with distance measurements. Although the link algorithm can test any group of objects, the number of pairs considered is restricted only to the nearest neighbors in the η-ϕ plane to reduce computing time.

3 Other physics objects scarcely used in this analysis, such as electron, photon, jets, and missing energy, are just briefly mentioned. For a full description of the PF algorithm see Ref. [107].

Tracks reconstructed in the tracker are linked to ECAL and HCAL barrel clusters in the η-ϕ plane, and ECAL endcap and preshower clusters in the x-y plane. This is achieved by extrapolating the tracks from their last measured hit in the tracking system to the cluster's area in the calorimeter. If several clusters are linked to the same track or vice versa, then only the link with the smallest distance in the η-ϕ plane is kept. Links between HCAL and ECAL clusters are established in a similar manner.

Finally, extrapolated tracks from the tracking system and segments in the muon detector are linked together to form muon track candidates as described in Section 5.3.

PARTICLE AND VERTEX RECONSTRUCTION The last step of the PF algorithm consists of forming particle hypotheses by using the PF objects reconstructed in the first step and the geometrical links evaluated in the second one. The procedure iteratively identifies physics objects in the following well-defined order, removing at each step the associated PF objects:

1. *Muons* are identified by a track in the tracking system linked to a track in the muon detectors. The expected energy deposited in calorimeter clusters is estimated and subtracted for further steps.

2. *Electrons* are identified by a track in conjunction with an ECAL cluster, and no HCAL deposits. To account for the emission of bremsstrahlung photons, energies of ECAL clusters compatible with the extrapolated tangent of the track are also considered in constructing the electron hypothesis.

3. *Charged hadrons* are identified by linking the remaining tracks to one or more calorimeter cluster. Geometrical ambiguities are resolved by comparing energy and momentum information. Track momenta that significantly exceed their linked calorimetric cluster and are related to weak activity in the muon system, are identified as low-quality muons. No attempt is made to distinguish between different species of hadrons[4].

4. *Neutral hadrons* and *photons* are identified by calorimeter clusters that are not associated with tracks in the tracker.

A general overview of the different interactions of these particle types has been already shown in Fig. 4.12 in Chapter 4.

The vertex reconstruction is performed entirely by the tracking system, which distinguished between *primary* interaction vertices (PV) and *secondary* decay vertices (SV). PVs result from pp interactions and are found within a cylinder of few millimeters around the beam axis, while SVs are related to the decays

4 Some species of charged hadrons are identifiable by exploiting the energy loss dE/dx of particles traversing the inner tracker [124]. However, this method is proven to be viable only for particles with momentum lower than 1 GeV. As such, this approach is not included in the PF algorithm and will not be further discussed.

of particles with relatively long lifetimes, such as b-hadrons. The vertex reconstruction procedures consist of three steps. First, reconstructed tracks are selected with quality criteria regarding their fit quality during the track finding procedure as well as their number of hits in the tracking system. Next, selected tracks are clustered using a deterministic annealing algorithm [125] based on their distance of closest approach to the interaction point (*beam spot*) in the longitudinal direction. This typically results in an arbitrary number of collision vertices per bunch crossing which depend on the instantaneous luminosity of the LHC and the inelastic cross section. Finally, the position of each vertex is determined using an adaptive fit procedure [126].

The collection of particle candidates reconstructed by the PF algorithm is then fed to higher-level object reconstruction algorithms in order to build higher-level physics objects such as jets, tau leptons, and missing transverse energy[5].

5.3 MUON RECONSTRUCTION

The PF algorithm reconstructs muons using information from both the tracking system and the muon spectrometer. The following three different types of muons can be reconstructed depending on which sub-detector information is used:

- *Standalone muons* are reconstructed by clustering hits in each DT or CSC chamber to form track segments, which are then used as seeds for the pattern recognition exploiting all DT, CSC, and RPC hits along the muon candidate trajectory. This type of muons typically has high purity, but also low precision in their momentum measurement. Since the tracking system is very efficient, standalone muons originating from collisions are very rare (≈1%) and most of them are cosmic ray muons traveling through the muon system[6].

- *Tracker muons* are reconstructed starting from inner tracks with $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV with at least one matched muon segment in the muon detectors. Multiple scattering processes with the detector material are taken into account in the track extrapolation to the muon system. The kinematic information of tracker muons is entirely estimated from the inner track, while the muon segment provides only the information on the particle type. Typically muons of this type have good purity, high reconstruction efficiency, and their momenta are measured with high precision[7].

5 As these objects are not used in this work, the respective reconstruction procedures are not described.

6 The muon system acceptance is $\approx 10^5$ times larger than the one of the silicon tracker.

7 Purity heavily depends on the kinematic region of interest. In the case of this analysis, where muons for flavor tagging are selected down to ≈2–3 GeV, tracker muons are significantly affected by misreconstructed light hadrons. See Chapter 8 and Appendix A, for more details.

- *Global muons* are build by extrapolating standalone-muon tracks inwards to the tracking system[8]. A link to an inner track is established if it yields a high compatibility as determined by a Kalman filter approach. The muon track is then refitted combining information from both sub-systems, thus improving the measurement of the muon momentum. This type of muons has high purity, while their reconstruction efficiency heavily depends on the muon momentum, reaching a plateau for $p_T > 8$ GeV.

For muons with momentum below 10 GeV the tracker muon reconstruction is more efficient than the global muon one, as muons are less likely to penetrate through more than one muon detector layer. Overall, about 99% of all muons produced within the geometrical acceptance of the muon system are reconstructed either as tracker muons or as global muons. Standalone muons are typically not considered. An example of a collision event in which four muons were reconstructed is shown in Fig. 5.3.

5.4 EVENT SIMULATION IN CMS

As in almost every physics analysis, this work makes a wide usage of simulated events to develop the various procedures and methods used to extract the physics results. In particular, simulations are used to parametrize acceptance and resolution effects, and to model background processes. The specific simulated samples used in this analysis are reported in Section 7.3.

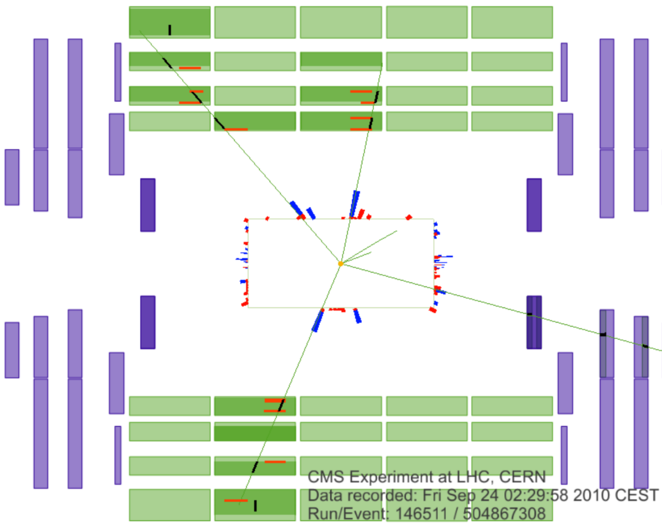
At CMS, the simulation of pp collision events is performed in multiple steps. The interaction of the two protons is view as a hard-scattering process of two constituent partons (i. e., gluons and quarks) using the so-called "factorization theorem", in which all possible parton combinations are factorized [128]. The cross section $\sigma(pp \to X \,|\, \sqrt{s})$ of a generic hard-scattering process $pp \to X$ at a given center-of-mass energy of $\sqrt{s}$, is computed as:

$$\sigma(pp \to X \,|\, \sqrt{s}) = \sum_{i,j=q,\bar{q},g} \int_0^1 \hat{\sigma}\left(ij \to X \,|\, \sqrt{\hat{s}}, \mu_R^2, \mu_F^2\right) f(x_1, \mu_F^2)\, f(x_2, \mu_F^2)\, dx_1\, dx_2 \;, \tag{5.1}$$

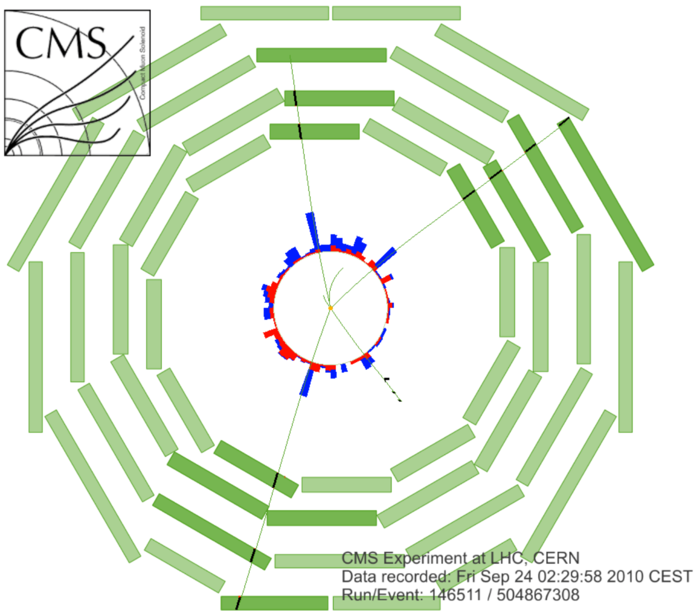
where the sum is performed over all possible parton combination (i, j), x_i is the fraction of the total hadron momentum carried by the parton i, $\hat{\sigma}$ is the parton-level cross section at the reduced center-of-mass-energy $\hat{s} = \sqrt{x_1 x_2 s}$, $f(x_i, \mu_F^2)$ is the parton distribution function of the parton i that carries a momentum fraction x_i at a given factorization scale μ_F[IX], and μ_R is the renormalization scale that

8 This approach is called *outside-in* and is opposite to the one used in reconstructing tracker muons (*inside-in*).

IX Interactions with $q^2 < \mu_F^2$ are absorbed in the parton distribution function, whereas those with $q^2 > \mu_F^2$ are contained in $\hat{\sigma}$.



(a) Longitudinal r-z view.



(b) Transverse r-ϕ view.

Figure 5.3: Longitudinal r-z (top) and transverse r-ϕ (bottom) views of a collision event in which four muons were reconstructed. The green curves in the inner cylinder represent tracks reconstructed in the inner tracker, those extending to the muon system represent tracks of muon reconstructed using hits in both the inner tracker and the muon system. Three muons were identified by the DTs and RPCs, the fourth one by the CSCs. Short black stubs in the muon system show fitted muon-track segments. Short red horizontal lines in the r-z view indicate positions of RPC hits. Energy depositions in the ECAL and HCAL are shown as red and blue bars, respectively. Figure from Ref. [127].

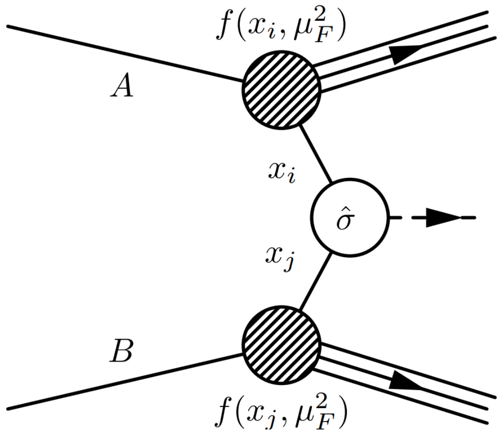


Figure 5.4: Schematic representation of the factorization in a hard-scattering process between two hadrons. Figure from Ref. [129].

defines the scale for the strong interaction coupling constant α_s[X]. A schematic representation of this factorization is shown in Fig. 5.4.

Parton distribution functions describe the probability of a parton to carry a certain momentum fraction of the initial hadron and depend on the type of the scattering parton and the total momentum q^2 transferred in the interaction. They are implemented in the simulation by Monte Carlo event generators which model each aspect of an event up to a given perturbation order [130]. Probabilistic processes, such as particle decays, are generated with set physics parameters, typically corresponding to the current world-average values, and with appropriately distributed random variables.

Above the QCD renormalization scale, i. e., when $q^2 > \mu_R^2$, the strong coupling constant is sufficiently small to allow QCD perturbative calculations. In these regimes, the parton showering approach is used to describe the evolution of QCD processes to a cutoff value by recursively adding $g \to q\overline{q}$ and $q \to qg$ splitting. At the renormalization scale, perturbation approaches cannot be used any longer and phenomenological methods are implemented to proceeds with the event description. The simulation continues until only stable particles remain, forming the final-state content of the event.

Additional stable particles are added to the final state as a result of the following residual contamination effects: multiple-parton scattering within the same pp interaction, interaction between particles produced in the hard process and remnants of the underlying event[11], and pileup interaction. At CMS, the effect of pileup is accounted for by overlaying the hard-scattering process to simulated minimum bias events, where the pileup multiplicity per event is based on the distribution observed in real data.

X Typically $\mu_F = \mu_R$ is chosen.

11 Underlying events are the remnants of scattering interactions, not directly related to the hard process.

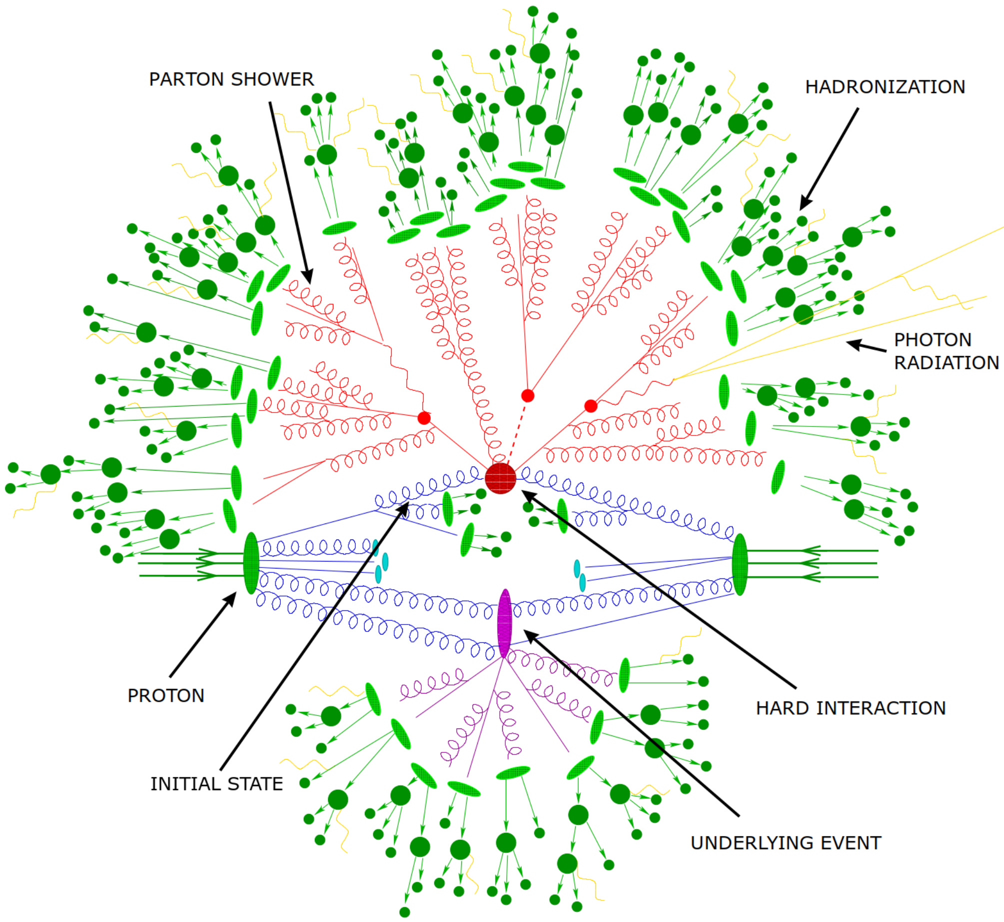


Figure 5.5: Sketch of a proton-proton collision from an event-generator point of view. Figure adapted from Ref. [131].

A schematic representation of a proton-proton collision from an event-generator point of view is shown in Fig. 5.5.

The response of the CMS detector and its sub-systems, given a set of simulated final-state particles, is simulated in the following steps. First, the final-state particles are propagated in a full simulation of the CMS detector using the Geant4 package [132], which models the propagation of particles through matter taking into consideration all types of interactions with matter, the production of secondary particles, and the influence of electromagnetic fields. Next, the readout of all sub-systems is simulated in the *digitization* step. At this point, the simulation consists of a collection of hits in the tracking and muon systems, and energy deposits in the calorimeters, resembling the detector response for a real collision. Finally, simulated samples are reconstructed using the same software as for collision data and can be analyzed with common analysis tools. Typically, the simulated samples contain also information on the simulated quantities, such as momenta, trajectories, decay time values, as well as ancestor histories, for the majority of the simulated particles of interest. This enables the development of analysis procedures over labeled data.

TECHNICAL DETAILS The simulated samples used in this work are produced using the PYTHIA 8.230 Monte Carlo event generator [133] with the underlying event tune CP5 [134] and the parton distribution function set NNPDF3.1 [135]. Final-state photon radiation is accounted for with PHOTOS 215.5 [136, 137]. The b-hadron decays are modeled with the EVTGEN 1.6.0 package [138].

Part III

DATA SAMPLE PREPARATION

The measurement of the CP-violating weak phase ϕ_s is based on a time-, angular- and flavor-dependent analysis of the $B_s^0 \rightarrow J/\psi\,\phi$ decay. In order to extract the physics parameters of interest, a large and pure data sample of candidate events must be collected and prepared. The probability of producing a $B_s^0 \rightarrow J/\psi\,\phi$ decay is very small. Approximately, only one bunch crossing every million produces a B_s^0 meson that decays into the desired final state and only a small fraction of B_s^0 mesons is actually contained in the CMS geometrical and kinematic acceptance. As such a multi-step selection procedure is needed to collect a large enough data sample with good purity. In this work, events are first selected by the CMS trigger system and later by offline optimized requirements. However, simply collecting the B_s^0 candidates is not enough, as several key properties of each decay need to be evaluated with high precision. The two fundamental quantities needed to measure the CP asymmetry are the B_s^0 decay time and flavor at the production time, each requiring dedicated tools, that need to be developed. The following chapters are dedicated to the description of the analysis strategy, the event selection procedure, and the flavor tagging algorithms. Once the data sample is defined and each event is characterized, the CP-violating weak phase ϕ_s can finally be measured using an unbinned maximum-likelihood fit.

6 ANALYSIS OVERVIEW

The goal of the analysis presented in this dissertation is to measure the properties of the $B_s^0 \to J/\psi\,\phi$ decay, when reconstructed in the final state where the J/ψ and ϕ mesons[1] decay to two muons and two charged kaons, respectively.

The differential decay rate for this process is described by Eq. 3.19, where the main physics parameters of interest are the CP-violating weak phase ϕ_s, the decay width (mass) splitting $\Delta\Gamma_s$ (Δm_s), the average decay width Γ_s, and the direct CPV observable $|\lambda|$.[II] Additionally, the transversity amplitude magnitudes $A_i = \left(|A_0|, |A_\perp|, |A_\parallel|\right)$, their respective strong phases $\delta_i = \left(\delta_0, \delta_\perp, \delta_\parallel\right)$, and the S-wave magnitude $|A_S|$ and strong phase δ_S are measured[3] in order to correctly model the data.

Equation 3.19 assumes the knowledge of the B_s^0 meson flavor at production time, which is usually not known and not inferable from the final state (as it is accessible from both B_s^0 and $\overline{B}_s^0$ mesons). In this analysis, the initial B_s^0 flavor is inferred on a probabilistic basis, by exploiting semileptonic $b \to \mu^- X$ decays of the other b-hadron in the event, under the assumption that b-quarks are produced in $b\overline{b}$ pairs. The principles and the development of this "flavor tagging" technique are described in detail in Chapter 8.

The ϕ_s measurement here presented is based on a data sample corresponding to an integrated luminosity of $96.4\,\mathrm{fb}^{-1}$, collected with the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2017–2018. Candidates are selected in two steps, first online by the CMS trigger system and then by offline selection requirements optimized to enhance the signal significance. The specific trigger used in this analysis requires a $J/\psi \to \mu^+\mu^-$ candidate plus an additional muon, which is used to infer the initial flavor of the B_s^0 meson, assuming that it is produced in the semileptonic decay of the other b-hadron in the event. A combination of rectangular cuts is then used to obtain the final sample of 65 500 $B_s^0 \to J/\psi\,\phi$ candidates. This selection procedure is described in detail in Chapter 7. A schematic representation of a generic $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ event used in this analysis is shown in Fig. 6.1.

The extraction of the physics parameters of interest is achieved through a maximum-likelihood fit to the data sample. The main fit inputs are the variables needed to parametrize the $B_s^0 \to J/\psi\,\phi$ decay rate, these are the B_s^0 proper

1 The ϕ mesons refers to the $\phi(1020)$ resonance.
II These parameters are defined in Equations 3.10, 2.29, 2.28, 2.27, and 2.41, respectively.
3 More specifically, $A_\parallel$ and δ_0 are not measured as the former is defined in terms of the other amplitudes as $|A_\parallel|^2 = 1 - |A_0|^2 - |A_\perp|^2$ and the latter is set to zero following a general convention.

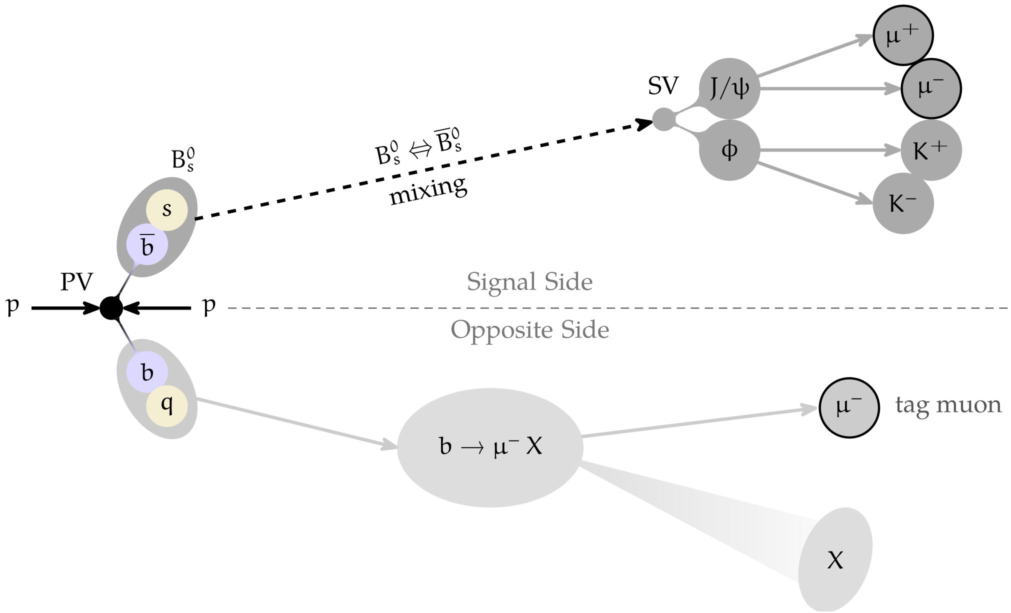


Figure 6.1: Schematic representation of a generic $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ event used for this analysis. The three triggered muons are indicated with a black border. The primary and secondary vertices are indicated with "PV" and "SV", respectively. The tag muon is used to infer the B_s^0 meson flavor at production time.

decay length[4], the three transversity angles, and the flavor tagging decision[5]. The following auxiliary variables are also used to improve the fit performances: the B_s^0 decay length uncertainty is included to better model the decay length[6], the mistag probability is included to enhance the tagging performance[7], finally the B_s^0 meson invariant mass is included to improve the signal/background discrimination. Time and angular efficiency functions are also included in the fit to take into account the distortions introduced by the detector acceptance and selection requirements. The fit model, efficiency parametrization, and fitting procedure are described in detail in Chapter 9.

Various systematic effects can affect the extraction of the physics parameters. These range from biases in the fitting model to limited statistics of the simulated samples. The evaluation of systematic uncertainties is discussed in Chapter 10.

Finally, the results and their combination with the previous CMS measurement at $\sqrt{s} = 8\,\text{TeV}$ are presented in Chapter 11.

4 The proper decay length is defined in as the proper decay time multiplied by the speed of light c and measured with Eq. 8.1.
5 The flavor tagging decision is defined in Eq. 8.1.
6 The main reason to include the decay length uncertainty is to avoid the so-called "Punzi bias" [139].
7 The reason for this improvement is explained in Section 8.1.1.

6.1 IMPROVEMENT WITH RESPECT TO PREVIOUS CMS RESULTS

The CMS Collaboration already measured the CP-violating phase ϕ_s in proton-proton collisions at $\sqrt{s} = 8$ TeV [64]. With respect to this previous result, this work presents a different strategy and several improvements. Due to the complexity of the measurement, it is impossible to mention all the changes and improvements. Some of the most relevant ones are:

- *Higher center of mass energy of the proton-proton collisions*, resulting in a higher production cross section for $b\overline{b}$ pairs (see Fig. 4.8).
- *New tagging focused trigger*, which requires an additional muon on top of the J/ψ candidate. The trigger used in the previous work required only a J/ψ candidate displaced from the luminous region.[8].
- *Physics parameters Δm_s and $|\lambda|$ free to float*. In the previous result Δm_s and $|\lambda|$ were constrained to the world-average and one, respectively. The constraint on $|\lambda|$ was, in particular, one of the leading sources of systematic uncertainty.
- *New method for evaluating the mistag probability*, which employs using a calibrated deep neural network.
- *Only muon tagging algorithm used*, in contrast with the previous result where both muon and electron tagging techniques were used[9]. This approach reflects the fact that the data sample used in this work is enriched of muons in the opposite side as a consequence of the trigger choice.
- *Dedicated term in the fit model to account for the peaking background from $B^0 \to J/\psi K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$*. In the previous result, only the combinatorial background was considered in the fit model.
- *Proper decay length efficiency included in the fit model*, while in the previous work it was not added as the efficiency was observed to be flat.
- *Additional sources of systematic uncertainties investigated*.

Among these improvements, the most important change is the trigger used to collect the data. By requiring an additional (third) muon at HLT level, the fraction of flavor-tagged events is greatly enlarged. In this work the tag decision is provided for $\approx$50% of the selected events, roughly six times higher than in the previous result ($\approx$8%). However, this comes at the cost of a reduced number of signal events selected at trigger level, as the $\mathcal{B}(b \to \mu) \sim 10\%$ inclusive branching ratio needs to be taken into account. As a result, this work is based on a similar number of B_s^0 candidates as the earlier one, even if the $pp \to b\overline{b}$

8 The trigger used in this work does *not* require a displaced J/ψ candidate.
9 Both methods are based on semileptonic $b \to \ell^- X$ decays of the other b-hadron in the event.

cross section has doubled and the integrated luminosity is five-time larger. This leads to an improvement in the uncertainty in quantities that are sensitive to flavor information, such as ϕ_s, while the uncertainties in those that do not use tagging (such as $\Delta\Gamma_s$) are not improved with respect to the 8 TeV result, as they depend mostly on the raw number of events.

7 CANDIDATE SELECTION

To perform a precise measurement of the CP-violating phase ϕ_s, a large and pure sample of $B^0_s \to J/\psi\,\phi$ decays is needed. However, only a small fraction of the proton-proton collisions provided by the LHC contains a B^0_s meson, with the majority of them being impossible to trigger. Moreover, only a small part of these B^0_s mesons decay into $J/\psi\,\phi$ inside the CMS detector geometrical acceptance.

The ensemble of techniques used to reconstruct and select a data sample for a given process is called "candidate selection". Its main goal is to maximize the *signal* statistics, while at the same time reducing the contamination from *background* sources. In doing so, several figures of merit can be defined, depending on the type of signal needed. A common choice used also in this work, is the "signal significance", defined as:

$$S = \frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}}, \tag{7.1}$$

where N_{sig} and N_{bkg} are the number of signal and background candidates, respectively.

This chapter describes the selection procedures for the definition of the data sample used in the measurement.

7.1 TRIGGER SELECTION

Events[1] are required to be collected in *good* experimental conditions, i. e., when the LHC and all the CMS subdetectors were in proper conditions to perform physics measurements. This certification is performed by the CMS Data Quality Monitoring and Certification group and applied to the data sets used in this analysis.

As explained in Section 5.1, only a fraction of pp collisions are selected by the CMS trigger system to be recorded for further analyses. In this work a high-level trigger optimized for the detection of b-hadrons decaying to J/ψ mesons, together with an additional muon potentially usable for flavor tagging, is used:

$$\texttt{HLT\_Dimuon0\_Jpsi3p5\_Muon2}. \tag{7.2}$$

This HLT is seeded by triple-muons L1 seeds that require at least three muons, with the minimum p_T requirement on the highest p_T and second-highest p_T muons of $p_T > 5$ and 3 GeV, respectively, and their invariant mass

1 The term "event" is used to indicate a recorded and reconstructed bunch crossing, see Chapter 5.

Table 7.1: L1 trigger selection requirements for $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ candidates. The highest p_T and second-highest p_T muons are indicated with μ_1 and μ_2, respectively.

Variable	Requirement
Min. number of muons	3
$p_T(\mu_1)$	> 5 GeV
$p_T(\mu_2)$	> 3 GeV
$m(\mu_1\mu_2)$	< 9 GeV

$m(\mu_1\mu_2) < 9$ GeV. There is no p_T requirement on the additional muons at L1. Momenta cuts are the most common selection criteria in high energy particle physics, as they reduce background from soft processes and restrict the kinematic acceptance to the region of interest. The L1 selection requirements are summarized in Tab. 7.1.

At HLT level, the three muons are required to be within the CMS muon system geometrical acceptance $|\eta| < 2.5$ to ensure good reconstruction properties. Two of these muons must have $p_T > 3.5$ GeV, be oppositely charged, form at least one J/ψ candidate with an invariant mass in the range 2.95–3.25 GeV, and have a probability to originate from a common vertex larger than 0.5%[II]. The vertex fit probability represents the probability that the observed χ^2 for a correct model exceeds the χ^2 value of the vertex fit. Low-quality vertex fits can be rejected with a requirement on this quantity. The third muon is required to have $p_T > 2$ GeV. The HLT algorithms do not use L1 information, as such, there is not a direct relationship between the L1 muon objects and the HLT ones. The HLT selection requirements are summarized in Tab. 7.2.

Events that pass the trigger selection are stored for further analyses. As explained in Section 5.1, HLT algorithms do not run on the same particle candidates used for the offline analysis, but "trigger objects", which are particle candidates built from a fast reconstruction algorithm. This algorithm trades precision in the reconstruction for speed in the execution of the code, to allow the trigger system to run fast enough to keep up with the stream of data. In order to avoid biases due to the trigger selection, all three trigger objects need to be matched to the respective muon candidates used in the analysis. In practice, the trigger matching consists of requiring the trigger object to lie in a cone centered on the direction of the muon candidate p_T. The size of the cone in the η-ϕ plane is equal to $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.1$.

II All combinations of the three muons are tried.

Table 7.2: `HLT_Dimuon0_Jpsi3p5_Muon2` selection requirements for $B^0_s \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ candidates. The J/ψ candidate is formed by a pair of oppositely charged muons μ^+ and μ^- that pass the listed requirements. The muon not used to form the J/ψ candidate is indicated with μ_3. The three muons are not required to have a direct relationships with the ones in Tab. 7.1.

Variable	Requirement
$\lvert\eta(\mu)\rvert$	< 2.5
$p_T(\mu^\pm)$	> 3.5 GeV
$p_T(\mu_3)$	> 2 GeV
$m(\mu^+\mu^-)$	$\in [2.95, 3.25]$ GeV
J/ψ vertex prob.	$> 0.5\%$

7.2 OFFLINE RECONSTRUCTION AND SELECTION

Events that pass the HLT selection are further analyzed offline. The offline analysis is performed on the full event reconstruction and it allows more optimization flexibility. The offline analysis is performed with a combination of several software packages, the main ones being: the CMS Offline Software (CMSSW) [140], ROOT [141], TMVA [142], and RooFit [143]. CMSSW and ROOT form the general analysis framework, while TMVA and RooFit are used for multivariate analysis (MVA) and fitting, respectively.

7.2.1 *Selection requirements optimization*

The numerical values of the selection requirements (hereafter "cuts") used in the reconstruction have been optimized with the help of a Genetic Algorithm (GA), available as part of the TMVA package. The GA is a random-based classical evolutionary algorithm that aims at optimizing a set of properties called "genomes", the cut values in this case, by maximizing a given *fitness function*. In this application, the signal significance, defined in Eq. 7.1, is used as the fitness function. In analogy with evolutionary biology, a set of cut values is called "individual", while the ensemble of individuals at a given iteration of the GA is called "population". The GA is an iterative process and proceeds as follows:

1. *Initialization:* the algorithm starts by assigning a set of initial cut values defined by the user, then the values are changed randomly within a predefined n-dimensional matrix of allowed ranges[3], to generate the individuals of the starting population.

3 Also set by the user.

2. *Evaluation and selection:* at each iteration each individual is evaluated using the fitness function and the worst-performing ones are discarded (in this application the sets of cut values with the lowest signal significance).

3. *Reproduction:* the surviving individuals are copied, crossed-over and mutated to create another population. The cross-over procedure combines several individuals, while mutations change the value of some cuts of some individuals randomly following a Gaussian distribution.

The evaluation, selection, and reproduction steps are repeated until the fitness function is maximized and the best configuration of selection cuts is achieved. The following selection variables are optimized: transverse momenta and pseudorapidities of all the final-state particles, B_s^0 meson transverse momentum, χ^2 probability of the B_s^0 vertex fit, and B_s^0 proper decay length. In defining the allowed range values, particular attention is paid to at least reproduce the hardness of the selection at HLT, as not to introduce threshold distortions. The best configuration of offline selection requirements is described in the text and reported in Tab. 7.3.

7.2.2 *Candidate reconstruction*

The reconstruction of B_s^0 candidates is obtained by combining $J/\psi \to \mu^+\mu^-$ and $\phi \to K^+K^-$ candidates.

First, J/ψ mesons are reconstructed using a pair of opposite-charge muons[4], with $p_T > 3.5$ GeV and $|\eta| < 2.4$, if compatible with originating from a common vertex[5]. Muons are required to be reconstructed as *global muons*, as described in Section 5.3. The J/ψ candidates are selected if their invariant mass is within 150 MeV of the world-average value[6].

Then, ϕ meson candidates are constructed from pairs of tracks with opposite charge, with $p_T > 1.2$ GeV and $|\eta| < 2.5$. Tracks must satisfy the *high-purity* requirement described in Ref. [120], and not being associated with the muons forming the J/ψ candidate; their invariant mass must be within ± 10 MeV of the world-average $\phi(1020)$ meson mass[7], assuming the charged kaon mass for both particles[8].

Finally, B_s^0 meson candidates are constructed by combining the selected J/ψ and ϕ. The "B_s^0 vertex", also called "decay vertex" or "secondary vertex" (SV), is obtained by fitting to a common vertex the two muon and two kaon tracks. The B_s^0 invariant mass is then computed with a kinematic fit, where the invariant

4 As a consequence of the trigger matching procedures, these two muon candidates are the same two selected at HLT.
5 The vertexing procedures are described in Section 5.2.1.
6 $m_{J/\psi}^{\text{w.a.}} = 3096.900 \pm 0.006$ MeV [34].
7 $m_{\phi(1020)}^{\text{w.a.}} = 1019.461 \pm 0.016$ MeV [34].
8 $m_{K^\pm}^{\text{w.a.}} = 493.677 \pm 0.016$ MeV [34].

Table 7.3: Final selection requirements for $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ candidates. The selection on the variables indicated with a dagger symbol (†) have been optimized with the GA, as described in Section 7.2.1.

Candidate	Variable	Requirement
Event	HLT_Dimuon0_Jpsi3p5_Muon2	✓
	PV found	✓
	Good conditions	✓
$B_s^0 \to J/\psi\,\phi$	Trigger match	✓
	$p_T(B_s^0)^\dagger$	> 11 GeV
	$ct(B_s^0)^\dagger$	> 70 µm
	$\sigma_{ct}(B_s^0)$	< 50 µm
	B_s^0 vertex prob.†	$> 2\%$
	$m(\mu^+\mu^-K^+K^-)$	$\in [5.24, 5.49]$ GeV
$J/\psi \to \mu^+\mu^-$	Muon reconstruction	*Global*
	$p_T(\mu)^\dagger$	> 3.5 GeV
	$\lvert\eta(\mu)\rvert^\dagger$	< 2.4
	$\left\lvert m(\mu^+\mu^-) - m_{J/\psi}^{\text{w.a.}}\right\rvert$	< 150 MeV
$\phi \to K^+K^-$	Track quality	*High-purity*
	$p_T(K)^\dagger$	> 1.2 GeV
	$\lvert\eta(K)\rvert^\dagger$	< 2.5
	$\left\lvert m(K^+K^-) - m_{\phi(1020)}^{\text{w.a.}}\right\rvert$	< 10 MeV

mass of the J/ψ candidate is constrained to the world-average value, while the mass of the ϕ candidate is left free to float, as its natural width is larger than the mass resolution[9]. B_s^0 candidates are selected if the invariant mass is in the 5.24–5.49 GeV range and the four-track vertex fit probability is higher than 2%.

PRIMARY VERTEX SELECTION The reconstruction and identification of the primary interaction vertex (PV) are of foremost importance for this analysis, as it is needed to measure the proper decay length. Due to the high instantaneous luminosity of proton-proton collisions delivered by the LHC at the CMS interaction point, several primary vertices, corresponding to different collisions taking place in the same bunch crossing, are reconstructed in each event. The reconstruction of the primary vertices per event is described in Section 5.2.1. The distribution of the number of primary vertices is shown in Fig. 7.1. Several approaches can be used to identify the primary vertex which generated the B_s^0 meson. In this work, the vertex that minimizes the so-called "pointing angle" is chosen. The pointing angle is defined as the angle between the B_s^0 momentum direction and the line connecting a given PV to the B_s^0 decay vertex. Another popular approach is selecting the vertex with the highest transverse momenta, calculated as the sum of the p_T of each track linked to the vertex. However, the pointing angle method is proven to be a more precise method for this analysis. Simulations are used to study if the PV selection procedure introduces any bias in the measurement: in about 96.5% of the events, the selected PV is also the closest one to the point of origin of the B_s^0 meson. The impact on the final results of choosing a different vertex in the remaining cases is estimated by fitting the simulated samples using always the *closest* PV and comparing the results with the ones obtained from the reference fit. No relevant bias is observed, hence no systematic uncertainty is assigned.

PROPER DECAY LENGTH MEASUREMENT The proper decay length ("ct") is measured in the transverse plane, using the selected PV, as:

$$ct = c\,\frac{m_{B_s^0}^{\text{w.a.}} \cdot L_{xy}}{p_T}\,, \tag{7.3}$$

where c is the speed of light, $m_{B_s^0}^{\text{w.a.}}$ is the world-average B_s^0 mass[10] and L_{xy} is the reconstructed transverse decay length:

$$L_{xy} \equiv \|\vec{r}_{xy}(\text{SV}) - \vec{r}_{xy}(\text{PV})\|\,, \tag{7.4}$$

with $\vec{r}_{xy}(\text{SV})$ and $\vec{r}_{xy}(\text{PV})$ being the secondary and primary vertex positions in the transverse plane, respectively. The proper decay length is measured in the

9 $\Gamma_{\phi(1020)}^{\text{w.a.}} = 4.249 \pm 0.013$ MeV [34].
10 $m_{B_s^0}^{\text{w.a.}} = 5366.88 \pm 0.14$ MeV [34].

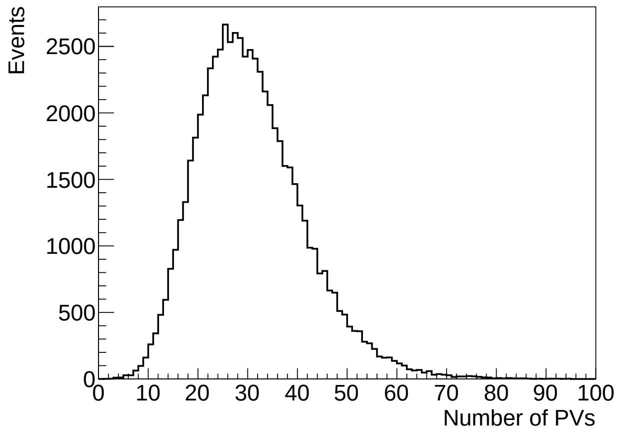


Figure 7.1: Distribution of the number of primary vertices reconstructed in the selected events. The average peak instantaneous luminosity is ≈ 15–20 Hz/nb depending on the data-taking period [93].

transverse plane as the vertex position resolution in the x-y plane is significantly higher than the one in the z direction.

The proper decay length uncertainty is obtained by propagating the uncertainties in the decay length and the p_T of the B_s^0 candidate, as:

$$\sigma_{ct} = ct\sqrt{\left(\frac{\sigma_{L_{xy}}}{L_{xy}}\right)^2 + \left(\frac{\sigma_{p_T}}{p_T}\right)^2}\,. \tag{7.5}$$

where the first term is by far dominating. The typical values of the quantities in Eq. 7.5 are $L_{xy} \sim \mathcal{O}(\text{mm})$, $\sigma_{L_{xy}} \sim 80\ \mu\text{m}$ and $\sigma_{p_T}/p_T \sim 1\%$, with the uncertainty in measuring the transverse decay length in turn is dominated by the spatial resolution on the position of the secondary vertex. The primary vertex is determined with significantly better precision as it is typically reconstructed using more tracks.

DECAY LENGTH SELECTION REQUIREMENTS A selection cut of $ct > 70\ \mu\text{m}$ for B_s^0 candidates is required in order to reduce the *prompt* background[11]. This type of background is related to processes produced in proximity of the primary vertex and with negligible lifetime. These events can pollute the data sample when the B_s^0 candidate is constructed by combining a prompt J/ψ meson with two random charged tracks with an invariant mass in the ϕ mass region. Since the amount of prompt J/ψ mesons is orders of magnitude higher than the ones produced in $B_s^0 \to J/\psi\,\phi$ decays, a non-negligible amount of these events can survive all the other selection requirements. A requirement on the minimum decay length is the easiest way to reduce such a background since genuine B_s^0

11 The requirement on ct is optimized with the procedure described in Section 7.2.1.

mesons have a relatively long average lifetime of $\tau_{B_s^0} \approx 1.5$ ps, which makes them travel $\mathcal{O}$(mm) before decaying. With respect to the previous CMS result [64], here the ct cut is used mainly to reduce background and not to reproduce the HLT selection, since in this work the HLT does not include a displacement requirement on the J/ψ candidate. The proper decay length distribution, without the $ct > 70$ µm cut, is shown in Fig. 7.2. Figure 7.3 show the invariant mass distribution of B_s^0 candidates before and after the ct cut, with all other selection requirements applied[12].

Additionally, the ct uncertainty is required to be lower than 50 µm, to exclude candidates with high uncertainty on the vertex position. Its distribution is shown in Fig. 7.4.

MULTIPLE CANDIDATES In about 2% of the events more than one candidate is selected. In these cases, the candidate with the highest vertex fit probability is chosen. The impact of this criterion on the measurement has been evaluated by re-performing the analysis taking the candidate with the *lowest* vertex fit probability. No relevant bias is observed, hence no systematic uncertainty is assigned.

DATA SAMPLES A total of 65 470 $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^- K^+K^-$ candidates are selected, divided into 25 220 and 40 250 candidates for the 2017 and 2018 data-taking periods, respectively. These data samples can be further split into sub-periods, according to the corresponding technical conditions of the CMS detector and the LHC machine. The integrated luminosity and selected number of events for each data-taking period are reported in Tab. 7.4. The later part of the 2017 data, corresponding to the data-taking period 2017-F, shows a slightly lower event rate with respect to the other periods. This is a consequence of an observed drop in tracking efficiency due to inactive pixel detector modules caused by a powering issue related to radiation-induced damage in the DC-DC converter ASICs [144]. This issue was promptly investigated and mitigated in subsequent data-taking periods.

The J/ψ and ϕ mesons invariant mass distributions for the selected candidates are shown in Fig. 7.5, where the latter is observed to be asymmetric as expected from a mass spectrum described by a relativistic Breit–Wigner distribution with spin 1. A CMS event display of a candidate event used in this analysis is also shown in Fig. 7.6. It is worth noticing that the invariant mass distribution of J/ψ candidates is practically background-free, which is not the case for B_s^0 and ϕ candidates. This observation suggests that the majority of J/ψ candidates are genuine and most of the $B_s^0 \to J/\psi\,\phi$ background comes from either combination of prompt J/ψ and two random charged tracks or $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ decays where the pion is assumed to be a kaon candidate.

12 The cut on ct is optimized simultaneously with the others, as explained Section 7.2.1. Figure 7.2 is just for visualization purpose.

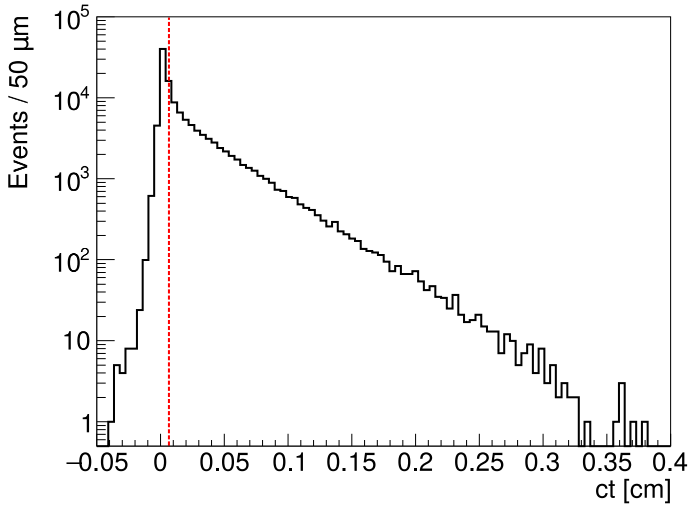
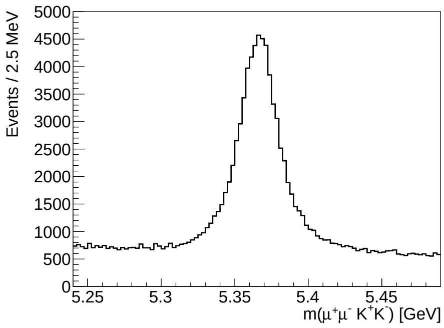
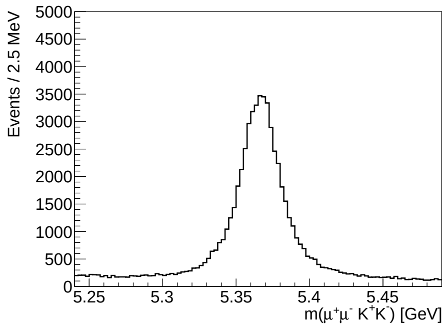


Figure 7.2: $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ proper decay length distribution in the selected data sample before the $ct > 70$ μm selection requirement. All other selection cuts are applied. The ct values can take negative values due to the uncertainty in its measurement. The red dashed vertical line represent the $ct > 70$ μm selection requirement.



(a) Mass distribution before the ct cut.



(b) Mass distribution after the ct cut.

Figure 7.3: $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ invariant mass distribution in the selected data sample before (left) and after (right) the $ct > 70$ μm selection requirement. All other selection cuts are applied.

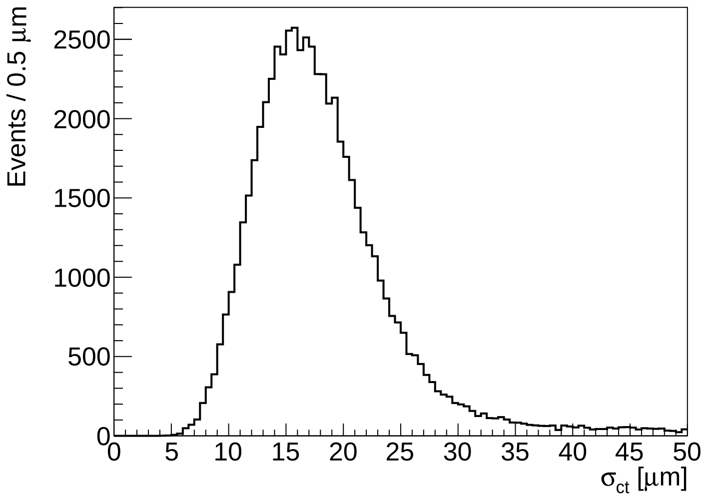
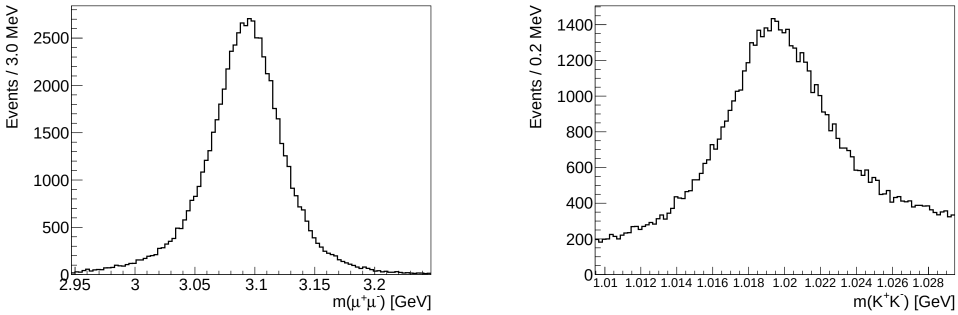


Figure 7.4: Proper decay length uncertainty distribution in the selected $B_s^0 \to J/\psi\,\phi$ data sample.

Table 7.4: Data sets used to perform the ϕ_s measurement. For each data-taking period the integrated luminosity and the number of selected candidates are also reported. The number of candidates is rounded to the nearest ten.

Data set	$\mathcal{L}_{\text{int}}$ [fb^{-1}]	Number of candidates
2017-C	9.6	7 100
2017-D	4.3	3 100
2017-E	9.3	6 890
2017-F	14.5	8 130
Total 2017	36.7	25 220
2018-A	14.0	9 710
2018-B	7.1	4 880
2018-C	6.9	4 510
2018-D	31.7	21 150
Total 2018	59.7	40 250
Total	96.4	65 470



(a) Invariant mass distribution of the selected $J/\psi \to \mu^+\mu^-$ candidates.

(b) Invariant mass distribution of the selected $\phi \to K^+K^-$ candidates.

Figure 7.5: Invariant mass distribution, after all the selection requirements, for the J/ψ (left) and ϕ (right) candidates. The x-axis ranges for each distribution correspond to the selected mass windows.

7.3 SIMULATED DATA SAMPLES

Simulated events are used to estimate the time and angular efficiency, a core component of the fit model[13], and to train the various MVA methods used in the flavor tagging algorithms[14]. Three different types of simulated $B_s^0 \to J/\psi\,\phi$ samples are produced:

- *Regular* $B_s^0 \to J/\psi\,\phi$ decays, which are used in all steps of the analysis, from the optimization of the tagging tools to the development of the fit model. These samples have been generated with values of physics parameters close to the world-averages ones. Hereafter, these samples will be indicated only as "simulated $B_s^0 \to J/\psi\,\phi$ samples", without further specification.

- $B_s^0 \to J/\psi\,\phi$ decays, where $\Delta\Gamma_s$ is forced to be equal to $0\ \text{ps}^{-1}$, which are used to measure the time and angular efficiency. The $\Delta\Gamma_s = 0$ condition is equivalent to setting the same mean lifetime for all the CP components of the final state, allowing the decoupling of the time and angular components of the efficiency[15]. These samples are referred to as "$B_s^0 \to J/\psi\,\phi\ (\Delta\Gamma_s = 0)$" hereafter.

- $B_s^0 \to J/\psi\,\phi$ decays where no selection filters are applied at generation level and no reconstruction is performed. This sample is also used in the efficiency measurement. The absence of generation filters, such as

13 See Section 9.2.3 and 9.2.4.
14 See Chapter 8.
15 If $\Delta\Gamma_s \neq 0$, different CP eigenstates would have different mean lifetime, leading to different time evolution of the angular components. This would make impossible to measure the time and angular efficiency separately.

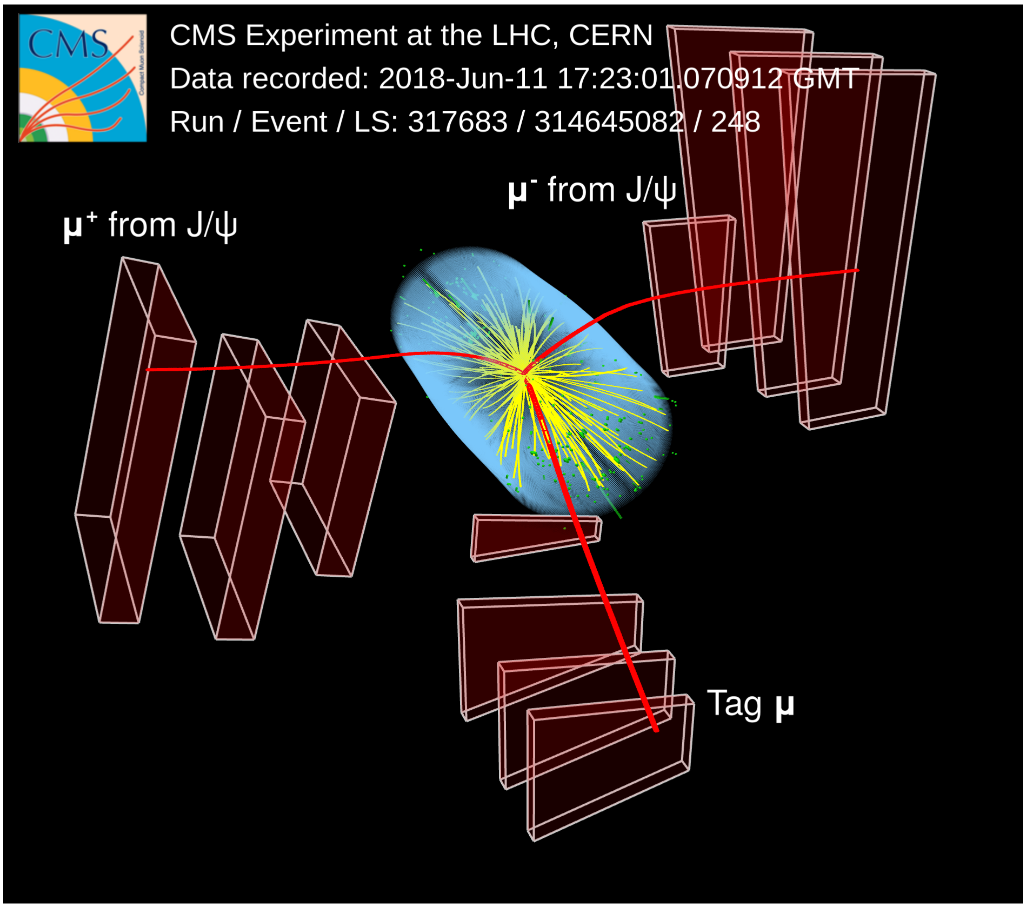


Figure 7.6: Event display of a $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ candidate event used in the CPV analysis. The red lines represent muons candidates reconstructed in the muon system, while the yellow lines represent charged tracks reconstructed in the tracking system. The red boxes represent hits in the muon DT chambers. Figure from Ref. [145].

asking that the final-state particles are contained in the CMS geometrical acceptance, is essential to avoid biases. This sample is generated with the same physics parameters as the $B_s^0 \to J/\psi\,\phi$ $(\Delta\Gamma_s = 0)$ sample. All other samples are produced requiring the final-state particles inside the CMS geometrical acceptance.

The following additional samples of other processes are also simulated:

- $B^0 \to J/\psi\,K^*(892)^0$ and $\Lambda_b^0 \to J/\psi\,K^- p$ samples, which are used to study and model background sources,
- $B^+ \to J/\psi\,K^+$ samples, which are used in the development of the flavor tagging algorithms,
- $J/\psi \to \mu^+\mu^-$ samples, which are used to study the proper decay length resolution.

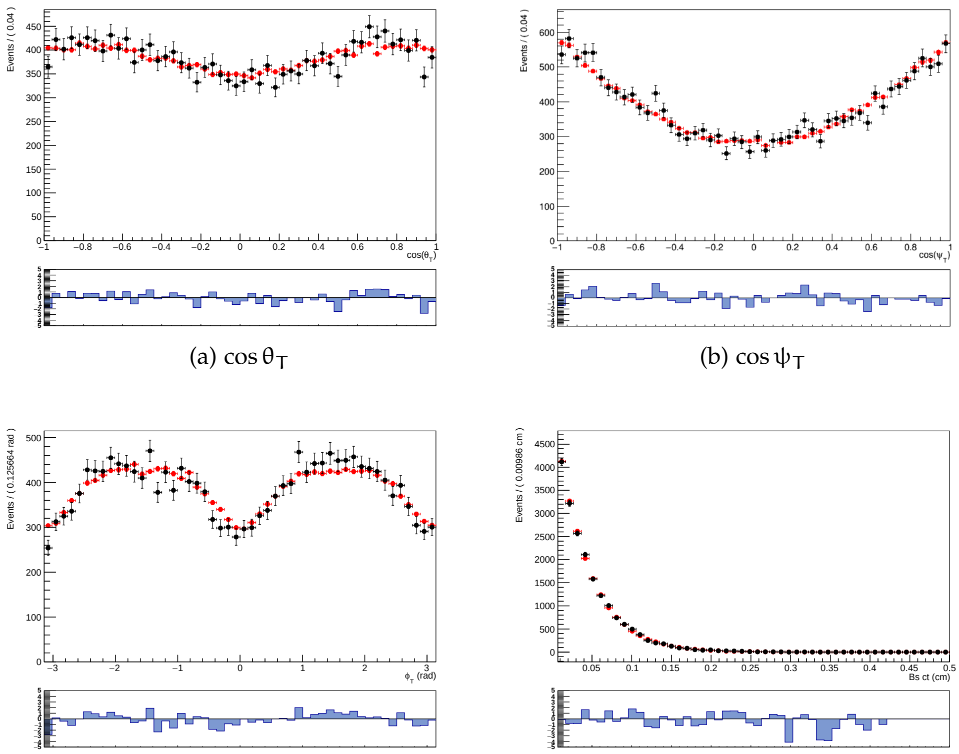
The simulated processes and the number of generated events are summarized in Tab. 7.5. All MC samples are produced with the procedure described in Section 5.4.

In this work, the most critical usage of simulated samples is the parametrization of the time and angular efficiency functions. Figure 7.7 and 7.8 show the comparison between distributions of the proper decay length and the three transversity angles in $B_s^0 \to J/\psi\,\phi$ data samples selected in data and MC. In the data sample, the background is subtracted with the *sPlot* technique [146], using the invariant mass as discrimination variable. This technique allows reconstructing the distributions of a given variable, independently for signal and background, without making use of any *a priori* knowledge, by exploiting the information available for the discriminating variable and using it to assign a weight to each event.

An overall good agreement is observed for all variables. The comparison is also repeated with similar results using the *sideband background subtraction* technique. Residual differences will be treated as a source of systematic uncertainty in Chapter 10.

Table 7.5: Simulated processes and number of generated events for each year of data-taking conditions. The number of events simulated with the 2017 and 2018 data-taking conditions is indicated with N_{evt}^{2017} and N_{evt}^{2018}, respectively. Numbers are rounded to the closest million. Charge conjugation is implied.

<table>
<tr><th>Simulated process</th><th>$N_{evt}^{2017} \cdot 10^6$</th><th>$N_{evt}^{2018} \cdot 10^6$</th></tr>
<tr><td>$B_s^0 \to J/\psi\,\phi$</td><td>76</td><td>73</td></tr>
<tr><td>$B_s^0 \to J/\psi\,\phi\ (\Delta\Gamma_s = 0)$</td><td>66</td><td>107</td></tr>
<tr><td>$B_s^0 \to J/\psi\,\phi$ (no filters)</td><td colspan="2">78</td></tr>
<tr><td>$B^+ \to J/\psi\,K^+$</td><td>24</td><td>24</td></tr>
<tr><td>$B^0 \to J/\psi\,K^*(892)^0$</td><td>67</td><td>55</td></tr>
<tr><td>$\Lambda_b^0 \to J/\psi\,K^- p$</td><td>/</td><td>74</td></tr>
<tr><td>$J/\psi \to \mu^+\mu^-$</td><td>29</td><td>250</td></tr>
</table>



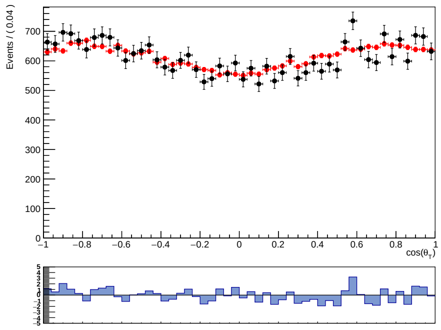
(a) $\cos\theta_T$

(b) $\cos\psi_T$

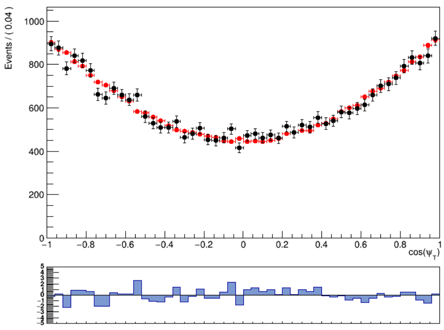
(c) φ_T

(d) $ct(B_s^0)$

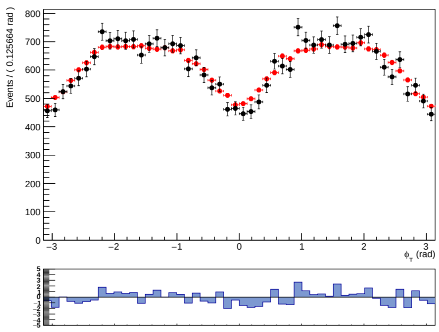
Figure 7.7: Comparison between the distributions of selected $B_s^0 \to J/\psi\,\phi$ decays in 2017 data (black) and simulated $B_s^0 \to J/\psi\,\phi$ events (red). In the data sample, the background is subtracted with the *sPlot* technique, using the invariant mass as discrimination variable. The pull distributions between the two distributions in each bin are shown in the lower panels.



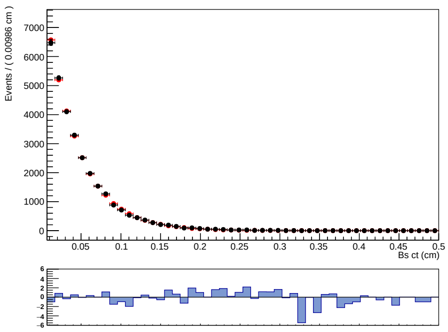
(a) $\cos\theta_T$



(b) $\cos\psi_T$



(c) φ_T



(d) $ct(B_s^0)$

Figure 7.8: Comparison between the distributions of selected $B_s^0 \to J/\psi\,\phi$ decays in 2018 data (black) and simulated $B_s^0 \to J/\psi\,\phi$ events (red). In the data sample, the background is subtracted with the *sPlot* technique, using the invariant mass as discrimination variable. The pull distributions between the two distributions in each bin are shown in the lower panels.

8 FLAVOR TAGGING

The determination of the flavor at production time of the B^0_s meson[1] is an essential ingredient to measure the CP violation in $B^0_s \to J/\psi\,\phi$ decays, since the terms more sensitive to ϕ_s in the decay rate model (Eq. 3.14) require the flavor information[2]. In practice, if no flavor information is available, there is no sensitivity to the last two terms of Eq. 3.14.

As the final state is a CP-eigenstate, accessible to both B^0_s and $\overline{B}^0_s$, the initial B^0_s flavor is not deducible by studying the final-state particles. The ensemble of techniques used to infer the flavor of a given particle is called "flavor tagging"[3]. Deducing the initial flavor is trivial in decays of charged mesons, e. g. in $B^+ \to J/\psi\,K^+$ decays where the flavor of the B^+ is given by the charge of the kaon, but way more complicated for neutral ones. Neutral b-mesons can still be tagged in some decay channels exploiting the final state, for example, the flavor of the B^0 meson in $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^-\,K^+\pi^-$ decays can be inferred from the charges of the final-state hadrons. However, in the case of $B^0_s \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ the final-state particles contain no information at all on the initial B^0_s flavor and indirect techniques must be adopted.

In this work, the B^0_s flavor at production time is determined with an *opposite-side* muon tagger, a popular technique that applies very well to the muon-enriched data sample. The tagging algorithm is based on machine learning techniques exploiting several properties of deep neural networks (DNN) to improve tagging performance by predicting the mistag probability on a per-event basis.

This chapter presents the fundamentals of flavor tagging in hadron colliders and describes the development of the opposite-side muon tagging algorithm. A brief introduction to DNNs is also given.

8.1 FLAVOR ALGORITHMS FUNDAMENTALS AND DEFINITIONS

Flavor tagging techniques (also called "taggers") work on a probabilistic basis and needs to be carefully optimized to ensure high performance. To facilitate the description of tagging techniques, several conventions and definitions are used in this chapter. The following discussion applies to all types of taggers.

1 Charge conjugation are implied thorough the chapter unless stated otherwise.
2 These terms change sign between decays of initial B^0_s and $\overline{B}^0_s$ mesons.
3 Not to be confused with "b-tagging", which are techniques used to identify b-jets.

The flavor tagging decision is assigned event by event and it is defined as:

$$\xi_{tag} = \begin{cases} +1 & \text{for } B\,, \\ -1 & \text{for } \overline{B}\,, \\ 0 & \text{if no tagging decision is made.} \end{cases} \tag{8.1}$$

Events are therefore first classified as "tagged" or "untagged" (UT), depending on the outcome of the tagging decision. A further classification can be made on whether the tagging decision matches the true flavor of the *signal* b-meson. If the inferred flavor is equal to the true flavor, events are classified as "rightly tagged" (RT), otherwise, they are referred to as "wrongly tagged" (WT) or "mistagged".

These categories can be used to construct several figures of merit, which help developing tagging algorithms. The most common ones are the *tagging efficiency*, the *mistag rate*[4] and the *tagging power*.

The *tagging efficiency* is defined as the fraction of tagged events ($\xi_{tag} = \pm 1$), regardless of the correctness of the inference, as:

$$\varepsilon_{tag} = \frac{N_{tag}}{N_{tot}} \tag{8.2}$$

where N_{tag} and N_{tot} are the numbers of tagged and total events, respectively. These quantities can also be defined as a function of the tagging decision as:

$$N_{tag} = N_{RT} + N_{WT}\,, \tag{8.3}$$

$$N_{tot} = N_{RT} + N_{WT} + N_{UT}\,. \tag{8.4}$$

Practically, one wants to maximize the tagging efficiency to have access, for as many events as possible, to the flavor sensitive terms of the $B_s^0 \to J/\psi\,\phi$ decay rate model, however, the tagging decision is not foolproof, as the inference is made only on probabilistic terms. The fraction of tagged events for which the tagging decision is wrong is called *mistag rate* and it is defined as:

$$\omega_{tag} = \frac{N_{WT}}{N_{tag}} = \frac{N_{WT}}{N_{RT} + N_{WT}}\,. \tag{8.5}$$

The higher the mistag rate, the larger is the performance loss.

The tagging efficiency can be rescaled into the so called *tagging power* in order to take into account the fraction of mistagged events, as:

$$P_{tag} = \varepsilon_{tag}\left(1 - 2\omega_{tag}\right)^2 = \varepsilon_{tag}\mathcal{D}_{tag}^2\,, \tag{8.6}$$

where $\mathcal{D}_{tag}$ is the *tagging dilution*, a measurement of the performance degradation due to mistagged events. Equations 8.5 and 8.6 define quantities averaged in a

4 Also called *mistag fraction*.

given data set. The tagging dilution and the tagging power can also be evaluated on a per-event basis as described later in Section 8.1.1.

The tagging power is the most used figure of merit in optimizing tagging tools, as it directly measures the statistical power of the tagged sample. It is constructed starting from the true and observed tagging asymmetries, which are respectively defined as:

$$\widetilde{A}_{tag} = \frac{\widetilde{N}_B - \widetilde{N}_{\overline{B}}}{\widetilde{N}_B + \widetilde{N}_{\overline{B}}} \quad \text{and} \quad A_{tag} = \frac{N_B - N_{\overline{B}}}{N_B + N_{\overline{B}}}, \tag{8.7}$$

where the "tilde" (~) indicates true quantities in contrast with measured ones. The observed number of B and $\overline{B}$ events can be expressed as:

$$\begin{aligned} N_B &= \varepsilon_B \widetilde{N}_B (1-\omega_B) + \varepsilon_{\overline{B}} \widetilde{N}_{\overline{B}} \omega_{\overline{B}}, \\ N_{\overline{B}} &= \varepsilon_{\overline{B}} \widetilde{N}_{\overline{B}} (1-\omega_{\overline{B}}) + \varepsilon_B \widetilde{N}_B \omega_B, \end{aligned} \tag{8.8}$$

where ε_B ($\varepsilon_{\overline{B}}$) and ω_B ($\omega_{\overline{B}}$) are the tagging efficiency and the mistag fraction for in the case of a genuine B ($\overline{B}$), respectively. In the typical case where the tagging performance does not depend on the initial flavor, that is if $\varepsilon_B = \varepsilon_{\overline{B}} = \varepsilon_{tag}$ and $\omega_B = \omega_{\overline{B}} = \omega_{tag}$, the observed asymmetry becomes:

$$A_{tag} = \frac{\varepsilon_{tag} \overbrace{(1-2\omega)}^{\equiv \mathcal{D}_{tag}} (\widetilde{N}_B - \widetilde{N}_{\overline{B}})}{\varepsilon_{tag} (\widetilde{N}_B + \widetilde{N}_{\overline{B}})} = \mathcal{D}_{tag} \widetilde{A}_{tag}. \tag{8.9}$$

Mistagged events effectively dilute the tagging asymmetry by a factor $\mathcal{D}_{tag} \equiv 1 - 2\omega_{tag}$. Neglecting the dilution's uncertainty, A_{tag} is measured with a statistical uncertainty of[5]:

$$\sigma_{A_{tag}} \approx \sqrt{\frac{1 - A_{tag}^2}{\varepsilon_{tag} N_{tot}}}, \tag{8.10}$$

where N_{tot} is the total number of events, as defined in Eq. 8.4. Using Eq. 8.10, the uncertainty on the true asymmetry can be evaluated as:

$$\sigma_{\widetilde{A}_{tag}} = \frac{\sigma_{A_{tag}}}{\mathcal{D}_{tag}} = \frac{\sqrt{1 - A_{tag}^2}}{\sqrt{\varepsilon_{tag} N_{tot}} \mathcal{D}_{tag}} \approx \frac{1}{\sqrt{\underbrace{\varepsilon_{tag} \mathcal{D}_{tag}^2}_{\equiv P_{tag}} N_{tot}}} = \frac{1}{\sqrt{P_{tag} N_{tot}}}, \tag{8.11}$$

where $\sqrt{1 - A_{tag}^2} \approx 1$ is assumed.

5 The uncertainty on A_{tag} is obtained by propagating the uncertainty on N_B and $N_{\overline{B}}$ in the A_{tag} definition reported in Eq. 8.7 (left). The number of events is assumed following a Poisson distribution.

The tagging power is, therefore, an *effective efficiency*, constructed in such a way that a data sample with tagging power P_{tag} holds the same statistical power of a perfectly tagged sample (i. e., $\omega_{tag} = 0$) with $\varepsilon_{tag} = P_{tag}$. Two immediate consequences of its definition are:

$$\omega_{tag} = 0.5 \leftrightarrow P_{tag} = 0\,, \tag{8.12}$$

$$\omega_{tag} = 0 \leftrightarrow P_{tag} = \varepsilon_{tag}\,. \tag{8.13}$$

The standard approach in the development of flavor tagging tools is to carefully balance tagging efficiency and mistag rate in order to maximize the tagging power.

8.1.1 *Flavor tagging with per-event mistag probability*

Tagging performance can be greatly improved by assigning a *per-event mistag probability* (ω_{evt}) instead of using an average mistag rate. This approach allows making full use of every single event by exploiting the dilution as a weight in the final fit model[6]. The effective tagging power $\widehat{P}_{tag}$ is defined as the sum of the tagging powers of each individual event, as[7]:

$$\widehat{P}_{tag} = \sum_{evt}^{N_{tag}} P_{evt} = \sum_{evt}^{N_{tag}} \underbrace{\varepsilon_{evt}}_{=\frac{1}{N_{tot}}} \mathcal{D}^2_{evt} = \frac{1}{N_{tot}} \sum_{evt}^{N_{tag}} \mathcal{D}^2_{evt}\,, \tag{8.14}$$

where the "evt" subscript indicates per-event quantities. Accordingly, an effective dilution and an effective mistag rate can be defined as:

$$\widehat{\mathcal{D}}_{tag} \equiv \sqrt{\frac{\widehat{P}_{tag}}{\varepsilon_{tag}}}\,, \tag{8.15}$$

$$\widehat{\omega}_{tag} = \frac{1}{2}\left(1 - \sqrt{\frac{\widehat{P}_{tag}}{\varepsilon_{tag}}}\right)\,, \tag{8.16}$$

where the total efficiency $\varepsilon_{tag} = \sum_{evt}^{N_{tag}} \varepsilon_{evt} = N_{tag}/N_{tot}$ is equivalent to the averaged one defined in Eq. 8.2. The average tagging dilution $\mathcal{D}_{tag}$, already introduced in Eq. 8.6, can also be defined in terms of $\mathcal{D}_{evt}$ as:

$$\mathcal{D}_{tag} = \frac{1}{N_{tag}} \sum_{evt}^{N_{tag}} \mathcal{D}_{evt}\,. \tag{8.17}$$

6 See Section 9.2.1.

7 The sum can be extended to N_{tot} by defining $\mathcal{D}_{evt}$ as equal to zero for untagged events. The efficiency of a single event $\varepsilon_{evt} = 1/N_{tot}$ does not depend on the event index and is treated as a constant term.

The tagging dilution variance can also be introduced to describe the spread of the $\mathcal{D}_{\mathrm{evt}}$ distribution in a given event sample and is defined as:

$$\begin{aligned}
\sigma^2_{\mathcal{D}_{\mathrm{tag}}} &= \frac{1}{N_{\mathrm{tot}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}(\mathcal{D}_{\mathrm{evt}} - \mathcal{D}_{\mathrm{tag}})^2 \\
&= \frac{1}{N_{\mathrm{tot}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}\mathcal{D}^2_{\mathrm{evt}} + \underbrace{\frac{1}{N_{\mathrm{tot}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}\mathcal{D}^2_{\mathrm{tag}}}_{\substack{=\frac{N_{\mathrm{tag}}}{N_{\mathrm{tot}}}\mathcal{D}^2_{\mathrm{tag}} \\ =\varepsilon_{\mathrm{tag}}\mathcal{D}^2_{\mathrm{tag}}}} - \underbrace{\frac{2}{N_{\mathrm{tot}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}\mathcal{D}_{\mathrm{evt}}\mathcal{D}_{\mathrm{tag}}}_{\substack{=\frac{2\mathcal{D}_{\mathrm{tag}}}{N_{\mathrm{tot}}}\frac{N_{\mathrm{tag}}}{N_{\mathrm{tag}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}\mathcal{D}_{\mathrm{evt}} \\ =2\varepsilon_{\mathrm{tag}}\mathcal{D}^2_{\mathrm{tag}}}} \\
&= \frac{1}{N_{\mathrm{tot}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}\mathcal{D}^2_{\mathrm{evt}} - \varepsilon_{\mathrm{tag}}\mathcal{D}^2_{\mathrm{tag}} \geqslant 0\,.
\end{aligned} \tag{8.18}$$

Equation 8.18 can also be used to demonstrate that, by construction, the effective tagging power $\widehat{P}_{\mathrm{tag}}$ is larger, or at least equal, to the average one defined in Eq. 8.6:

$$\widehat{P}_{\mathrm{tag}} = \frac{1}{N_{\mathrm{tot}}}\sum_{\mathrm{evt}}^{N_{\mathrm{tag}}}\mathcal{D}^2_{\mathrm{evt}} = \sigma^2_{\mathcal{D}_{\mathrm{tag}}} + \underbrace{\varepsilon_{\mathrm{tag}}\mathcal{D}^2_{\mathrm{tag}}}_{=P_{\mathrm{tag}}} \geqslant P_{\mathrm{tag}}\,. \tag{8.19}$$

The performance improvement, compared to the averaged approach, is given by $\sigma^2_{\mathcal{D}_{\mathrm{tag}}}$. The larger the variance of the dilution distribution, the larger the performance enhancement. This translates to the necessity of, not only evaluate ω_{evt}, but also maximize the variance of its distribution, that is separate as much as possible right- and wrong-tag muons.

8.2 DEEP NEURAL NETWORKS

Deep learning (DL) is a broad family of learning techniques based on artificial neural networks (ANN) attempting to model data with complex architectures combining different nonlinear transformations.

8.2.1 *Artificial neural networks*

Artificial neural networks are computing systems, vaguely inspired by the biological neural networks, composed of connected elementary computational nodes called "artificial neurons". Each neuron has several inputs and produces a single output which is sent to multiple other neurons. The network consists of weighted connections, each connection providing the output of one neuron as one of the inputs to another neuron. Neurons are organized in *layers*, where each neuron is connected only to neurons of the immediately preceding and immediately following layers. The first layer, which receives the data, is called

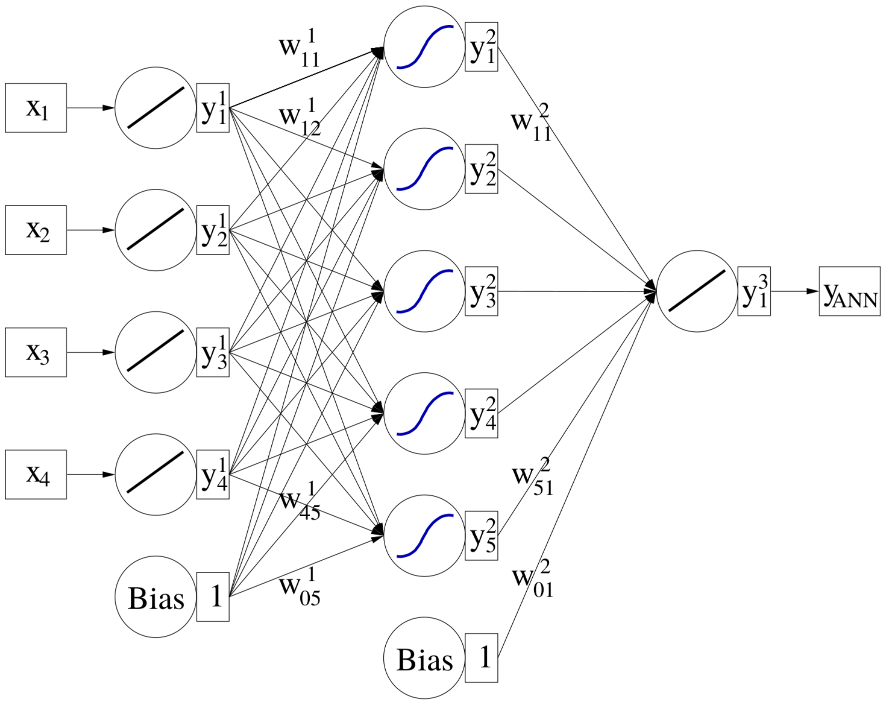


Figure 8.1: Schematic representation of an Artificial Neural Network with one hidden layer. The inputs of the network are indicated x_i, while the output of the i-th neuron in the j-th layer is indicated with y_i^j. The weights between the i-th neuron of the k-th layer and the j-th neuron of the next $(j+1)$ layer are indicated with w_{ij}^k. The bias term is assigned an index of zero. The output of the network is indicated with y_{ANN}. The naming and index conventions are slightly different from what is used in the text and refer only to this figure.

"input layer", while the last layer is called "output layer". All the other layers are called "hidden", as they are not accessed from outside the network. A schematic representation of an ANN is shown in Fig. 8.1. An ANN with multiple hidden layers is called a "deep neural network" (DNN). In this work, DNNs are considered only for binary classification tasks, in which methods are optimized to predict one of two *classes* (e.g. signal/background, right-/wrong-tag) for a given example of input data.

ARTIFICIAL NEURONS Mathematically a neuron is a function f of the input vector $\vec{x}$ weighted by a vector of weights $\vec{w}$, that produces an output y. This can be expressed as:

$$y = f(\vec{x}) = \phi(\vec{x} \cdot \vec{w} + \vec{b}) \,, \tag{8.20}$$

where $\vec{b}$ is a bias vector and ϕ is the so-called neuron "activation function". The weights and bias are randomly initialized and optimized through *learning* processes. Numerous activation functions can be considered, each with its ad-

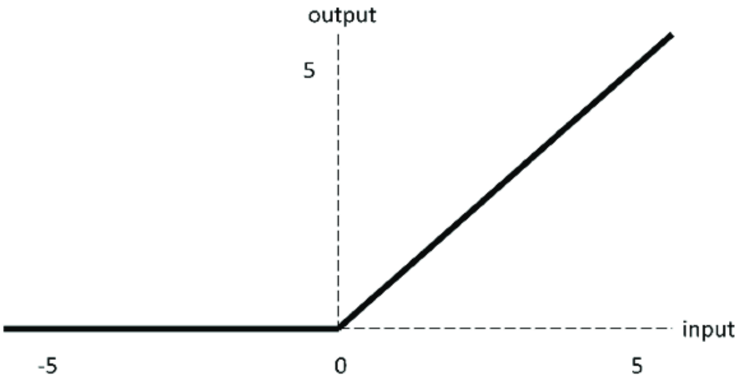


Figure 8.2: Plot of the ReLU function near $x = 0$.

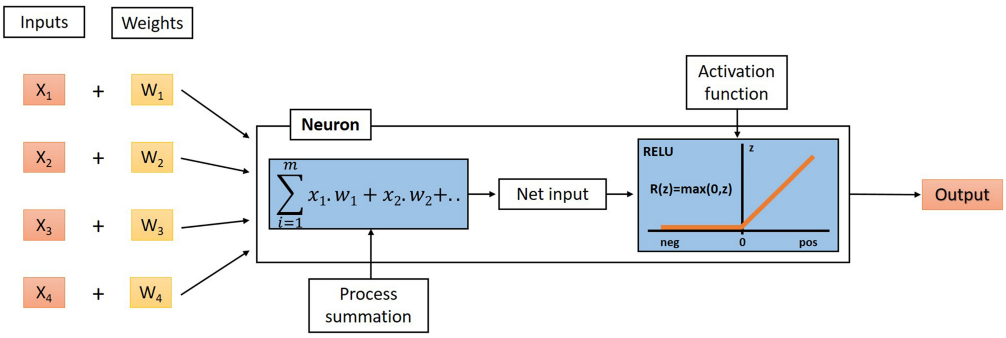


Figure 8.3: Schematic representation of an artificial neuron with ReLU activation function. Figure from Ref. [149].

vantages and different mathematical properties. In this work, the only activation function used in hidden layers is the *rectified linear unit* (*ReLu*), defined as:

$$R(x) = \max(0, x) , \tag{8.21}$$

and plotted in Fig. 8.2. As one of the most popular modern activation functions, the ReLu has several mathematical justification [147] and it has been demonstrated to improve the training of deeper networks [148]. A schematic representation of a neuron with ReLU activation is shown Fig. 8.3. Neurons in the output layer require special handling and their activation function is described later in the text[8].

LAYER TYPES Artificial neurons can be arranged in several different types of layers, which are used to tackle different types of tasks. All DNNs used in this work are built from a combination of *Dense* and *Dropout* layers.

8 Equation 8.22.

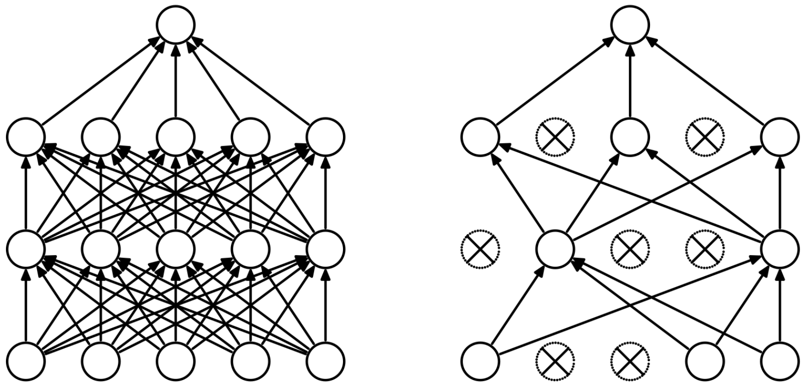


Figure 8.4: Artificial neural network with two hidden layers before (left) and after (right) applying dropout during the training phase. Figure from Ref. [151].

Dense layers are regular deeply connected layers where each neuron is connected to all the neurons of the previous and following layers. Dense layers are the most common type of layers in ANNs and DNNs, they can be arbitrarily large and are easily combined one after the other. The number of neurons in each layer and the total number of layers are hyperparameters to optimize case-by-case.

Dropout layers randomly set the input to zero with a given frequency called "dropout rate". They are typically applied after other types of layers as regularization tools to avoid overfitting the model [150, 151]. Dropout layers are only active during the training phase[9] and no values are dropped during inference. A schematic representation of a DNN with dropout is shown in Fig. 8.4.

OUTPUT The last layer in a DNN is responsible to produce the output score of the model and it is built differently from the others. The standard structure for an output layer in multiclassification tasks consists in one output neuron for each class with a *softmax* activation function [152][10]. The multidimensional softmax function is used to normalize the output of the network to a probability distribution over the predicted output classes, based on the Luce's choice axiom [153][11], and it is defined as:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{for } i = 1, \ldots, K \tag{8.22}$$

9 The training process is described in Section 8.2.2.

10 In the case of binary classification a single output neuron with *sigmoid* activation is mathematically equivalent to two neurons with *softmax* activation.

11 Luce's choice axiom states that the probability of selecting one item over another from a pool of many items is not affected by the presence or absence of the other items. Mathematically, it states that the probability $P(i)$ of selecting item i from a pool of j items is $P(i) = \frac{w_i}{\sum_j w_j}$, where w are a set of weights, so that if $P(i) > P(j)$ for the set (i, j), then $P(i) > P(j)$ for all sets $(i, j, \ldots, n)$.

where $\vec{z} = (z_1, \ldots, z_K)$ is the input vector of the last layer and K is the number of output classes. In practice, Eq. 8.22 takes an input vector of dimension K and normalizes it into a probability distribution consisting of K probabilities in the interval [0, 1] that add up to 1. A neural network is called "calibrated" if the output probability associated with the predicted class reflects its true posterior probability. Typically, simple ANNs produce well-calibrated probabilities on binary classification tasks [154], while more complex DNNs are found to be poorly calibrated [155].

8.2.2 *Neural networks training*

Learning (or *training*) is the process through which a network is optimized by analyzing training examples. In practice, this is achieved by adjusting the weights of the model to minimize a given *loss* function. In this work only supervised training is considered, where methods are optimized to produce a desired output for each given input.

LOSS FUNCTIONS The loss function used in all DNNs considered in this work is the so-called "cross-entropy", which for a single event is defined as:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{K} t_i \log(s_i)\,, \tag{8.23}$$

where t_i and s_i are the truth label and network output score, respectively, for each of the K classes. In supervised learning, the truth label is always 1 for the true class and 0 for the others. The cross-entropy can be interpreted as the negative log-likelihood for the conditional probability $P(\vec{t}\,|\,\vec{s})$ of the true class $\vec{t}$ given the output $\vec{s}$.

OPTIMIZERS One of the simplest optimizer used to minimize the loss function is the so-called "Stochastic Gradient Descent" (*SGD*), an iterative algorithm that searches for the minimum of a function by taking *steps* proportional to the negative of the function gradient. The stochastic approximation replaces the gradient, typically calculated from the entire data set, with an estimate calculated from a subset usually called "batch" in DL applications. At each iteration SGD updates the vector of weights $\vec{w}$, according to the gradient of the loss function Q, as:

$$\vec{w} \to \vec{w} - \alpha \nabla Q(\vec{w})\,, \tag{8.24}$$

where α is the learning rate, typically set by the user, and Q is evaluated in a given batch, usually constructed randomly.

Many improvements on the basic stochastic gradient descent algorithm exist, in particular, to adjust the learning rate that, being fixed in SGD, may be problematic for complex training applications. A popular SGD extension is the so-called "momentum method", which includes the previous update ($\Delta\vec{w}$) and uses it to determine the next one as [156]:

$$\vec{w} \rightarrow \vec{w} - \alpha\nabla Q(\vec{w}) + \beta\Delta\vec{w}\,, \tag{8.25}$$

where β is an exponential decay factor between 0 and 1. Momentum has been used for decades to improve the training convergence speed.

Another popular SGD extension is *Adam*, an adaptive learning rate optimization algorithm specifically designed for training deep neural networks [157]. While stochastic gradient descent maintains a single learning rate in all iterations, Adam computes individual learning rates for each weight using the first and second moments of the gradient. The method is computationally efficient and well suited for problems that are large in terms of data or parameters.

OVERFITTING AND REGULARIZATION Due to their complexity, deep neural networks are prone to *overfitting*, that is situations when the model adapts too well to the training data set and perform poorly outside of it. Overfitting can be detected by evaluating at each iteration the loss function in a validation data set. When the performance of the validation data set starts to decline it means that the model is no longer learning general features, but only specific characteristics of the training data set, which degrades the network generalization capabilities. An example of learning curves for underfitted and overfitted models are shown in Fig. 8.5.

The most direct way to reduce overtraining is to use larger training data sets or to simplify the model. However, these approaches are not always feasible, hence other regularization methods have to be used. Two very popular techniques, often used in conjunction, are the already described *dropout* and the *early stopping*.

Dropout randomly simplifies the model in each training iteration, forcing the network to not favor any specific feature, but to learn general properties. The dropout rate is the key hyperparameter to optimize in order to reduce overfitting without losing performance. *Early stopping* is the practice of stop the learning process after a given number of iterations during which no improvements on the validation loss have been made. A graphical representation is shown in Fig. 8.6. Due to its simplicity and straightforwardness, early stopping is a staple of the modern training of DNNs.

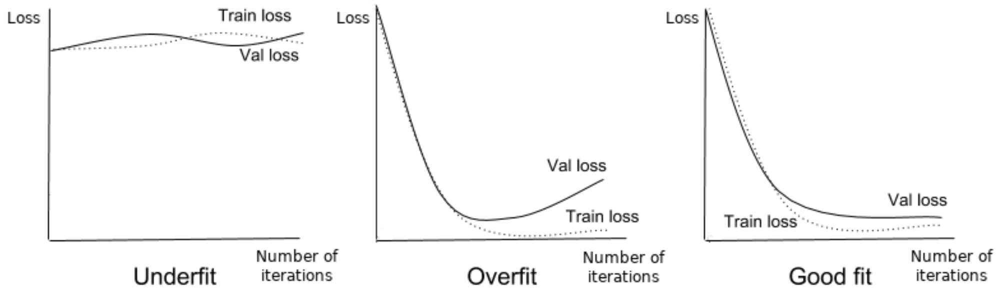


Figure 8.5: Learning curves for underfitted (left), overfitted (middle) and well fitted (right) models.

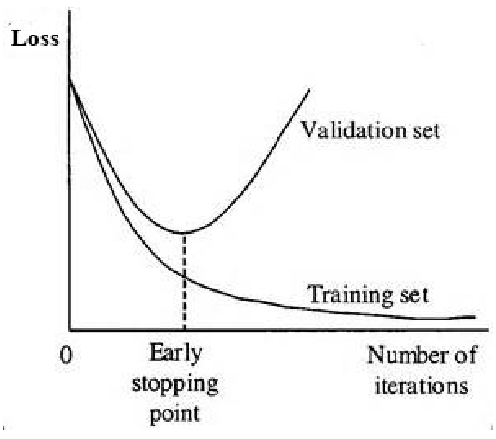


Figure 8.6: Graphical representation of the early stopping regularization method.

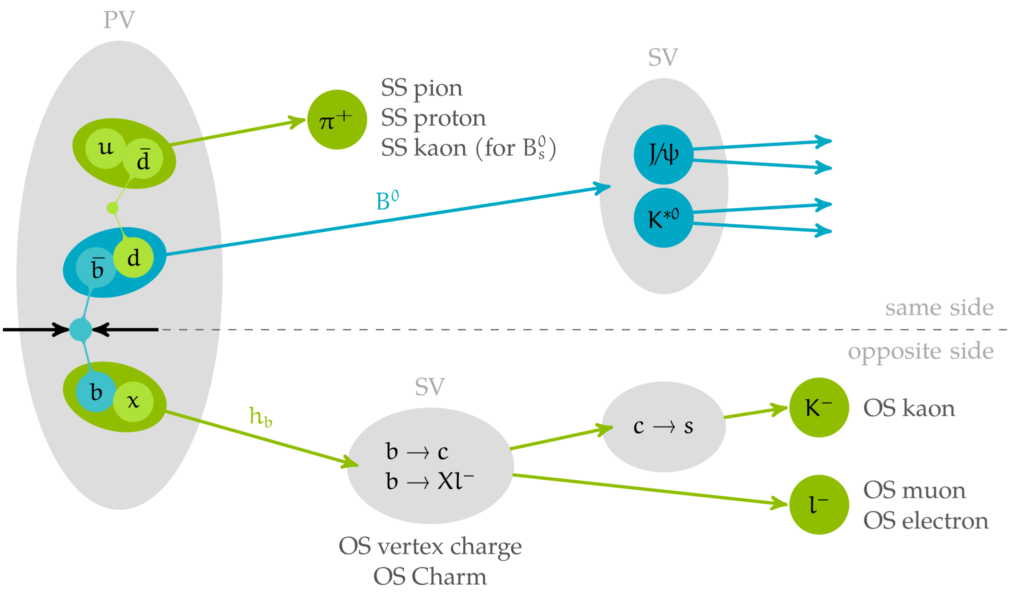


Figure 8.7: Schematic overview of various flavor tagging techniques. Same-side taggers exploit the properties of particles produced in the hadronization of the reconstructed ("signal") b-meson (in this case a B^0 meson). Opposite-side taggers exploit the decay products of the other b-hadron (h_b in the figure, "OS b-hadron" in the text) in the event, under the hypothesis of $b\bar{b}$ pair production. Figure adapted from Ref. [158].

8.3 TAGGING STRATEGIES

Flavor tagging can be achieved with a plethora of different techniques depending on the specific characteristics of the detector and the physics process of interest. Two complementary approaches can be defined depending on which *side* of the event is exploited: *same-side* (*SS*) and *opposite-side* (*OS*) taggers. A schematic overview of various SS and OS techniques is shown in Fig. 8.7.

8.3.1 *Same-side tagging*

Same-side techniques take advantage of the flavor correlations between the signal b-meson and nearby particles produced in its hadronization. In the case of the formation of a B_s^0 meson, the bottom antiquark bounds with a strange quark from a $s\bar{s}$ pair, leaving an isolated $\bar{s}$ quark. In about half of the times, the strange antiquark hadronizes into a K^+, whose charge can be used to infer the flavor of the B_s^0 meson with the tag decision logic:

$$K^+ \xrightarrow{\text{tag}} \text{signal } B_s^0$$
$$K^- \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0$$

In the case of a signal B^0 meson, analogous techniques can used exploiting charged pions and protons.

Same-side methods all require the identification of soft hadrons, which makes them experimentally more challenging than the opposite-side counterpart, which usually exploits leptonic decays. As such, SS taggers have been so far implemented in full-fledged analyses only by experiments with hadron particle identifications capabilities, such as CDF [159] and LHCb [69, 160, 161]. The feasibility of SS kaon tagging for B_s^0 mesons with CMS is currently under active study.

8.3.2 *Opposite-side tagging*

Opposite-side techniques exploit the decay products of the other b-hadron present in the event (hereafter "OS b-hadron"), relying on the fact that b-quarks are predominantly produced in $b\bar{b}$ pairs[12]. To obtain the highest possible tagging efficiency, OS taggers do not fully reconstruct the OS b-hadron, but instead exploit inclusive flavor-sensitive processes, where only a few decay products need to be actually analyzed. Once the flavor of the OS b-hadron is known, the tag decision is simply given by charge conjugation as:

$$\begin{aligned} \text{OS } b &\xrightarrow{\text{tag}} \text{signal } \bar{b} \\ \text{OS } \bar{b} &\xrightarrow{\text{tag}} \text{signal } b \end{aligned} \quad . \tag{8.26}$$

LEPTON TAGGING The most popular and performing opposite-side technique is "OS lepton tagging", where semileptonic $b \to \ell X$ ($\ell = e, \mu$) decays of the OS b-hadron are exploited. The tag decision is then given by the charge of the lepton (hereafter "OS lepton"), which is correlated with the flavor of the b-quark as:

$$\begin{aligned} \text{OS } \ell^- \to \text{OS } b &\xrightarrow{\text{tag}} \text{signal } \bar{b} \\ \text{OS } \ell^+ \to \text{OS } \bar{b} &\xrightarrow{\text{tag}} \text{signal } b \end{aligned} \quad . \tag{8.27}$$

Opposite-side lepton tagging is affected by several sources of dilution, the most prominent being:

- *cascade decays of the OS b-hadron*, where $b \to cX \to \ell^+ X'$ chains produce leptons with opposite charge correlation with respect to the expected one;
- *leptons from pileup interactions* (PU leptons) and *fake leptons* (consisting mostly in misidentified light hadrons and photons), which have no charge correlation with the flavor of the OS b-hadron;
- *double gluon splitting*, where two pairs of heavy quarks are produced ($b\bar{b}\,b\bar{b}$ or $b\bar{b}\,c\bar{c}$) and the selected OS lepton comes from a heavy quark uncorrelated with the signal b-meson;

12 See Section 4.1.2.

- *mixing of the OS b-hadron*, which reverses the correlation between the lepton charge and the OS b-hadron initial flavor.

The background[13] from cascade decays can be reduced by studying in detail the kinematic properties of the OS b-hadron decay, while contamination from PU and fake leptons can be reduced with selection requirements. Double gluon splitting and mixing of the OS b-hadron are, however, almost irreducible backgrounds, that inevitably dilute all taggers based on the opposite-side approach. These two effects combined set a lower limit for the mistag rate of OS techniques at $\omega_{\text{tag}}^{\text{min}} \approx 15\%$ ($\mathcal{D}_{\text{tag}}^2 \approx 0.5$). In practice, average mistag rates for OS lepton tagging techniques are about $\omega_{\text{tag}} \sim 30\%$ ($\mathcal{D}_{\text{tag}}^2 \sim 0.15$).

The maximum tagging efficiency for OS lepton taggers is limited by the semileptonic branching ratio $\mathcal{B}(b \to \ell X) \approx 20\%$ [34] and further reduced by OS hadrons that decay outside the CMS acceptance. However, data samples enriched of additional leptons, such as the one used in this work, can lead to ε_{tag} values higher than the semileptonic hadron branching fraction. As an example, if an additional lepton is required at trigger level and then it is *always* used for flavor tagging, the tagging efficiency would result $\varepsilon_{\text{tag}} = 100\%$, as each event in the data sample is tagged.

Even without much optimization, lepton taggers are well-balanced techniques, with good tagging efficiency and moderate mistag rates. As such, they are usually the first choices in most flavor dependent analyses.

JET TAGGING "OS jet tagging" is another popular opposite-side technique, where charge correlations in the fragmentation of the OS b-hadron in b-jets are exploited, by combining the charges of the b-jet components in the so-called "jet charge". The jet charge is usually computed as a p_T-weighted sum, in order to exploit the higher p_T of the b-quark products, as:

$$Q_{\text{jet}} = \frac{\sum_i p_{T,i}^k \cdot Q_i}{\sum_i p_{T,i}^k}, \tag{8.28}$$

where i is an index that runs over the b-jet components and k is a weighting exponent to be optimized (usually of the order of unity). The jet charge can also be computed with more sophisticated methods, such as with deep neural networks that exploit more complex charge correlations between the jet components. Typically, the flavor of the OS b-hadron is then determined from the sign of Q_{jet} with the tag decision logic:

$$\begin{aligned} Q_{\text{OS jet}} < 0 \to \text{OS } b \xrightarrow{\text{tag}} \text{ signal } \overline{b} \\ Q_{\text{OS jet}} > 0 \to \text{OS } \overline{b} \xrightarrow{\text{tag}} \text{ signal } b \end{aligned} \quad . \tag{8.29}$$

13 Here and through this chapter, *background* refers to dilution sources that can produce a tagging muon and not the background for $B_s^0 \to J/\psi\, \phi$ decays.

The sources of dilution discussed for lepton tagging techniques apply also for jet taggers, with the exception of cascade semileptonic decays and fake leptons. Usually, taggers have higher tagging efficiency than their lepton relatives, but also higher mistag rate, which combined may lead to an overall lower tagging power[14].

8.4 OPPOSITE-SIDE MUON TAGGING FOR $B_s^0 \to J/\psi\,\phi$

Opposite-side muon tagging is the flavor tagging technique used in the analysis presented in this dissertation. The choice of using only this tagger is motivated by the selected muon-enriched data sample, which provides at least one muon candidate for each event that can be used for tagging. However, the HLT selection does not assure that the additional muon is actually from the OS b-hadron. As such, additional selection requirements need to be used to ensure good performance, as described in Section 8.4.2. No other techniques are used, as their inclusion would lead to a negligible performance improvement in this specific data sample[15].

8.4.1 *Flavor tagging strategy*

The muon tagging algorithm developed for the $B_s^0 \to J/\psi\,\phi$ CP violation analysis is an opposite-side lepton tagger with per-event mistag probability. As such, the tagging inference is performed through three key steps:

1. *tagging muon selection*,
2. *tagging decision*, which following Eq. 8.27 is given by:

$$\begin{aligned} \text{OS } \mu^- &\to \text{OS } b \xrightarrow{\text{tag}} \text{signal } B_s^0 \\ \text{OS } \mu^+ &\to \text{OS } \bar{b} \xrightarrow{\text{tag}} \text{signal } \bar{B}_s^0 \end{aligned} \,, \tag{8.30}$$

3. *per-event mistag probability estimation*.

The muon selection procedure has the role of reducing the dilution sources while ensuring a large tagging efficiency, in order to maximize the final tagging power. On the other hand, estimating the mistag probability on an event-by-event basis maximizes the performance by exploiting the individual tagging power of each event, as described in Section 8.1.1.

---

14 In CMS, this has been observed in preliminary internal studies. Similar observations can also be made from the tagging performance quoted in two most recent ϕ_s results by ATLAS [62, 63]. In both cases, the jet tagging algorithms have the highest tagging efficiency, but also the highest effective mistag rate.

15 The combined tagging power of the OS electron and OS jet taggers is estimated to be more than one order of magnitude smaller than the one of the OS muon tagger, in the selected data sample.

The algorithm is developed on simulated $B_s^0 \to J/\psi\,\phi$ events[16], where the true initial B_s^0 flavor is known from MC information ("MC truth"), and calibrated on a data sample of self-tagging $B^+ \to J/\psi\,K^+$ decays, where the charge of the kaon is related to the charge and flavor of the B^+ meson. The correction for the so-called "simulation bias" [162] is implied in all the simulated samples[17].

8.4.2 *Tagging muon selection*

In understanding the tagging muon selection procedure, it is important to remember that, due to the trigger used, each event presents at least one muon candidate in addition to the two from the signal $J/\psi \to \mu^+\mu^-$ decay. As such, if no additional selection requirement is applied, the tagging efficiency should be $\varepsilon_{tag} \approx 100\%$[XVIII]. While this approach ensures the highest possible tagging efficiency, it is not ideal, as it leads to a high mistag rate ($\omega_{tag} \approx 40\%$) and poor overall tagging power ($P_{tag} \approx 3.5\%$).

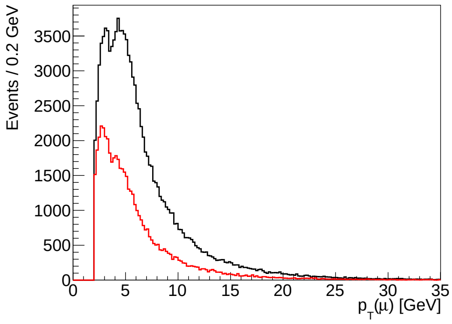
To reduce the dilution, the tagging muon candidates are selected with a combination of reconstruction and quality requirements. All candidates need to be reconstructed by the Particle Flow algorithm and have a matching track in the tracking system. The tracking information provides a precise measurement of the muon charge. Tracks linked to the reconstructed signal b-meson decay are explicitly excluded from consideration. All tagging muon candidates are also required to pass the HLT kinematic requirements of $p_T > 2$ GeV and $|\eta| < 2.4$.

SELECTION OPTIMIZATION The tagging muon selection is optimized with a multidimensional grid scan on the selection cut values, to maximize the tagging power. All selection configurations within 5% of the best one are considered equally good and the one that maximizes the tagging efficiency is chosen. The distributions of the selection variables in simulated $B_s^0 \to J/\psi\,\phi$ events are shown in Fig. 8.8, while the chosen configuration is reported in Tab. 8.1, and discussed hereafter. In ≈3% of the events more than one tagging muon candidate pass all the selection requirements, in these cases, the one with the highest p_T is chosen.

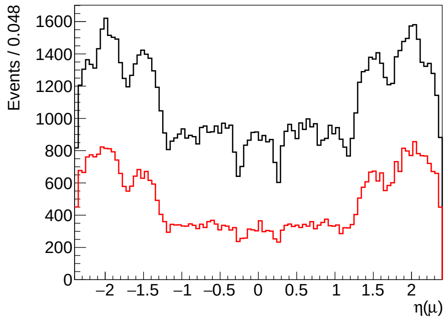
16 The MC samples are described in detail in Section 7.3.

17 In flavor tagging, "simulation bias" refers to the phenomenon for which the flavor composition of the *unconstrained* side in simulated events where a specific process is forced (e. g. $B_s^0 \to J/\psi\,\phi$) differs from that which would be observed in a generic $b\bar{b}$ production. This bias can be easily corrected by reweighing a portion of the events as described in Section 3.4.2.2 of Ref. [162]

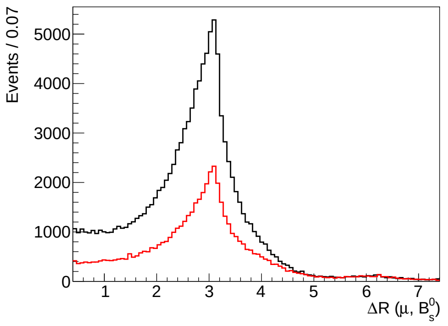
XVIII The tagging efficiency, in the case where no additional selection is applied, is not exactly 100% since a small fraction of trigger muon objects does not end up to be reconstructed as muon by the Particle Flow algorithm.



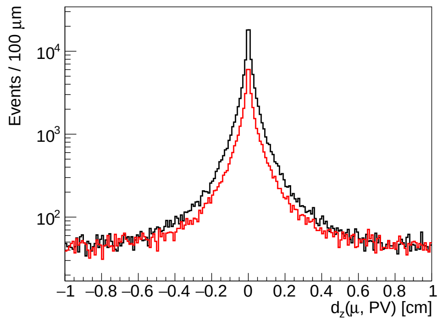
(a) Transverse momentum distribution.



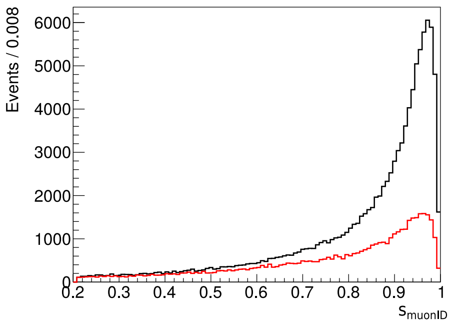
(b) Pseudorapidity distribution.



(c) $\Delta R_{\eta,\phi}(\mu, B_s^0)$ distribution.



(d) Longitudinal impact parameter distribution.



(e) DALIHA discriminator score distribution. This variable is defined later in the text and described in detail in Appendix A.

Figure 8.8: Distribution of the tagging muon selection variables in simulated $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ events for right-tag muons (black line) and wrong-tag muons (red line). based on generation information. All selection requirements reported in Tab. 8.1 are applied. Histograms are not stacked.

Table 8.1: Optimized selection requirements for tagging muon candidates. Tracks associated with the reconstructed b-meson are explicitly excluded from the pool of candidates. The output score value of the DNN against light hadrons (DALIHA) discriminator is indicated with s_{muonID}. This variable is defined later in the text and described in detail in Appendix A. All requirements are optimized.

Variable	Requirement
Muon reconstruction	*Global muon*
p_T	> 2 GeV
$\|\eta\|$	< 2.4
$\|d_z\|$	< 1 cm
$\Delta R_{\eta,\phi}(\mu, B^0_s)$	> 0.4
DALIHA discriminator score	$s_{\mathrm{muonID}} > 0.21$ (*Loose WP*)

RECONSTRUCTION Tagging muon candidates are required to be reconstructed as *global muons* by the PF algorithm, as described in Section 5.3, and to have a distance from the direction of the signal b-meson momentum in the (η, ϕ) plane of $\Delta R > 0.4$.

PILEUP MITIGATION The largest reducible source of dilution, by far, are muons originating in pileup interactions. Luckily, this background can be greatly reduced by exploiting differences in the longitudinal impact parameter (d_z) distributions between muons from OS b-hadrons and PU muons. The 3D impact parameter (d_{xyz}) of a track is defined as the distance of closest approach with respect to the selected primary vertex. As shown in Fig. 8.9, the d_z distribution for tagging muons clearly shows two different trends:

- muons originating from b-hadrons are Gaussian distributed around $d_z \approx 0$, as the majority of them come from the PV,
- on the other hand, muons that are not associated to b-decays show two components: a Gaussian core, associated to prompt muons, and a very large flat component, associated with muons from pileup interactions.

Therefore, the number of pileup muons can be greatly reduced with a requirement on the maximum absolute value of the longitudinal impact parameter. In this algorithm, tagging muon candidates are required to have $|d_z| < 1$ cm.

FAKE MUON MITIGATION In the CMS detector hadrons are usually detected by being completely absorbed in the ECAL and HCAL. However, light hadrons, such as kaons or pions, can also decay in flight into muons or travel enough through the calorimeters and the solenoid to leave a signature in the muon

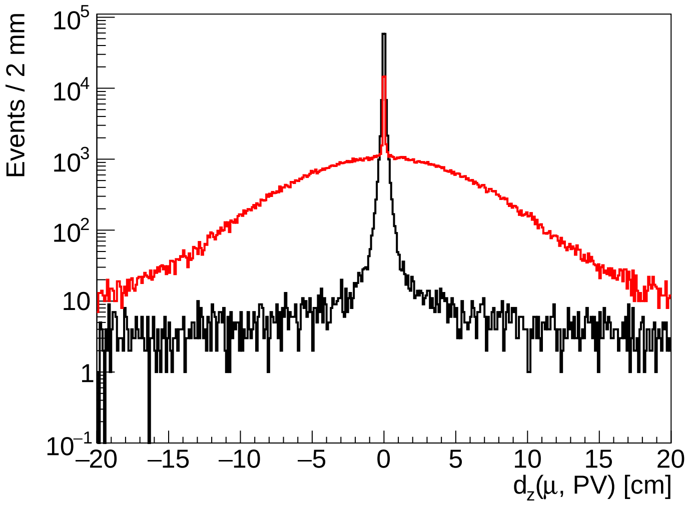


Figure 8.9: Distribution of the impact parameter in the longitudinal direction for tagging muon candidates originating from b-hadrons (black line) and from any other source (red line). Data extracted from $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ simulated events using MC truth. No other selection is applied to the tagging muon candidate, with the exception of the HLT kinematic requirements.

system. Due to the enormous amount of particles produced in pp collisions, the number of hadrons misreconstructed as muons is not negligible, especially at low p_T values. Through the years, the CMS Collaboration has developed several muon *discriminators*, which are able to reduce this type of background by a combination of selection requirements on the quality of the muon reconstruction. However, due to the very low momentum of the muons used by this tagging algorithm, no existing muon discriminator was found to be suitable for the task. As such, a new discriminator based on deep neural networks has been developed using the KERAS library [163] within the TMVA toolkit. This discriminator, called "DNN against light hadrons" (DALIHA) in the following, uses 25 input features related to the muon kinematics and reconstruction quality and is trained with 3.5×10^6 simulated muon candidates of which 2.5×10^5 are misreconstructed hadrons[19]. The development of this discriminator is described in detail in Appendix A. Tagging muon candidates are required to pass a working point of the DNN output that has an efficiency of ≈98% for genuine muons and ≈33% for misreconstructed light-flavor hadrons when evaluated using global muon candidates. This working point is referred to as "*Loose* working point", due to its high efficiency for genuine muons and corresponds to a requirement on the output score of $s_{muonID} > 0.21$.

19 As described in Appendix A, the genuine muon data sample is reweighed by a factor of 0.1 to balance the training data set.

TAGGING PERFORMANCE The average tagging performances obtained, after the optimized tagging muon selection requirements are applied, are reported in Tab. 8.2, alongside the performance at each step of the selection procedure (*cut flow*).

Table 8.2: Average flavor tagging performance in simulated $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ events when different additional selection requirements are applied to the tagging muon candidate. The $B_s^0 \to J/\psi\,\phi$ candidates are selected with the same selection requirements used for the CPV analysis and described in Chapter 7. "*HLT kinematics*" stands for offline p_T and $|\eta|$ requirements that reflect the selection at HLT level on the additional muon ("*no additional selection*" scenario). "*ID*" refers to requiring both the muon candidate to be reconstructed as a global muon and to pass the score requirement on the DALIHA discriminator. The numerical values of the various selection requirements are reported in Tab. 8.1. In all the scenarios, the muon candidate needs to be reconstructed by the Particle Flow algorithm and to be linked to a track reconstructed in the tracking system.

Selection requirements	ε_{tag}	ω_{tag}	P_{tag}
HLT kinematics	98%	40%	3.6%
Kinematic + $\Delta R_{\eta,\phi}$	93%	40%	3.7%
Kinematic + $\Delta R_{\eta,\phi}$ + ID	80%	38%	4.5%
Kinematic + $\Delta R_{\eta,\phi}$ + ID + d_z (final)	52%	32%	6.6%

8.4.3 *Per-event mistag probability*

The evaluation of the mistag probability of each event greatly enhances tagging performance, as described in Section 8.1.1. In this algorithm the mistag probability is predicted with a DNN, hereafter called "muon tagger DNN", that also further discriminates right- and wrong-tag muons. The model has been developed using the KERAS machine learning library within the TMVA toolkit and it uses 14 input features related to the muon kinematics and surrounding activity.

The main advantage of using a DNN is the possibility to exploit the properties of the output score to directly predict the mistag probability with very high accuracy. As described in Section 8.2, the *softmax* function can be used as activation function in the network output layer to obtain the probability of the two classes given the input of the output layer $\vec{z} = (z_1, z_2)$, as:

$$\begin{aligned} \sigma(z_1, z_2)_1 &= \frac{e^{z_1}}{e^{z_1} + e^{z_2}} \\ \sigma(z_1, z_2)_2 &= \frac{e^{z_2}}{e^{z_1} + e^{z_2}} \end{aligned} \quad \text{with } \sigma_1 + \sigma_2 = 1\,. \tag{8.31}$$

In this model the right-tag muons are defined as *class-1*, while the wrong-tag ones are defined as *class-2*[XX]. However, while the softmax function output can be interpreted as a probability distribution[21], it is not guaranteed that it is equal to the true probability distribution of the classes.

The *cross-entropy* loss function can be used, in conjunction with softmax outputs, to produce well-calibrated networks, as it can be interpreted as the negative log-likelihood for the conditional probability $P(\vec{t}\,|\,\vec{s})$, where $\vec{s}$ is the model output score vector and $\vec{t}$ the classes truth labels. In supervised learning, where the class labels are known, the truth label vector is typically *one-hot* encoded, that is every entry of $\vec{t}$ is equal to 0 but the one corresponding to the true class, which is equal to 1. The output score vector $\vec{s}$ contains at every entry the corresponding predicted class probability.

In binary classifications the cross-entropy of a given event is equal to:

$$\mathcal{L}_{\text{CE}} = \begin{cases} -\log(s_1) & \text{if } \vec{t} = (1,\,0) \\ -\log(1-s_1) & \text{if } \vec{t} = (0,\,1) \end{cases}, \tag{8.32}$$

where $\vec{s} = (s_1, s_2)$ the normalization relationship $s_1 + s_2 = 1$ has been used. When evaluated on a data set of size n the cross-entropy becomes:

$$\mathcal{L}_{\text{CE}}^{\text{dataset}} = -\sum_{i=1}^{n} \left[-t_{1,i} \log(s_{1,i}) - (1-t_{1,i}) \log(1-s_{1,i}) \right]. \tag{8.33}$$

The smaller the total cross-entropy in Eq. 8.33 is, the better the predicted probabilities resembles the true probabilities. Since predicting the mistag probability is of primary interests, the muon tagger DNN is optimized by minimizing $\mathcal{L}_{\text{CE}}^{\text{dataset}}$. Once the model is trained, the mistag probability can be calculated as

$$\omega_{\text{evt}} = 1 - s_1\,, \tag{8.34}$$

where complement to one of s_1 is used, instead of s_2, since the TMVA framework provides only the output score of the class labeled as *Signal*, which in this case corresponds to right-tag muons.

8.4.3.1 *Training data set*

The DNN model is trained on $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ events simulated in the 2018 data-taking conditions. The B_s^0 meson is identified using MC truth and selected with requirements on the generated momenta and pseudorapidity values. Muons from $J/\psi \to \mu^+\mu^-$ decays are required to be generated with $p_T > 3.5$ GeV and $|\eta| < 2.4$, while the B_s^0 meson is required to have $p_T > 10$ GeV and $|\eta| < 2.5$. This selection, summarized in Tab. 8.3, ensures signal b-mesons in

XX Class-1 (right-tag muons) corresponds to the TMVA *Signal* class, while class-2 (wrong-tag muons) correspond to the TMVA *Background* class.

21 The softmax output satisfies the properties of a probability $\sigma_i \in [0,\,1]$ and $\sum_i^K \sigma_i = 1$.

the same kinematic region of the B_s^0 candidates used in the CPV analysis. The tagging muon candidates are selected with the same requirements described in Section 8.4.2.

A total of 562 582 muon candidates are selected for training, out of which 169 226 are muons that carry the wrong flavor information ("wrong-tag"). The data set is then further split into *training*, *validation* and *testing* subsets with respective fractions of 50%, 25% and 25%. The training data set is used to train the DNN model, the validation data set is used to evaluate the performance at each training iteration (also called "epoch") to control overfitting and trigger early stopping, while the testing data set is used to evaluate the performance and overtraining after the training is finished.

Table 8.3: B_s^0 selection requirement for the muon tagger DNN training data set. All requirements are applied to the generated quantities.

Variable	Requirement
Gen matched	✓
$p_T(\mu_{gen})$	> 3.5 GeV
$\lvert\eta(\mu_{gen})\rvert$	< 2.4
$p_T(B^0_{s,gen})$	> 10 GeV
$\lvert\eta(B^0_{s,gen})\rvert$	< 2.5

8.4.3.2 *DNN input features*

The muon tagger DNN combines several input features in order to discriminate right-tag muon from wrong-tag ones and infer the mistag probability. A total of 14 variables have been chosen for their discrimination power against various dilution sources.

KINEMATICS

- p_T: muon transverse momentum
- η: muon pseudorapidity

Information on the muon candidate kinematics are helpful in discriminating kaons and pions misidentified as muon. In fact, light hadrons are produced for the most part in the forward region and with softer momenta compared to muons from $b \to \mu^- X$ decays, as shown in Fig. 8.10a. Since these variables do not carry flavor related information they are mostly helpful to identify muons from dilution sources with $\omega_{evt} \approx 50\%$.

HADRON DISCRIMINATOR

- s_{muonID}: output score of the DALIHA discriminator

The output score of the muon discriminator is added for the same reason as the kinematic variables, that is to discriminate misidentified muons that carry no flavor information ($\omega_{\mathrm{evt}} \approx 50\%$).

IMPACT PARAMETER

- d_{xy}: signed muon impact parameter in the transverse plane
- d_z: muon longitudinal impact parameter
- $\sigma(d_{xy})$: uncertainty on d_{xy}
- $\sigma(d_z)$: uncertainty on d_z

The impact parameter is very sensitive to the nature of the decay process that generates the muon candidate. High values of the transverse impact parameter are associated with decays far from the primary vertex, such as decay in flight of light hadrons or decay of long-lived b-hadrons, while low values are associated with decays of c-hadrons and other short-lived species. Moreover, muons associated with small impact parameters are more likely of the correct flavor tag if originating from OS B^0 mesons, as this species is subject to time-dependent flavor mixing[22]. While, in theory, the same reasoning applies also to B_s^0 mesons, their high oscillation frequency makes it practically impossible to use this type of information.

On the other hand, the impact parameter in the longitudinal direction is a strong discriminating variable against muons from pileup interactions, as already described in Section 8.4.2 and shown in Fig. 8.9. In combination with other variables, d_z can also help in identifying prompt hadrons misidentified as muons, since they have a small impact parameter.

In summary, the impact parameter distribution is made up of several different components that depend on the specificity of the muon origin, which are difficult to disentangle using only the d_{xyz} distribution. Hence, the various components of the impact parameter make good candidates to be included in a multivariate method. The uncertainty on the impact parameter is also given to improve its modeling.

SEPARATION

- $\Delta R_{\eta,\phi}(\mu, B)$: distance in the η-ϕ plane between the muon and the signal b-meson momentum.

The angular separation between the tagging muon candidate and the signal b-meson is used in the selection as a veto for the same-side activity and can be used by the muon tagger DNN to construct more complex variables that relate the muon features to the B_s^0 flavor.

22 Bottom quarks hadronize into a B^0 meson ≈40% of the times [34].

ISOLATION

- Iso_μ: muon isolation, defined as

$$\mathrm{Iso}_\mu = \frac{\sum_{h^\pm} p_T + \max\Big(0,\ \sum_{h^0} E_T + \sum_\gamma E_T - 0.5 \sum_{h^\pm_{PU}} p_T\Big)}{p_T^\mu} \tag{8.35}$$

  where

  - $\sum_{h^\pm} p_T$ is the sum of the transverse momenta of the charged tracks linked to the same vertex as the muon and contained in a $\Delta R_{\eta,\phi} < 0.4$ cone around the muon momentum;
  - $\sum_{h^0} E_T$ and $\sum_\gamma E_T$ are the sum of the transverse energies of the neutral hadrons and the photons, respectively, contained in the $\Delta R_{\eta,\phi} < 0.4$ cone;
  - $\sum_{h^\pm_{PU}} p_T$ is the sum of the transverse momentum of the charged tracks associated with pileup interactions, contained in the $\Delta R_{\eta,\phi} < 0.4$ cone, and $-0.5 \sum_{h^\pm_{PU}} p_T$ is the so-called "$\Delta\beta$" correction for contamination from PU neutral particles[23].

The muons isolation is a variable widely used to separate muons produced in the decay of light hadrons, as they are more likely to be surrounded by particles from the hadronization of the light-quark jet and from muons produced in the decay of heavy quarks. This feature is therefore useful in discriminating $b \to \mu^- X$ direct decays from $b \to cX \to \mu^+ X'$ cascade decays, which are related to mistag probabilities $>50\%$.

SURROUNDING ACTIVITY To further study the muon surrounding activity, a jet-like object (hereafter "muon cone") is constructed from all the Particle Flow candidates contained in a $\Delta R_{\eta,\phi} < 0.4$ cone, centered in the muon momentum, with $p_T > 0.5$ GeV, $|\eta| < 3$ and $d_z < 1$ cm[XXIV]. Unless stated otherwise, the tagging muon is assumed to be part of the cone.

The following variables are defined combining the muon and muon cone properties:

- p_T^{cone}: total transverse momentum of the muon cone, evaluated as the sum of the transverse momenta of its components,
- $p_T^{\mathrm{rel}} = \vec{p}_T^{\,\mu} \cdot (\vec{p}^{\,\mathrm{cone}} - \vec{p}^{\,\mu})$: muon $\vec{p}_T$ projected on the cone axis, after subtracting the muon momentum,

23 The estimation that the amount of E_T carried by PU neutral hadrons and photons combined is half the sum of p_T of the PU charged hadrons comes from the fact that the majority of the hadrons produced in pp collisions consists in pions and, due to isospin symmetry, the production ratio between charged pions and neutral pions is 2:1. Charged PU particles can be more easily removed, since they are not associated with the PV.

XXIV A *home-made* cone is used instead of jet objects since these are not reconstructed in the kinematic region of interest.

- $\Delta R_{\eta,\phi}(\mu, \text{cone})$: distance in the η-ϕ plane between the muon and the cone axis[25],
- E_μ/E_{cone}: ratio of the muon energy to the cone energy[26],

All these variables have been proven to be helpful in separating direct semileptonic decays ($b \to \mu^- X$) from cascade ones ($b \to cX \to \mu^+ X'$) [162, 164].

CONE CHARGE

- Q_{cone}: cone charge, computed as[27]:

$$Q_{\text{cone}} = Q_\mu \frac{\sum_i^{\text{cone}} p_{T,i} \cdot Q_i}{\sum_i^{\text{cone}} p_{T,i}}\,, \tag{8.36}$$

where Q_μ is the charge of the tagging muon, used as multiplicative factor to remove the dependence of Q_{cone} on the tagging decision[28]. The cone charge is constructed following the same general strategy of the jet tagging techniques described in Section 8.3.2 and it helps identifying the OS b-hadron flavor.

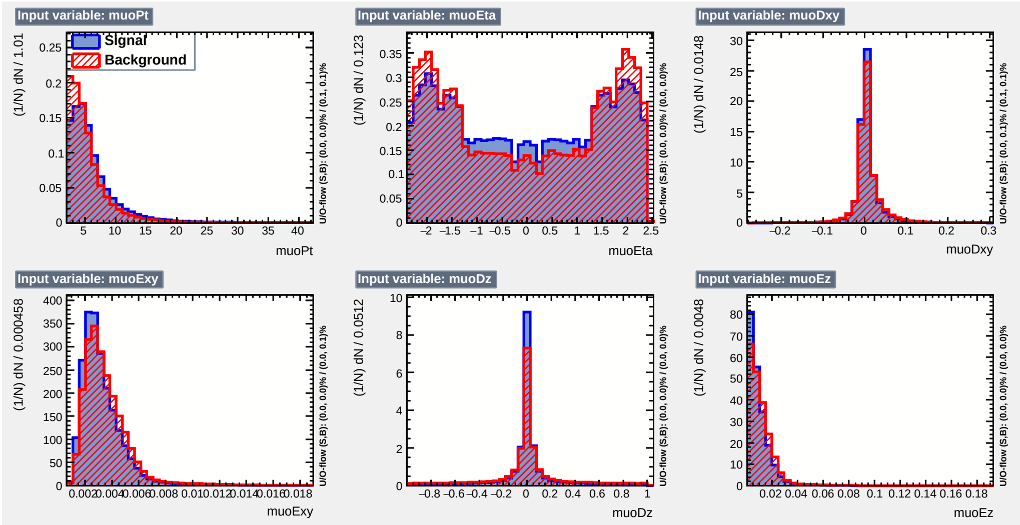
The distributions of the 14 input features of the muon tagger DNN are shown in Fig. 8.10.

FEATURE PREPROCESSING The input features are preprocessed in order to increase the performance of the DNN method. The variables are first scaled in the range $[-1, 1]$, then their distribution is transformed into a Gaussian ("Gaussianisation")[29] and finally normalized again. The scaling is helpful to the training process because it ensures that both positive and negative values are used as inputs, which makes learning more flexible, and that the network's regards all features to a similar extent without favoritism. Gaussianisation is found to speeds up learning and to improve the convergence properties when combined with a Gaussian random initialization of the model weights. The preprocessing procedures are handled with the TMVA toolkit.

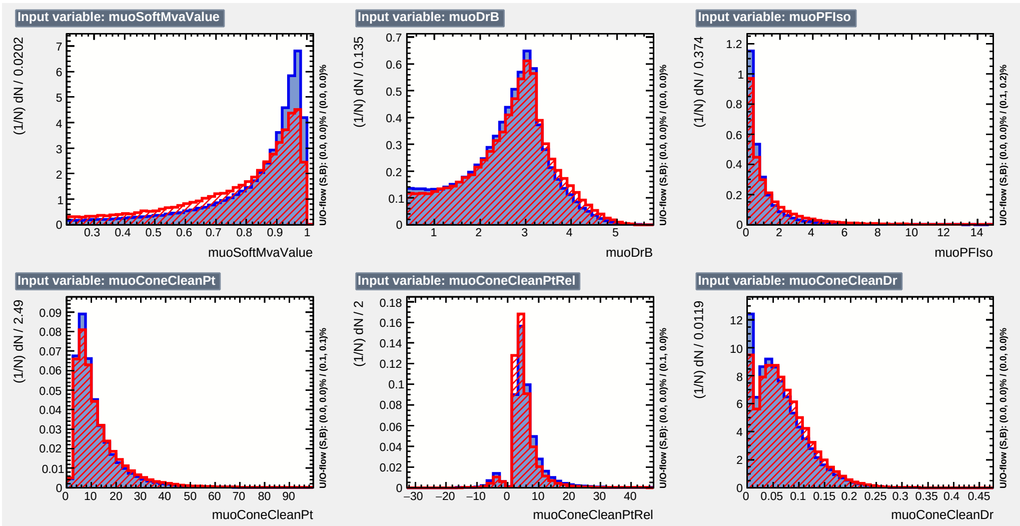
8.4.3.3 *DNN model optimization, training, and performance*

The DNN model is built with the KERAS library and trained within the TMVA framework. In order to construct the network with the best right- vs wrong-tag

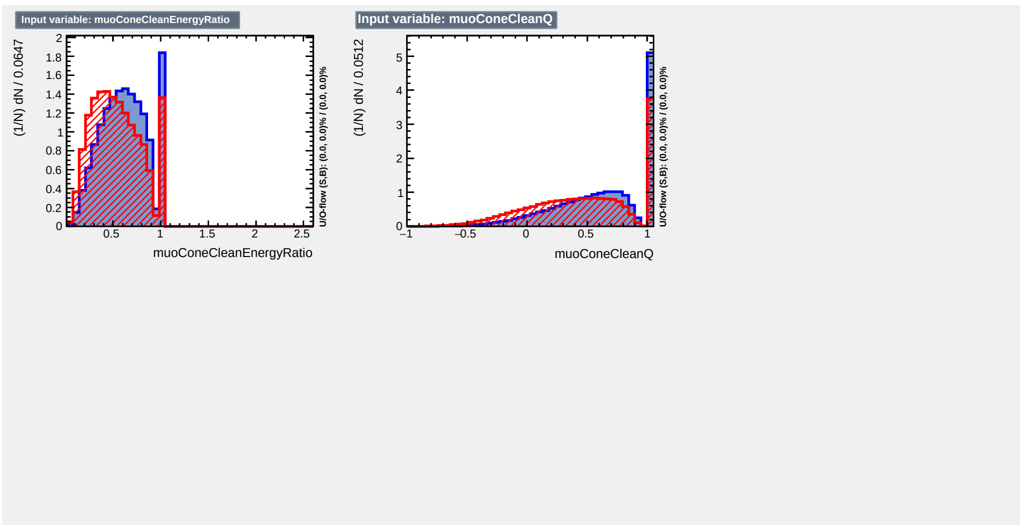
25 The cone axis is defined as the direction of the total momentum of the cone.
26 Since the tagging muon is not removed from the cone, E_μ/E_{cone} values of one are possible.
27 Since the tagging muon is not removed from the cone, Q_{cone} values of one are possible.
28 The muon tagger DNN task is to discriminate right- and wrong-tag, not identifying the flavor of the OS b-hadron.
29 Gaussianisation proceeds in two steps: first the input distribution is transformed into a uniform distribution using its cumulative distribution function. Next, the inverse error function is used to transform the uniform distribution into a Gaussian with zero mean and unity width. For more detail see the TMVA Users Guide found in Ref. [142].



(a) From left to right, top to bottom: p_T, η, d_{xy}, $\sigma(d_{xy})$, d_z and $\sigma(d_z)$.



(b) From left to right, top to bottom: s_{muonID}, $\Delta R_{\eta,\phi}(\mu, B)$, Iso_μ, p_T^{cone}, p_T^{rel} and $\Delta R_{\eta,\phi}(\mu, \mathrm{cone})$.



(c) From left to right: $E_\mu / E_{\mathrm{cone}}$ and Q_{cone}.

Figure 8.10: Muon tagger DNN input features distributions for right- (blue) and wrong-tag (red) tagging muon candidates. The distributions are normalized. The TMVA plotting tool automatically calls class-1 (in this case right-tag) "Signal" and class-2 (in this case wrong-tag) "Background". However, this naming scheme is not related to any other definition of signal and background given in this dissertation.

separation, several hyperparameters related to the DNN structure are optimized with a grid scan. The parameter configuration that maximizes the area under the *Receiver Operating Characteristic* (*ROC*) curve, with no observable overfitting, is chosen. The ROC curve is a graph that illustrates the ability of a binary classifier at all classification thresholds, with the area under the curve (AUC) being a widely used proxy to measure the overall performance[30]. Between equally performing models, the simpler one is chosen. The tested values in the grid scan and the chosen best parameter configuration are reported in Tab. 8.4.

The optimal configuration is found to be three *Dense* layers of 200 neurons, each followed by a *Dropout* layer with a dropout rate of 40%. The activation function of the neurons in the Dense layers is the *ReLU*. The model weights are optimized with the *Adam* optimizer using a batch size of 1024 events[31]. The loss function is the *cross-entropy*, to ensure the best probability prediction. The maximum number of training epoch is 50, however, to avoid overfitting the training is stopped early after 10 epochs with no accuracy improvement, and the model with the smaller loss is kept. The number of epochs, batch size, and optimizer learning rate are mildly optimized by testing different values with the chosen model. The full list of training hyperparameters is reported in Tab. 8.5.

Table 8.4: Muon tagger DNN model hyperparameters, optimized through grid scan, and their tested values. The values belonging to the optimal configuration are underlined.

Hyperparameter	Tested values
Number of Dense layers	[2, <u>3</u>, 4, 5]
Neurons for each layer	[100, 150, <u>200</u>, 300]
Dropout probability	[0.1, 0.2, 0.3, <u>0.4</u>, 0.5]

Table 8.5: Muon tagger DNN training hyperparameters.

Hyperparameter	Value
Optimizer	Adam
Optimizer learning rate	0.001
Batch size	1024
Number of training epochs	50
Early stopping epochs	10

30 A perfect classifier has an AUC of 1, while one that randomly guesses has an AUC of 0.5.
31 The batch size is the number of training examples used for a single gradient step during the training process.

MUON TAGGER DNN PERFORMANCE The performance of the trained model is evaluated in a test data set statistically independent from the training and validation samples. The ROC curve for the trained model and the output score distribution are shown in Fig. 8.11. No signs of overfitting are observed when comparing the output distributions obtained from the training and testing data sets.

Since the inner workings of a neural network are typical of difficult interpretations, it is not possible to study how the input features are combined by the muon tagger DNN and what are the prominent properties of the various sources of dilution. However, the output score distribution shown in Fig. 8.11b can be still qualitatively commented, remembering that the output score s_{dnn} is equal to $1 - \omega_{\text{evt}}$. The network is clearly able to partially identify tagging muons that carry no flavor information, such as muons PU and fakes, by clustering them around $\omega_{\text{evt}} \approx 50\%$. This is a very helpful feature, as now the flavor inference on these events can be ignored in the final fit to extract the physics parameters. The network is also able to recover the correct flavor inference of a portion of mistagged events, by assigning a mistag probability higher than 50%. These are likely muons from cascade decays, as it is the only identifiable dilution source that uniquely inverts the flavor correlation. Lastly, most of the right-tag muons are grouped around $\omega_{\text{evt}} \approx 20\%$, which makes sense for tagging muon from $b \to \mu^- X$ decays, as $\omega_{\text{tag}}^{\text{min}} \approx 15\%$ is the lower boundary for the average mistag rate due to mixing of the OS b-hadron[32].

In summary, the muon tagger DNN is able to partially separate the various dilution sources by exploiting their different properties. It is also able to directly predict the right-tag probability s_{dnn} and by consequence the mistag probability as

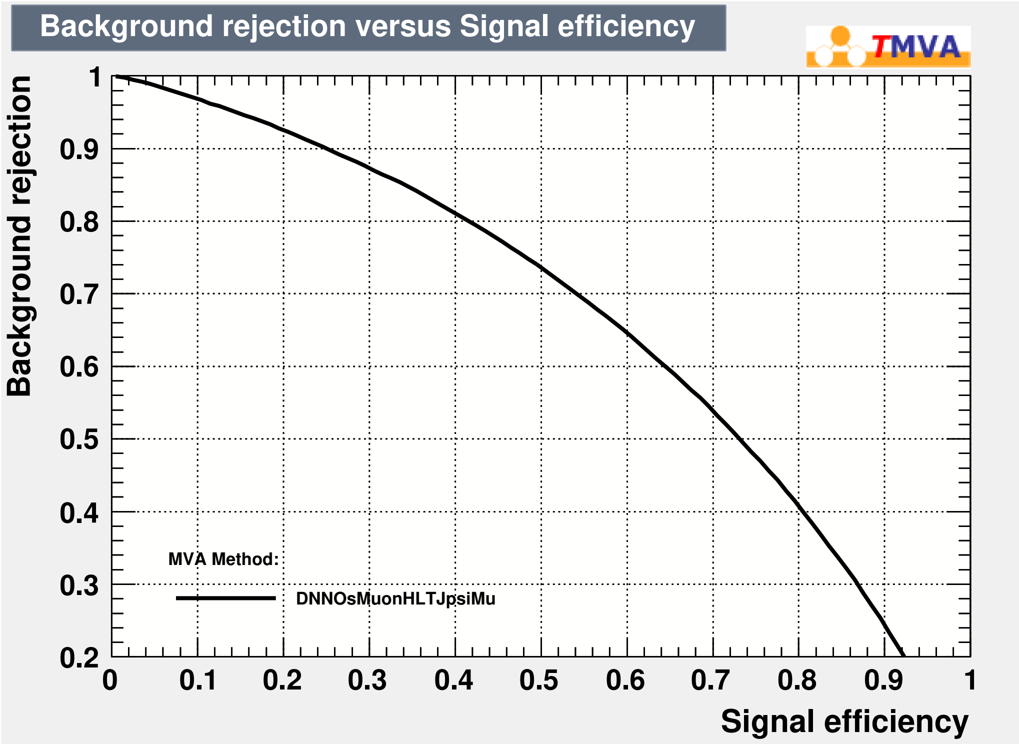
$$\omega_{\text{evt}} = 1 - s_{\text{dnn}} \,. \tag{8.37}$$

The predicted mistag probability can finally be used to re-weight the flavor inference and better exploit the data.

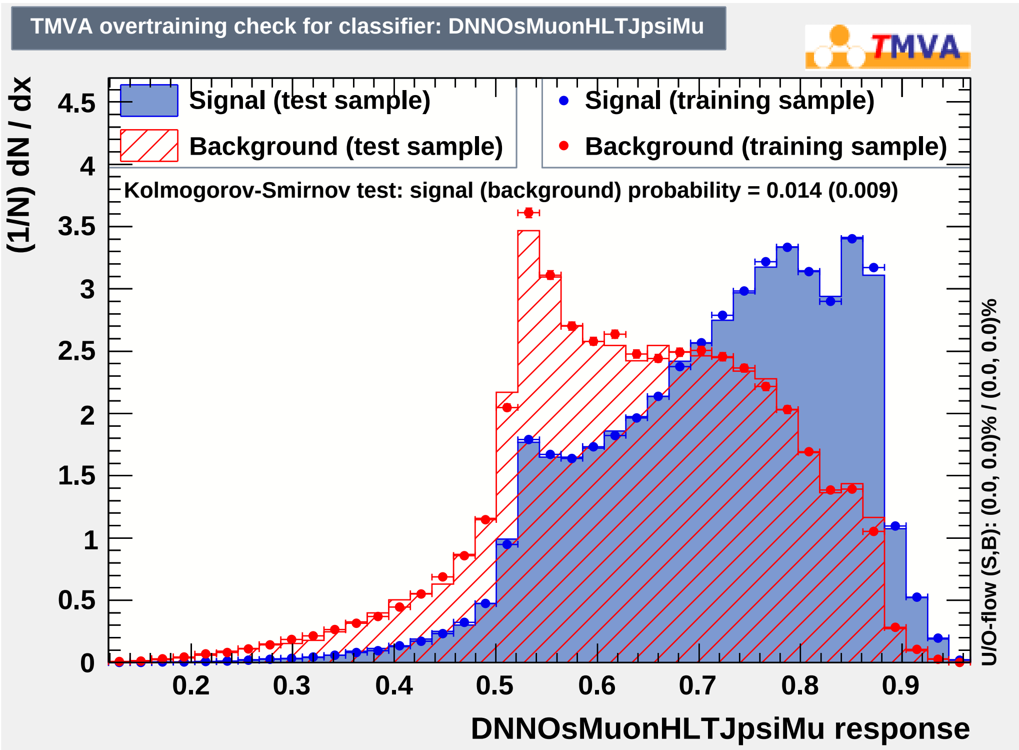
8.4.3.4 *Mistag probability calibration*

The muon tagger DNN is constructed specifically to directly predict the mistag probability. The softmax activation function ensures that the output score can be interpreted as a probability, while the cross-entropy loss function allows the model to be optimized to produce the true probability distribution of the input classes. However, this alone does not guarantee that the trained DNN is calibrated [155]. Moreover, the ultimate goal of the muon tagger DNN is to predict ω_{evt} on $B_s^0 \to J/\psi\,\phi$ data collected in pp collisions, not in the simulated events where the model was trained. Due to unavoidable simulation

32 The small fraction of events with $\omega_{\text{evt}} < 15\%$ are likely well-reconstructed muons with small transverse impact parameter, which is correlated to lower mixing probabilities, or muons from decays of charged b-hadrons, which do not oscillate.



(a) Muon tagger DNN ROC curve.



(b) Muon tagger DNN normalized output score.

Figure 8.11: Muon tagger DNN performance plots: ROC curve (top) and normalized DNN output score (bottom) for right- (blue) and wrong-tag (red) muons for the training (markers) and in the test (histograms) data sets. The TMVA plotting tool automatically calls class-1 (in this case right-tag) "Signal" and class-2 (in this case wrong-tag) "Background". However, this naming scheme is not related to any other definition of signal and background given in this dissertation.

mismodeling, it is not guaranteed the predicted mistag probability is equal to the one observed in data.

For these reasons, ω_{evt} is calibrated in data using self-tagging $B^+ \to J/\psi\, K^+$ decays, where the charge of the final-state kaon unambiguously determines the flavor of the B^+ meson[33]. This calibration strategy is possible because opposite-side tagging techniques are expected to be independent of the signal b-meson species. This assumption is tested by comparing the calibration in $B_s^0 \to J/\psi\, \phi$ and $B^+ \to J/\psi\, K^+$ simulated decays. Possible differences are treated as systematic uncertainties, as described in Section 10.8. The calibration is performed separately for the 2017 and 2018 data samples.

CANDIDATE SELECTION $B^+ \to J/\psi\, K^+$ candidates are selected with a procedure similar to the one used for the $B_s^0 \to J/\psi\, \phi$ decay. The same trigger and J/ψ candidate reconstruction requirements as for the B_s^0 signal sample are applied[34]. A charged particle with $p_T > 1.6$ GeV is assumed to be a kaon and combined in a kinematic fit with the dimuon pair to form the B^+ candidate. Candidates are selected if their invariant mass is in the range 5.10–5.65 GeV. A total of ≈1.4 M $B^+ \to J/\psi\, K^+$ candidates are selected, out of which about half are flavor tagged. Their $J/\psi\, K^+$ invariant mass distribution is shown in Fig. 8.12.

MISTAG MEASUREMENT For each tagged $B^+ \to J/\psi\, K^+$ candidate the muon tagger DNN is used to estimate the probability of it being mistagged. The B^+ events are then divided according to their predicted ω_{evt} into 100 equal bins and the number of right- and wrong-tag events are separately counted in each bin to measure the corresponding ω_{meas} as:

$$\omega^i_{meas} = \frac{N^i_{WT}}{N^i_{WT} + N^i_{RT}}\,, \tag{8.38}$$

where i is the bin index.

The number of B^+ signal events in each bin is obtained with a binned likelihood fit to the $J/\psi\, K^+$ invariant mass distribution in the 5.10–5.65 GeV range. The $B^+ \to J/\psi\, K^+$ signal distribution is modeled with two Gaussian distributions with the same mean, while the background is parametrized with a linear function and a complementary error function. An example of one of such fits is shown in Fig. 8.13. Uncertainties on the events counts are obtained from the fit and propagated to the bin measured mistag.

MISTAG CALIBRATION The predicted mistag probability is calibrated by fitting the measured mistag fractions with the linear function

$$\omega_{meas} = a + b \cdot \omega_{evt}\,. \tag{8.39}$$

33 Charge conjugation is implied.

34 The B_s^0 meson selection requirements are described in Section 7.2 and summarized in Tab. 7.3

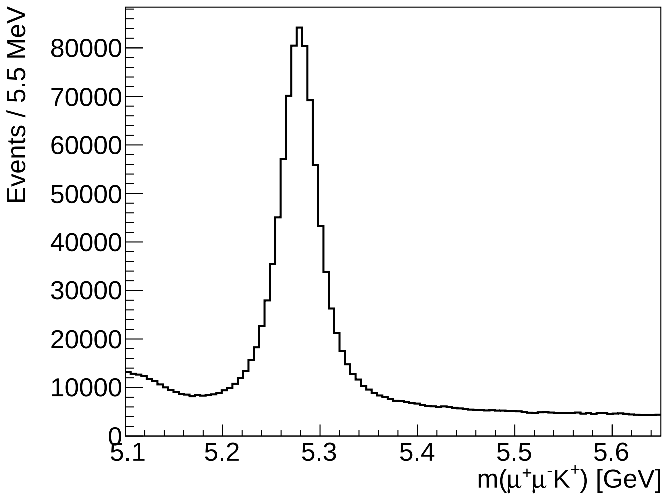


Figure 8.12: Invariant mass distribution for $B^+ \to J/\psi\,K^+$ candidates for mistag probability calibration.

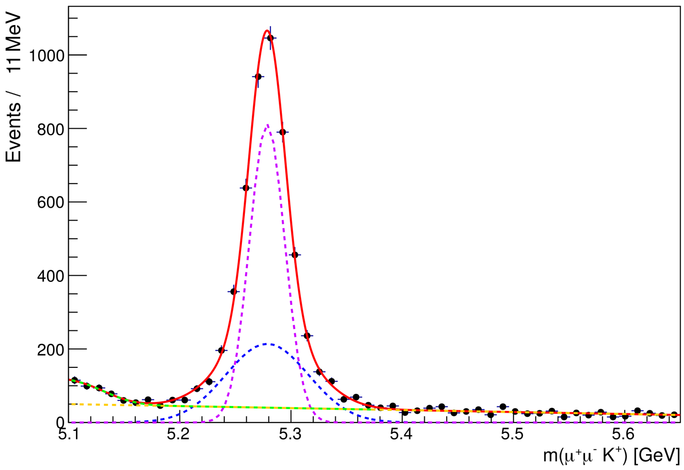


Figure 8.13: Example of fit to the $B^+ \to J/\psi\,K^+$ invariant mass distribution to extract the number of right-tag events in the predicted mistag probability bin $0.26 < \omega_{evt} < 0.27$. The blue and violet dashed lines represent the two Gaussian distributions that model the signal, while the orange and green line represent the linear and complementary error function distributions, respectively, that model the background.

The calibration fits for the 2017 and 2018 data samples are shown in Fig. 8.14, while calibration parameters are reported in Tab. 8.6. In both data samples, the measured mistag is predicted very accurately by ω_{evt}, with calibration parameters statistically compatible with a unit slope and zero offset.

The calibration of the muon tagger DNN output is also verified with the same procedure using an independent sample of simulated $B_s^0 \to J/\psi\,\phi$ and $B^+ \to J/\psi\,K^+$ events, where the reconstructed B_s^0 and B^+ mesons are matched to the generated ones to find their true flavor at production time. In this case, the number of events in each ω_{evt} bin is simply counted using the generation information and no fit to the invariant mass distribution is performed. The results of the calibration fits are reported in Tab. 8.7 and shown in Fig. 8.15.

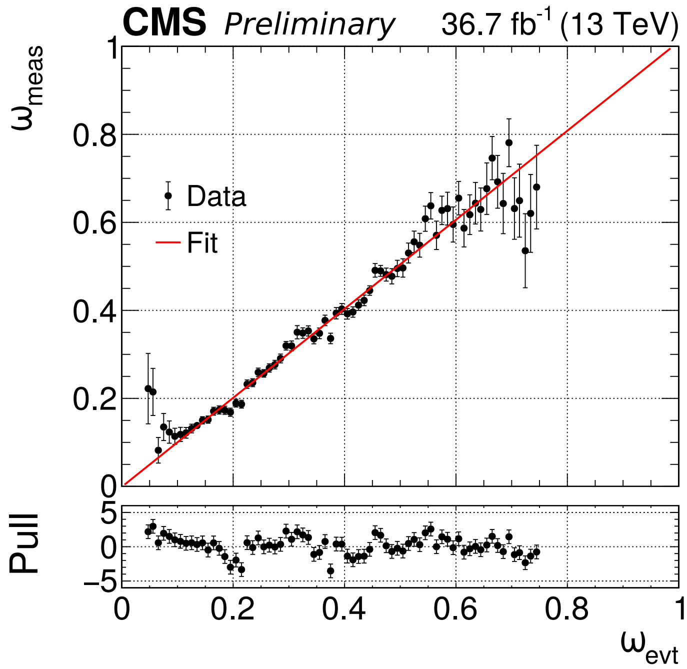
In general, the measured mistag probability is predicted very accurately by ω_{evt} over the entire range for all the examined simulated and data samples. In all cases more than 90% of the tagged events falls in the $\omega_{\text{evt}} = [0.1, 0.5]$ interval. Residual differences are well approximated by the linear calibration with slopes close to unity and offsets consistent with zero. The normalized χ^2/d.o.f. values for all fits are below two.

It is observed that at very low and very high values of ω_{evt} the measured mistag fractions consistently diverge from the predicted ones. This is likely due to the muon tagger DNN model being slightly overfitted at the edges of the mistag distribution where the training statistic is lower[35]. This is not considered an issue given that it concerns only a small fraction of events and is found to not affect the measurement results. It is observed that the calibration for the $B^+ \to J/\psi\,K^+$ (2017) samples is in slight tension with the others. This is likely due to the fact that this is the sample *further* from the training one, as both the data-taking conditions and decay channel are different. Differences in the mistag probability calibration between the $B_s^0 \to J/\psi\,\phi$ and $B^+ \to J/\psi\,K^+$ samples, as well as possible variations from linearity of the calibration function and the statistical uncertainties in the calibration fits, are considered as systematic uncertainties and described in Chapter 10.

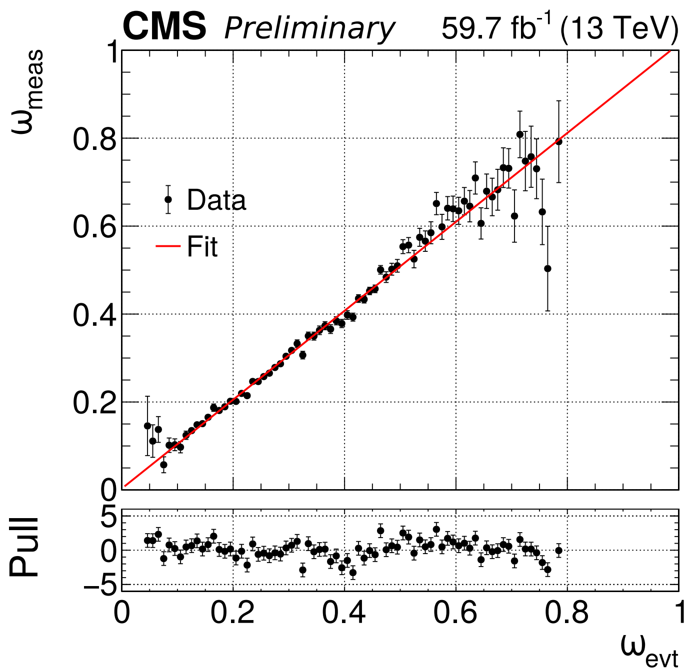
Table 8.6: Results of the calibration fit of the per-event mistag probability ω_{evt} on $B^+ \to J/\psi\,K^+$ decays. The fitted model is $\omega_{\text{meas}} = a + b \cdot \omega_{\text{evt}}$.

Data sample	a	b
2017	-0.0010 ± 0.0040	1.012 ± 0.013
2018	0.0031 ± 0.0031	1.011 ± 0.010

35 See Fig. 8.11.

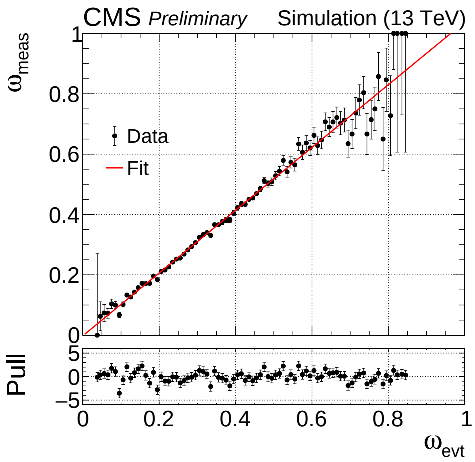


(a) 2017 data sample.

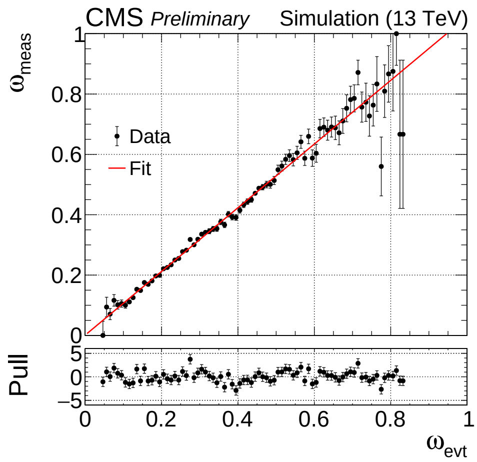


(b) 2018 data sample.

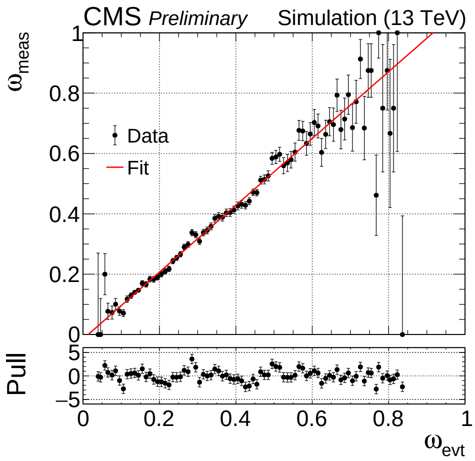
Figure 8.14: Results of the calibration of the per-event mistag probability ω_{evt} on $B^+ \to J/\psi\,K^+$ decays from the 2017 (top) and 2018 (bottom) data samples. The vertical bars represent the statistical uncertainties. The solid line shows the linear calibration fit to data (solid markers). The pull distributions between the data and the fit function in each bin are shown in the lower panels. Figures from Ref. [1].



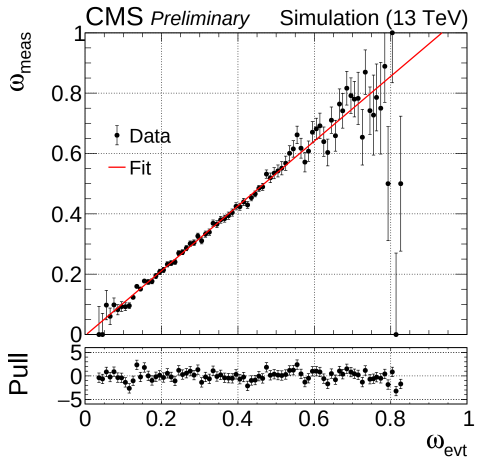
(a) $B_s^0 \to J/\psi\,\phi$ MC (2017 conditions).



(b) $B_s^0 \to J/\psi\,\phi$ MC (2018 conditions).



(c) $B^+ \to J/\psi\,K^+$ MC (2017 conditions).



(d) $B^+ \to J/\psi\,K^+$ MC (2018 conditions).

Figure 8.15: Results of the calibration of the per-event mistag probability ω_{evt} on simulated $B_s^0 \to J/\psi\,\phi$ (top) and $B^+ \to J/\psi\,K^+$ (bottom) decays in different data-taking conditions: 2017 (left) and 2018 (right). The vertical bars represent the statistical uncertainties. The solid line shows the linear calibration fit to data (solid markers). The pull distributions between the data and the fit function in each bin are shown in the lower panels.

Table 8.7: Results of the control calibration fits of the per-event mistag probability ω_{evt} on simulated $B_s^0 \to J/\psi\,\phi$ and $B^+ \to J/\psi\,K^+$ decays. The fitted model is $\omega_{meas} = a + b \cdot \omega_{evt}$.

MC sample	a	b
$B_s^0 \to J/\psi\,\phi$ (2017)	-0.0019 ± 0.0025	1.0042 ± 0.0082
$B_s^0 \to J/\psi\,\phi$ (2018)	0.0010 ± 0.0028	1.0552 ± 0.0090
$B^+ \to J/\psi\,K^+$ (2017)	-0.0146 ± 0.0038	1.107 ± 0.012
$B^+ \to J/\psi\,K^+$ (2018)	-0.0032 ± 0.0036	1.074 ± 0.012

8.4.4 *Final tagging performance*

The final tagging performance is evaluated on the same $B^+ \to J/\psi\,K^+$ data samples used for the mistag probability calibration. The total tagging power, when using per-event mistag probability, is equal to the sum of the tagging power of each individual tagged event, as[36]:

$$\widehat{P}_{tag} = \sum_{evt} P_{evt} = \frac{1}{N_{tot}} \sum_{evt}^{N_{tag}} (1 - 2\omega_{evt})^2 \,, \tag{8.40}$$

where $1/N_{tot}$ represents the tagging efficiency of a single event. However, this calculation is possible only in a pure sample of b-decays, in this case, $B^+ \to J/\psi\,K^+$. In a realistic data sample, where background is present, Eq. 8.40 is approximated by the sum of the tagging power of each individual ω_{evt} bin[37] as:

$$\widehat{P}_{tag} = \sum_{evt} P_{evt} \simeq \sum_{bin} P_{bin} = \sum_{bin} \varepsilon_{bin} (1 - 2\omega_{bin})^2 \,, \tag{8.41}$$

where $\omega_{bin} = N_{WT}^{bin}/N_{tag}^{bin}$ is the bin mistag fraction and $\varepsilon_{bin} = N_{tag}^{bin}/N_{tot}$ is the fraction of tagged B^+ candidates contained in the bin. All numbers refer only to $B^+ \to J/\psi\,K^+$ signal events. Accordingly, the total tagging efficiency can be evaluated as:

$$\varepsilon_{tag} \simeq \sum_{bin} \varepsilon_{bin} \,. \tag{8.42}$$

Simulated $B_s^0 \to J/\psi\,\phi$ and $B^+ \to J/\psi\,K^+$ events have been used to verify that the total tagging power evaluated with Eq. 8.40 and Eq. 8.41 are in agreement within statistical uncertainties.

The final tagging performance measured in the $B^+ \to J/\psi\,K^+$ data sample with Eq. 8.41 is reported in Tab. 8.8. For comparison, the tagging performance obtained using the average mistag rate $\omega_{tag} = N_{WT}/N_{tag}$, hence with no usage of

36 The "hat" indicate the *effective* tagging power (see Eq. 8.14).
37 The same binning used on the mistag probability calibration is used.

per-event information, is reported in Tab. 8.9. A tagging efficiency of ≈50% and a tagging power of ≈10% are achieved in both the 2017 and 2018 data samples. Per-event mistag probability is found to increase the muon tagging performance by ≈50% compared to the averaged approach, where only the muon charge was exploited[38]. The calibrated mistag probability distribution, as measured in the $B_s^0 \to J/\psi\,\phi$ data samples, is shown in Fig. 8.16.

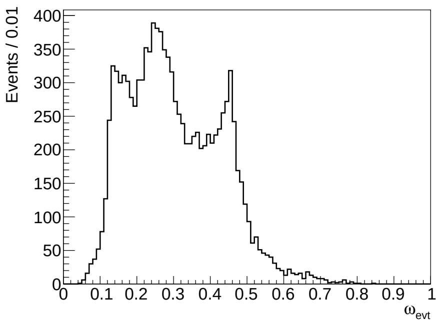
Table 8.8: Calibrated opposite-side muon tagger performance evaluated using $B^+ \to J/\psi\,K^+$ events in the 2017 and 2018 data samples using the per-event mistag probability approach. The tagging efficiency ε_{tag} and the total tagging power $\widehat{P}_{tag}$ are computed with Eq. 8.42 and Eq. 8.41, respectively. The effective mistag rate ($\widehat{\omega}_{tag}$) and effective tagging dilution ($\widehat{\mathcal{D}}_{tag}$) are defined in Eq. 8.16. The quoted uncertainties are statistical only. The performance obtained averaging the mistag rate are reported in Tab. 8.9.

Data sample	ε_{tag} [%]	$\widehat{\omega}_{tag}$ [%]	$\widehat{\mathcal{D}}_{tag}$	$\widehat{P}_{tag}$ [%]
2017	45.7 ± 0.1	27.1 ± 0.1	0.456 ± 0.002	9.6 ± 0.1
2018	50.9 ± 0.1	27.3 ± 0.1	0.454 ± 0.002	10.5 ± 0.1

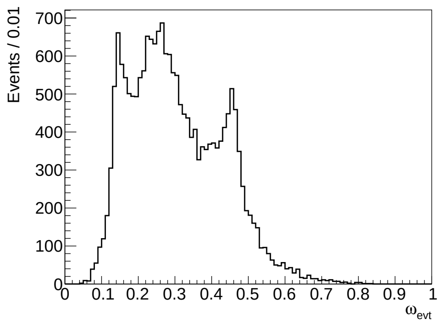
Table 8.9: Averaged opposite-side muon tagger performance evaluated using $B^+ \to J/\psi\,K^+$ events in the 2017 and 2018 data samples. The tagging efficiency ε_{tag}, mistag fraction ω_{tag}, and tagging power P_{tag} are computer with Eq. 8.2, Eq. 8.5, and Eq. 8.6, respectively. The quoted uncertainties are statistical only. The performance obtained with the per-event approach are reported in Tab. 8.8.

Data sample	ε_{tag} [%]	ω_{tag} [%]	$\mathcal{D}_{tag}$	P_{tag} [%]
2017 (no ω_{evt})	45.8 ± 0.1	30.7 ± 0.2	0.385 ± 0.004	6.8 ± 0.1
2018 (no ω_{evt})	51.0 ± 0.1	30.8 ± 0.2	0.384 ± 0.003	7.5 ± 0.1

38 Note that the tagging efficiency is not affected by the per-event approach.



(a) 2017 data sample.



(b) 2018 data sample.

Figure 8.16: Mistag probability distribution as measured in the selected $B_s^0 \to J/\psi\,\phi$ data samples. The calibration described in Section 8.4.3.4 is applied. No background subtraction is performed.

Part IV

THE MEASUREMENT

The CP-violating phase ϕ_s and the other physics parameters are finally estimated by performing a time-dependent and flavor-tagged angular analysis of the $B_s^0 \to J/\psi\,\phi$ decay rate on the selected data sample. The experimental data are described by probability density functions that model the distribution of the input observables for the signal and background components. Maximum-likelihood estimation methods are used to extract the parameter values that best describe the data sample. The analysis is completed by the evaluation of the systematic effects induced by the assumptions made in the fit model as well as in the analysis methods. Finally, the sensitivity on ϕ_s is maximized by statistically combining the measurement results with those obtained by the CMS collaboration at $\sqrt{s} = 8$ TeV. The following chapters are dedicated to the description of the fit model and the studies performed to estimate the systematic uncertainties. The resolution and acceptance effects that need to be taken into account to obtain unbiased results are also discussed, as well as the combination strategy.

9 FITTING PROCEDURE AND RESULTS

The physics parameters that govern the $B_s^0 \to J/\psi\,\phi$ decay are extracted with an *unbinned multidimensional extended maximum-likelihood* (UML) fit to the combined 2017+2018 data sample. In order to properly model the decay rate described in Eq. 3.19, the following observables are used as input: the three decay angles $\Theta = (\theta_T, \psi_T, \varphi_T)$ describing the event topology in the transversity basis, the flavor tag decision ξ_{tag}[1], the calibrated mistag probability ω_{evt}, the proper decay length of the B_s^0 candidate ct, and its uncertainty σ_{ct}. The B_s^0 candidate invariant mass $m_{B_s^0}$ is also used to discriminate signal from background.

The mistag probability and proper decay length uncertainty are included to take into account the dilution factors described in Section 9.2. From the UML fit, the following physics parameters of primary interest are determined: ϕ_s, $\Delta\Gamma_s$, Γ_s, Δm_s, $|\lambda|$ [II], the squares of amplitudes $|A_\perp|^2$, $|A_0|^2$, $|A_S|^2$, and the strong phases $\delta_\parallel$, $\delta_\perp$, and $\delta_{S\perp}$, where $\delta_{S\perp}$ is defined as the difference $\delta_S - \delta_\perp$[III].

ASSUMPTIONS AND CONVENTIONS With respect to the decay rate described by Eq. 3.19 the number of free parameters is reduced by two. The value of δ_0 is set to zero[4], since only phase differences are physically meaningful, and the normalization relationship $|A_0|^2 + |A_\perp|^2 + \left|A_\parallel\right|^2 = 1$ is used in redefining $\left|A_\parallel\right|^2 = 1 - |A_0|^2 - |A_\perp|^2$. The value of $\Delta\Gamma_s$ is also constrained to be positive, based on the LHCb measurement in Ref. [52]. The fit input observables, the main physics parameters extracted and the conventions used are summarized in Tab. 9.1.

This chapter describes the construction of the final fit model and presents the results obtained on the combined 2017 and 2018 data samples. Particular attention is given to the various resolution and acceptance effects that have been taken into account in order to obtain unbiased results. A brief introduction on maximum likelihood estimation is also given.

1 The flavor tagging decision ξ_{tag} is defined in Eq. 8.1.
II These parameters are defined in Equations 3.10, 2.29, 2.27, 2.28, and 2.41, respectively.
III The transversity basis is described in Section 3.2.1.
4 This follows a general convention first used in Refs. [56, 58].

Table 9.1: Summary of the UML fit input observables, main physics parameters to be determined and used assumption. The rows of this table carry no meaning.

Input observables	Fitted parameters	Constraints
θ_T, ψ_T, φ_T	ϕ_s, $\Delta\Gamma_s$, Γ_s, Δm_s, $\lvert\lambda\rvert$	$\Delta\Gamma_s > 0$
ct, σ_{ct}, $m_{B^0_s}$	$\lvert A_\perp\rvert^2$, $\lvert A_0\rvert^2$, $\lvert A_S\rvert^2$	$\lvert A_\parallel\rvert^2 = 1 - \lvert A_0\rvert^2 - \lvert A_\perp\rvert^2$
ξ_{tag}, ω_{evt}	$\delta_\parallel$, $\delta_\perp$, $\delta_{S\perp}$	$\delta_0 = 0$, $\delta_{S\perp} = \delta_S - \delta_\perp$

9.1 MAXIMUM LIKELIHOOD ESTIMATION

The experimental data are described by a *probability density function* (PDF), which describes the distributions of all the observables. A PDF is a function that represents a continuous probability distribution and can be used to infer the probability of a continuous random variable x of falling in a given range $[a, b]$ of the phase space as:

$$\int_a^b f(x)\,dx = P(a < x < b)\,. \tag{9.1}$$

Unlike the simpler probability functions, the output of a PDF is not a probability value. By definition, probability density functions can only assume positive values and are normalized to unity.

The *likelihood function* measures the goodness of fit of a statistical model to a sample of data given the values of some parameters, typically unknown. It is constructed from the joint probability of the sample, as a function of the unknown parameters[5]. The likelihood of a set of parameters $\vec{\theta}$, given a set of n observations $\vec{x} = (x_1, \ldots, x_n)$ that are distributed following a PDF $f(\vec{x}, \vec{\theta})$, is defined as the product of the PDF values for each event as:

$$\mathcal{L}(\vec{\theta}\,|\,\vec{x}) = \prod_{i=1}^{n} f(x_i, \vec{\theta})\,. \tag{9.2}$$

The likelihood function is maximized by the combination of model parameters values for which the joint probability of all the observables is maximal [165][6]. The approach of maximizing $\mathcal{L}$ to estimate the parameters $\vec{\theta}$ is called *maximum likelihood estimation* (MLE):

$$\widehat{\theta} = \arg\max \mathcal{L}(\vec{\theta}\,|\,\vec{x}) \quad \leftrightarrow \quad \left.\frac{d\mathcal{L}(\vec{\theta}\,|\,\vec{x})}{d\vec{\theta}}\right|_{\widehat{\theta}} = 0\,. \tag{9.3}$$

5 Since the likelihood is a function of the model parameters, and not of the observables, it is *not* a PDF and it is *not* normalized.

6 It is not guaranteed that the likelihood function has a maximum.

If the product in Eq. 9.2 is performed over all the events in the data sample, not collected in histograms, the MLE is referred to as *unbinned*. Unbinned MLEs are more precise but also computationally heavier than their binned counterpart.

The common practice in MLE is to minimize the negative log-likelihood (NLL) function, which is defined as:

$$-\ln \mathcal{L}(\vec{\theta}\,|\,\vec{x}) = -\sum_{i=1}^{n} \ln f(x_i, \vec{\theta})\,. \tag{9.4}$$

Because the natural logarithm is an monotonic function, maximizing the log-likelihood is equivalent to maximizing the likelihood, with the advantages of dealing with sums instead of products in the derivative in Eq. 9.3. The problem is ultimately transformed into a minimization, as optimizers in statistical packages usually work by minimizing the result of a function, instead of maximizing it.

In this work NLL functions are constructed with the RooFit library [143] and minimized with Minuit minimization framework [166].

9.1.1 *Extended term*

Regular ML estimation uses only shape information, as the PDF of each event is normalized. If the number of events in the sample is of interest, the *extended* maximum likelihood (EML) method [167] can be used to vary the normalization of the PDF as well, assuming that the number of events n is itself a random variable governed by Poisson statistics.

In this assumption the likelihood and NLL functions become:

$$\mathcal{L}(\nu, \vec{\theta}\,|\,\vec{x}) = \frac{e^{-\nu}\nu^{n}}{n!}\mathcal{L}(\vec{\theta}\,|\,\vec{x}) = \frac{e^{-\nu}}{n!}\prod_{i=1}^{n} \nu\, f(x_i, \vec{\theta})\,, \tag{9.5}$$

$$-\ln \mathcal{L}(\nu, \vec{\theta}\,|\,\vec{x}) = \nu - \sum_{i=1}^{n} \ln\!\big[\nu\, f(x_i, \vec{\theta})\big] + \text{const.}\,, \tag{9.6}$$

where ν is the expected number of events. If ν is independent of $\vec{\theta}$, its estimate is equal to the number of observed events:

$$\frac{d\big[-\ln \mathcal{L}(\vec{\theta}\,|\,\vec{x})\big]}{d\nu} = 1 - \sum_{i=1}^{n} \frac{1}{\nu} = 0 \quad \rightarrow \quad \widehat{\nu} = n\,. \tag{9.7}$$

9.1.2 *Uncertainties evaluation*

The uncertainties on the best values of the model parameters can be estimated by expanding the NLL around its minimum $\vec{\theta} = \widehat{\theta}$, where the first derivative is zero. Restricting, for the sake of simplicity, the discussion to a model with a single parameter (i. e. $\vec{\theta} = \theta$), the NLL can be described as[7]:

$$-\ln \mathcal{L}(\theta) = f(\theta) \approx f(\widehat{\theta}) + \underbrace{\left.\frac{df(\theta)}{d\theta}\right|_{\widehat{\theta}} (\theta - \widehat{\theta})}_{=0} + \frac{1}{2} \left.\frac{d^2 f(\theta)}{d\theta^2}\right|_{\widehat{\theta}} (\theta - \widehat{\theta})^2 + \mathcal{O}\Big((\theta - \widehat{\theta})^3\Big) . \tag{9.8}$$

Using Eq. 9.8, the likelihood function at the minimum can be evaluated as:

$$\mathcal{L}(\theta) = e^{-f(\theta)} \approx e^{-f(\widehat{\theta})} \cdot e^{-\frac{1}{2} \left.\frac{d^2 f(\theta)}{d\theta^2}\right|_{\widehat{\theta}} (\theta - \widehat{\theta})^2} \cdot e^{-\mathcal{O}((\theta - \widehat{\theta})^3)} . \tag{9.9}$$

If the shape of the likelihood is *sufficiently* Gaussian around its maximum, that is if the higher order terms can be neglected, Eq. 9.9 can be expressed as:

$$\mathcal{L}(\theta) \approx \mathcal{L}_{\max} \cdot e^{-\frac{(\theta - \widehat{\theta})^2}{2\sigma_{\theta}^2}} , \tag{9.10}$$

where σ_{θ}^2 is the variance of θ, estimated as:

$$\sigma_{\theta}^2 = \left(\left.\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta^2}\right|_{\widehat{\theta}} \right)^{-1} , \tag{9.11}$$

and $\mathcal{L}_{\max} = e^{\ln \mathcal{L}(\widehat{\theta})} = \mathcal{L}(\widehat{\theta})$.

From Eq. 9.10 it immediately follows that the NLL near its minimum is approximated by the parabolic expression:

$$-\ln \mathcal{L}(\theta) \approx -\ln \mathcal{L}_{\max} + \frac{(\theta - \widehat{\theta})^2}{2\sigma_{\theta}^2} , \tag{9.12}$$

which satisfies the following important relationships:

$$-\ln \mathcal{L}(\widehat{\theta} \pm \sigma_{\theta}) = -\ln \mathcal{L}_{\max} + 0.5 , \tag{9.13}$$

$$-\ln \mathcal{L}(\widehat{\theta} \pm 2\sigma_{\theta}) = -\ln \mathcal{L}_{\max} + 2.0 . \tag{9.14}$$

If the NLL is not parabolic, then $\widehat{\theta} \pm \sigma_{\theta}$ is not a good estimate for the limits of the 68% (1σ) confidence interval (CI). In these cases the standard approach is to construct the confidence intervals with the relationships reported in Equations 9.13 and 9.14, as illustrated in Fig. 9.1[8]. This method typically produces asymmetric errors.

MULTI-PARAMETERS MODELS In the case of a model with multiple parameters, the approach is the same as described above. The NLL is expanded around

7 The dependence on the observables x is dropped from the following equations for simplicity.
8 This construction does not guarantee adequate coverage.

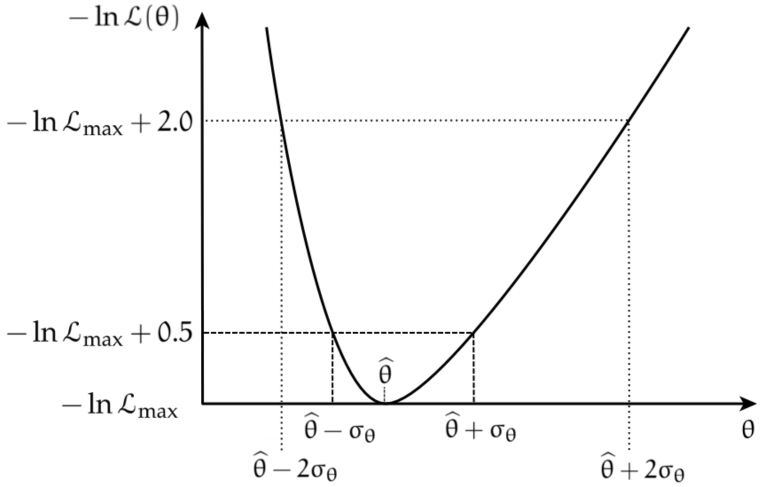


Figure 9.1: Confidence intervals (CI) construction for ML estimators with non-parabolic NLL function. The construction procedure for the 68% and 95% CIs is indicated by the dashed and dotted lines, respectively.

its minimum and the second derivatives are used to estimate the uncertainty on the parameters and the covariance between them.

In practical terms, an Hessian matrix H is constructed from the second derivatives of the log-likelihood function, as[9]:

$$H_{ij} = \left. \frac{\partial^2 \ln \mathcal{L}(\vec{\theta})}{\partial\theta_i \partial\theta_j} \right|_{\hat{\theta}}, \tag{9.15}$$

which is then used to calculate both the parameters uncertainty and the covariance, as:

$$\begin{aligned} \sigma_{\theta_i} &= \sqrt{(H^{-1})_{ii}}, \\ \mathrm{cov}(\theta_i, \theta_j) &= \left(H^{-1}\right)_{ij}, \end{aligned} \tag{9.16}$$

where (H^{-1}) is also called "covariance matrix".

As in the monodimensional case, this method is accurate only if the NLL is parabolic near the minimum. If not, the method described above can be extended to define the boundaries for the 68% or the 95% confidence intervals for each parameter[10].

In this work, the uncertainties on the fit parameters and the correlations between them are evaluated with the MINOS and HESSE methods, respectively, as implemented by the MINUIT fitting framework [168].

9 H_{ij} indicates the elements of the Hessian matrix H.
10 More details on the exact procedures are contained in the MINUIT documentation [168].

9.2 EXPERIMENTAL EFFECTS

An accurate parameterization of the transversity angles, the proper decay length, and the flavor information is not sufficient to properly fit the $B_s^0 \to J/\psi\,\phi$ decay rate in data. In order to not only obtain unbiased results but also to improve the performance, any distortion due to the reconstruction and selection procedures needs also to be taken into account. These experimental factors can be divided into *resolution* and *acceptance* effects.

Resolution distortions are caused by the finite precision of measurements of the various observables, typically leading to a dilution of the observed CP asymmetry and reducing the statistical power of the analysis. Lifetime resolution and mistagged events are by far the dominant effects, whereas the resolution on the angular variables can be neglected as it is observed to be smaller than the angular structures to resolve[11].

Acceptance effects are caused by the reconstruction and selection of the candidates and alter the shape of the variable distributions. In order to avoid bias in the results, they are taken into account in the fit model by dedicated efficiency functions.

9.2.1 *Flavor tagging dilution*

Mistagged events directly dilute the observed CP asymmetry by a factor

$$\mathcal{D}_{\text{tag}} = (1 - 2\omega_{\text{tag}})\,.$$

The origin and estimation of the tagging dilution are extensively described in Chapter 8. This effect is incorporated into the fit model by modifying the time-dependent functions $\mathcal{O}_i$ of the differential decay rate model[12] to include the flavor tagging information, as:

$$\begin{aligned} O_i(\alpha, t) = N_i e^{-\Gamma_s t} \Bigg[& a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ & + c_i \xi_{\text{tag}} (1 - 2\omega_{\text{evt}}) \cos(\Delta m_s t) \\ & + d_i \xi_{\text{tag}} (1 - 2\omega_{\text{evt}}) \sin(\Delta m_s t) \Bigg]\,, \end{aligned} \tag{9.17}$$

where the tagging decision ξ_{tag} and the per-event dilution $\mathcal{D}_{\text{evt}} = (1 - 2\omega_{\text{evt}})$ are applied to each of the c_i and d_i terms[13].

---

11 The resolution on the three transversity angles is measured, using simulated events, to be ≈30 mrad.

12 The $\mathcal{O}_i$ functions have been first defined in Eq. 3.20.

13 The c_i and d_i terms change sign between $B_s^0 \to J/\psi\,\phi$ and $\overline{B}_s^0 \to J/\psi\,\phi$ decays, as can be seen from Eq. 3.14.

9.2.2 *Decay length resolution*

In order to properly resolve the flavor oscillation of the B_s^0–$\overline{B}_s^0$ system, the proper decay time[14] resolution needs to be significantly smaller than the oscillation period $T = 2\pi/\Delta m_s \sim 350$ fs.

Being able to properly observe the B_s^0 oscillations is of primary importance in this analysis, as the part of the decay rate most sensitive to ϕ_s is the asymmetry between the B_s^0 and $\overline{B}_s^0$ decays, which is roughly given by[15]:

$$a_{CP}(t) \propto \sin\phi_s \sin(\Delta m_s t)\,. \tag{9.18}$$

The impact of the decay time resolution (hereafter indicated with δ_t) can be estimates assuming a Gaussian resolution function, as [169]:

$$a_{CP}^{meas}(t) = \int_{-\infty}^{+\infty} a_{CP}(t') \frac{1}{\sqrt{2\pi}\delta_t} e^{-\frac{(t-t')}{2\delta_t^2}}\, dt' = e^{-\frac{\delta_t^2 \Delta m_s^2}{2}} a_{CP}(t)\,, \tag{9.19}$$

From Eq. 9.19 follows that the decay time resolution effectively damps the flavor oscillations amplitude, and by consequence the observed CP asymmetry, by a dilution factor equal to:

$$\mathcal{D}_{time} = e^{-\delta_t^2 \frac{\Delta m_s^2}{2}}\,. \tag{9.20}$$

In analogy to the tagging dilution, the sensitivity on ϕ_s is directly proportional to $\mathcal{D}_{time}$[XVI].

The decay time resolution distribution in the selected $B_s^0 \to J/\psi\,\phi$ data sample is shown in Fig. 9.2. The typical δ_t values are of the order of 50∼80 fs, which correspond to dilution factors of about 0.4∼0.7. However, since the decay time resolution is evaluated on a per-event basis, the *effective* dilution factor is likely to be higher, in analogy with the per-event mistag probability reasoning presented in Section 8.1.1.

ANALYSIS SENSITIVITY Taking into consideration all dilution factors, and neglecting background, the overall sensitivity on ϕ_s can be expressed as:

$$S_{\phi_s} \propto \mathcal{D}_{tag}\mathcal{D}_{time}\sqrt{\varepsilon_{tag} N_{B_s^0}}\,, \tag{9.21}$$

which can be used to compare different experimental strategies. As an example, the ratio of S_{ϕ_s} as evaluated for this work to the one evaluated for the previous CMS result [64] is roughly equal to $S_{\phi_s}^{13\text{ TeV}}/S_{\phi_s}^{8\text{ TeV}} \approx 3.5$.[XVII]

14 The proper decay length ct and the proper decay time t differ only by a factor c. As such, sometimes these two terms are interchanged depending on what unit of measure is more suitable for the discussion.

15 See Eq. 3.13.

XVI This is only true as a first approximation, as some terms of the decay rate are not directly related to the B_s^0 mixing. These terms are however less sensitive to ϕ_s and can be neglected in this discussion.

XVII This work: $\mathcal{D}_{tag} \approx 0.46$, $\mathcal{D}_{time} \approx 0.55$, $\varepsilon_{tag} \approx 0.5$. Ref. [64]: $\mathcal{D}_{tag} \approx 0.40$, $\mathcal{D}_{time} \approx 0.45$, $\varepsilon_{tag} \approx 0.08$. The number of B_s^0 candidates is assumed equal, which is the case in a first approximation.

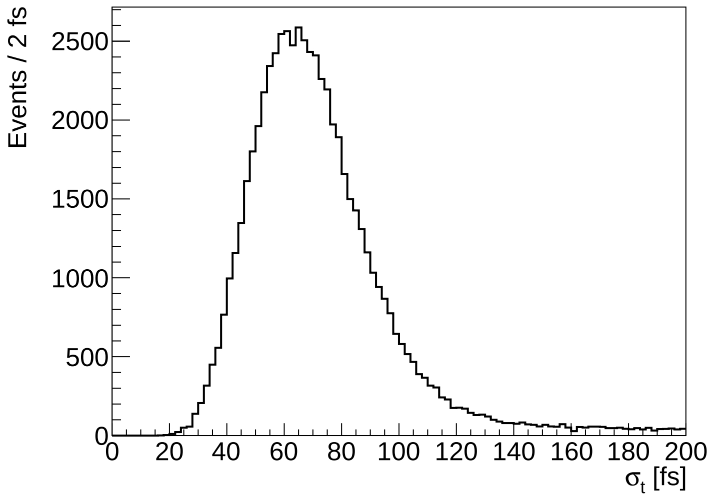


Figure 9.2: Decay time resolution distribution in the selected $B_s^0 \to J/\psi\,\phi$ data sample, evaluated as $\delta_t = \frac{\kappa_{ct}\sigma_{ct}}{c}$. The κ_{ct} scaling factor, roughly equal to ≈ 1.2, is discussed in the text in Section 9.2.2.1.

9.2.2.1 *Resolution estimation*

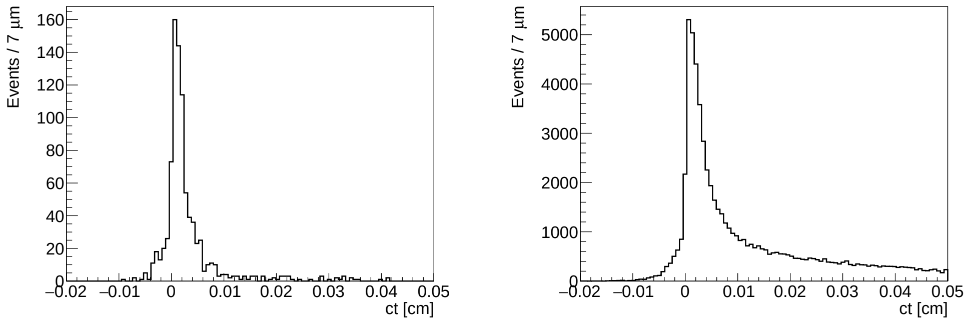
The proper decay length uncertainty, measured in each event is (Eq. 7.5):

$$\sigma_{ct} = ct\sqrt{\left(\frac{\sigma_{L_{xy}}}{L_{xy}}\right)^2 + \left(\frac{\sigma_{p_T}}{p_T}\right)^2}. \tag{9.22}$$

As explained in Section 7.2.2, the dominant part in Eq. 9.22 is the term related to the uncertainty in measuring the decay length L_{xy}, which in turn is dominated by the uncertainty on the measurement of the position of the secondary vertex. The measured σ_{ct} distribution in the selected $B_s^0 \to J/\psi\,\phi$ data sample is shown in Fig. 7.4, with average values of the order of 15–20 μm. In this work, the measured ct uncertainty is used to estimate the ct resolution on a per-event basis. It is worth remembering that the ct uncertainty and the ct resolution are not the same quantity. The former is a per-event estimate of the uncertainty in measuring ct, while the latter is the main parameter of the resolution function (Eq. 9.19) that needs to be convolved with the original ct distribution to obtain the measured one[18].

PROPER DECAY LENGTH UNCERTAINTY CORRECTION To account for any underestimate or overestimate of the proper decay length uncertainties, a scaling factor κ_{ct} is introduced as constant multiplying σ_{ct} in the fit model. The scaling factor can be determined by comparing event-by-event the measured σ_{ct} with

18 This reasoning assumes a flat efficiency.



(a) $J/\psi \to \mu^+\mu^-$ simulated sample.

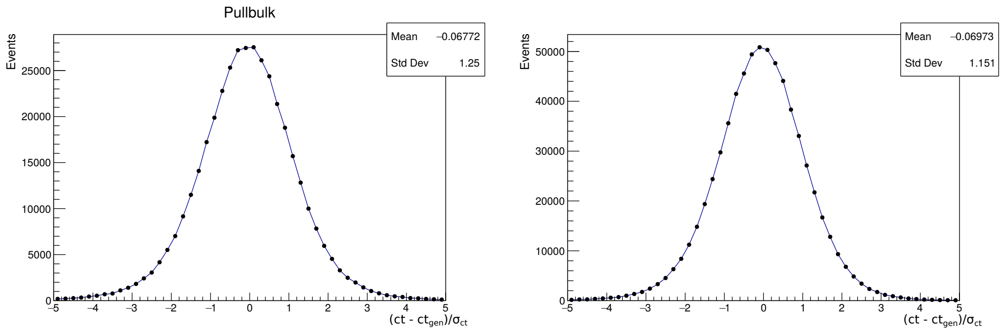
(b) $B_s^0 \to J/\psi\,\phi$ data sample.

Figure 9.3: Proper decay length distribution around $ct \sim 0$ cm of $B_s^0 \to J/\psi\,\phi$ candidates as reconstructed in $J/\psi \to \mu^+\mu^-$ simulated events (left) and in the data sample (right). In both cases the $B_s^0 \to J/\psi\,\phi$ analysis trigger and offline reconstruction steps are applied, with the exception of the requirement on ct.

the difference between the measured and *true* proper decay length. Typically this can be achieved either in simulations, where the true ct is known from MC truth, or in data exploiting prompt J/ψ mesons that have a lifetime comparable with zero.

In the context of this analysis, κ_{ct} cannot be reliably determined in data, as the trigger selection strongly suppresses the prompt background. The requirement of a third tagging muon, combined with the other selection requirements, is found to systematically reduce any background source not related to b-hadrons from $b\bar{b}$ pairs. This has been cross-checked in simulations, by applying the trigger and offline reconstruction in simulated $J/\psi \to \mu^+\mu^-$ events. Figure 9.3 shows the ct distribution for reconstructed $B_s^0 \to J/\psi\,\phi \to \mu^+\mu^-K^+K^-$ candidates in $J/\psi \to \mu^+\mu^-$ simulated events[19] and in the data. Both samples are constructed applying the analysis reconstruction procedure described in Chapter 7, except for the requirement on ct [XX]. In both cases, the observed distribution is found to be compatible with a sum of exponential distributions convolved with Gaussian resolution functions. The prompt component is found to be too suppressed to be properly modeled. For this reason, only $B_s^0 \to J/\psi\,\phi$ simulated events are used to estimate κ_{ct}, with the Gaussian resolution functions fitted in data (Fig. 9.3b) only used to roughly cross-check the obtained values and to evaluate a related systematic uncertainty[21].

19 See Section 7.3 for more details on simulated samples.
XX Selection requirement on the proper decay length: $ct > 70$ µm.
21 See Section 10.6.



(a) 2017 simulations. (b) 2018 simulations.

Figure 9.4: Proper decay length pull distribution as measured in simulated $B_s^0 \to J/\psi\,\phi$ events for 2017 (left) and 2018 (right) data-taking conditions. The entries of the distributions are calculated as $(ct - ct_{gen})/\sigma_{ct}$, where ct_{gen} is the generated ct value of the selected B_s^0 meson. For each data-taking period, the scaling factor κ_{ct} is defined as the RMS of the resulting distribution.

The scaling factor κ_{ct} is estimated separately for the 2017 and 2018 data-taking conditions as the RMS of proper decay length *pull* distribution:

$$\kappa_{ct} = \mathrm{RMS}\left(\frac{ct - ct_{gen}}{\sigma_{ct}}\right). \tag{9.23}$$

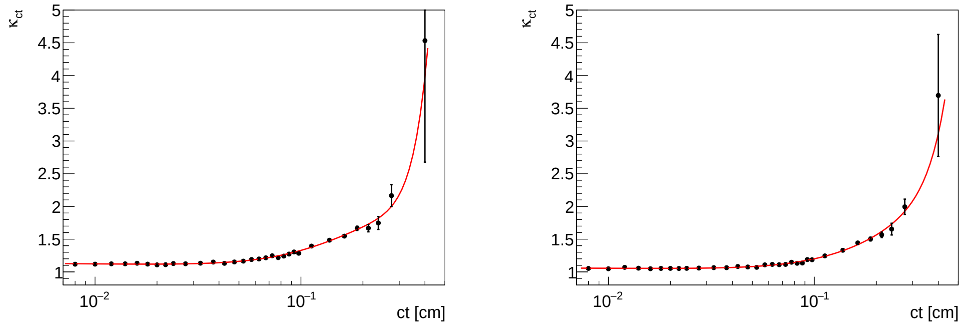
The resulting distributions are shown in Fig. 9.4. The scaling factor κ_{ct} is measured to be 1.25 and 1.15 for the 2017 and 2018 data-taking conditions, respectively, while the central values are compatible with zero within 0.05. These values are roughly compatible within ±10% with the values observed in data.

The resolution scaling factor is also evaluated as a function of the proper decay length. The procedure is the same as above but performed in several bins of ct. The results of this study are shown in Fig. 9.5. The value of κ_{ct} is found to increase with increasing values of ct, especially after $ct \approx 0.2$ cm. In the final fit model, the resolution scaling factors for the two data-taking periods are fixed to their averages, as the majority of events are found to be well described by such values. The effect of the κ_{ct} dependence on ct is treated as a systematic uncertainty, as described in Section 10.6.

PROPER DECAY LENGTH RESOLUTION FIT MODEL The proper decay length resolution is included in the fit model through a Gaussian resolution function which makes use of the scaled per-event σ_{ct} as:

$$G(ct, \sigma_{ct}) = e^{-\frac{1}{2}\left(\frac{ct}{\delta_{ct}}\right)^2} \approx e^{-\frac{1}{2}\left(\frac{ct}{\kappa_{ct}\sigma_{ct}}\right)^2}. \tag{9.24}$$

In the final fit Eq. 9.24 is then convolved with the $B_s^0 \to J/\psi\,\phi$ decay rate, effectively smearing the ct distribution.



(a) 2017 simulations. (b) 2018 simulations.

Figure 9.5: Dependence on ct of the resolution scaling factor κ_{ct}, as measured in simulated $B_s^0 \to J/\psi\,\phi$ samples. The markers represent the κ_{ct} value as evaluated in each ct bin, while the red line represents a 4-th order polynomial fit to the experimental points.

The σ_{ct} PDF is also included in the fit model. It is parametrized as the sum of two *gamma distributions*, each defined as:

$$P_\Gamma = \frac{(\sigma_{ct} - \mu)^{\gamma-1}\, e^{\frac{-(\sigma_{ct}-\mu)}{\beta}}}{\Gamma(\gamma)\,\beta^{\gamma}}\,, \tag{9.25}$$

where $\Gamma(\gamma)$ is the gamma function $\Gamma(\gamma) = (\gamma - 1)!$ and μ, γ and β are three shape parameters[22]. The σ_{ct} distribution obtained in $B_s^0 \to J/\psi\,\phi$ simulated events is shown in Fig. 9.6 with the fit results overlayed. In both distributions, *structures* are clearly visible in the pulls. This is interpreted as residual differences between the observed distributions and the fit model, enhanced by the very high statistics of the MC samples. At the level of statistics available in the data, these structures should not be visible[23] and are not considered an issue.

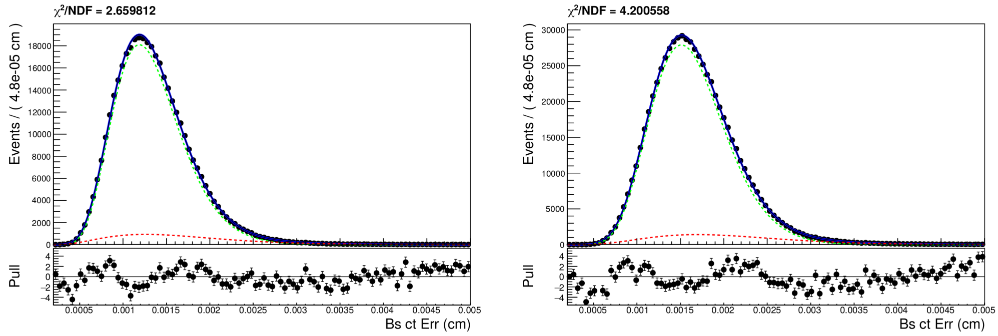
9.2.3 *Proper decay length efficiency*

The efficiency in reconstructing a $B_s^0 \to J/\psi\,\phi$ candidates depends on the decay length of the B_s^0 meson. Especially at large ct values, the efficiency in reconstructing secondary decay vertices is expected to decrease [120, 170].

The proper decay length efficiency ("ct efficiency") is estimated using simulated $B_s^0 \to J/\psi\,\phi$ events, as:

$$\varepsilon(ct) \propto \frac{N(ct)}{N(ct_{gen}) \otimes G(ct, \sigma_{ct})}\,, \tag{9.26}$$

22 These shape parameters are not related to other quantities indicated with the same symbols. This applies to all PDF described in this chapter.
23 See Fig. 9.16c later in the chapter.



(a) 2017 simulations.

(b) 2018 simulations.

Figure 9.6: Proper decay length uncertainty distribution as measured in 2017 (left) and 2018 (right) $B_s^0 \to J/\psi\,\phi$ simulated events. The two colored dashed lines in each plot represent the two gamma distributions.

where

- $N(ct)$ is the ct distribution as *measured* in $B_s^0 \to J/\psi\,\phi$ simulated samples[24] with $\Delta\Gamma_s = 0$,
- $N(ct_{gen})$ is the ct distribution as *generated* in $B_s^0 \to J/\psi\,\phi$ simulated sample where no geometrical acceptance filter is applied at generation level[25],
- $G(ct, \sigma_{ct})$ is the Gaussian resolution function defined in Eq. 9.24.

The proper decay length efficiency is estimated in several bins of ct separately for the 2017 and 2018 data samples.

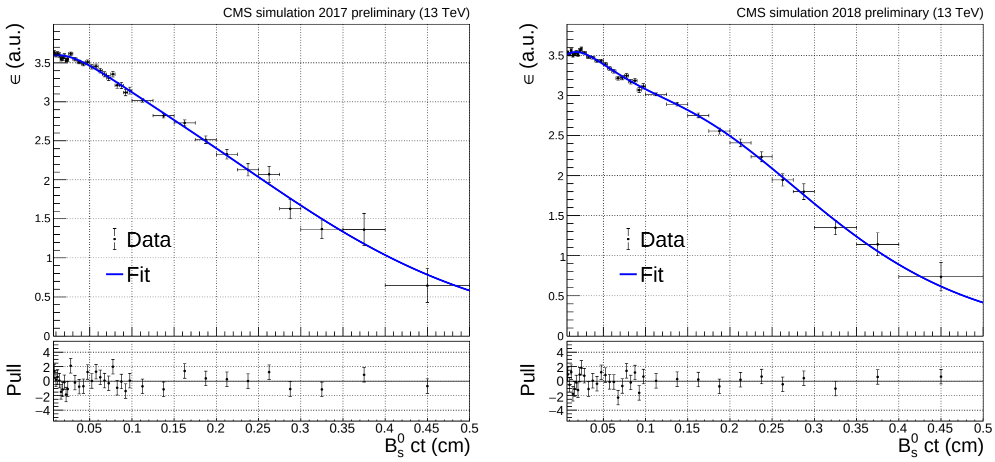
The acceptance histograms are then parametrized in the 0.007–0.5 cm ct range with the following function:

$$\varepsilon(ct) = a e^{-b\,ct} \cdot \sum_{i=1}^{4} c_i T_i(ct) \tag{9.27}$$

where T_n are the Chebyshev polynomials of the first kind and a, b, and c_i are a set of parameters to be optimized. Higher order polynomials have been tested and found to not improve the parametrization.

The efficiency histograms and the obtained efficiency functions are shown in Fig. 9.7. As expected the proper decay length efficiency decrease with increasing values of ct.

24 The trigger selection and offline reconstruction procedure is applied.
25 This is equivalent of using a $e^{-\Gamma_s t}$ distribution.



(a) 2017 data-taking conditions.

(b) 2018 data-taking conditions.

Figure 9.7: Proper decay length efficiency functions as evaluated in 2017 (left) and 2018 (right) $B_s^0 \to J/\psi\,\phi$ simulated events. The solid line represents the fit results (Eq. 9.27) to the efficiency values (solid markers) in the various ct bins.

9.2.3.1 *Efficiency validation*

The procedure used to estimate the proper decay length efficiency is validated by measuring the lifetime of the $B^+ \to J/\psi\,K^+$ control channel in data.

Efficiency functions are evaluated in $B^+ \to J/\psi\,K^+$ simulated events and then used to fit the ct distribution in data. Eight different data-taking periods are considered, each with the number of B^+ candidates comparable with the one of signal candidates in the B_s^0 sample used in the analysis[26]. The results of the fits in the eight data samples are reported in Tab. 9.2. In all cases the results are in good agreement with the world-average B^+ meson lifetime, with differences no larger than 1.6 standard deviations, showing no large bias or instabilities during the data taking.

The measured efficiency functions are also cross-checked in simulations in the regular $B_s^0 \to J/\psi\,\phi$ simulated samples[27]. The measured ct is divided by the efficiency functions and the obtained distributions are found to be in excellent agreement with an exponential with a time constant equal to the B_s^0 lifetime value set in the simulation.

9.2.4 *Angular efficiency*

The reconstruction and selection procedures distort the shape of the angular distributions of the final-state particles. As such, in order to obtain unbiased

26 The data-taking periods are the same reported in Tab. 7.4.

27 These are the regular $B_s^0 \to J/\psi\,\phi$ simulated samples, *not* the ones with $\Delta\Gamma_s = 0$ used to estimate the efficiency functions.

Table 9.2: $B^+ \to J/\psi\, K^+$ lifetime measurement in different data-taking periods. The data sets are the same reported in Tab. 7.4. The quoted uncertainties are statistical only. The pull with respect to the world-average value $c\tau^{\text{w.a.}}_{B^+\to J/\psi\, K^+} = 491.1 \pm 1.2\ \mu\text{m}$ [34] are also reported. The quoted uncertainties are smaller than the world average because no systematic uncertainty is considered.

Data sample	$c\tau_{B^+\to J/\psi\, K^+}$ [μm]	Pull w.r.t. WA [s.d.]
2017-C	493.8 ± 2.4	$+1.0$
2017-D	494.8 ± 3.5	$+1.0$
2017-E	494.7 ± 2.3	$+1.4$
2017-F	489.5 ± 1.7	-0.7
2017 (total)	492.9 ± 1.1	$+1.1$
2018-A	489.3 ± 2.0	-0.8
2018-B	495.7 ± 2.7	$+1.6$
2018-C	489.2 ± 1.4	-1.0
2018-D	493.2 ± 1.3	$+1.2$
2018 (total)	492.78 ± 0.97	$+1.1$

results, the angular efficiency needs to be taken into account in the final fit model.

The angular efficiency is defined as the ratio of the observed 3D distribution of the three variables $\cos\theta_T$, $\cos\psi_T$ and φ_T to the true angular distribution:

$$\varepsilon(\Theta) = \varepsilon(\cos\theta_T, \cos\psi_T, \varphi_T) \propto \frac{N(\cos\theta_T, \cos\psi_T, \varphi_T)}{N_{\text{gen}}(\cos\theta_T, \cos\psi_T, \varphi_T)}\,. \tag{9.28}$$

In analogy with the ct efficiency, $N(\Theta)$ is measured in $B^0_s \to J/\psi\,\phi$ simulated samples generated with $\Delta\Gamma_s = 0$ to disentangle angles from time evolution, while $N_{\text{gen}}(\Theta)$ is extracted from the generation information of the $B^0_s \to J/\psi\,\phi$ samples where no geometrical acceptance filter is applied. All the other MC samples are simulated requiring the final-state particles to be generated inside the CMS geometrical acceptance. While this is not an issue for reconstructed events, it can cause border effects when generation information is used, as reconstructed events can migrate in and out the geometrical acceptance due to finite resolution in the determination of the track positions.

EFFICIENCY PARAMETRIZATION Angular efficiency functions are constructed separately for the 2017 and 2018 data samples with the following procedure.

First, Eq. 9.28 is evaluated as a 3D ratio histogram obtained from the $N(\Theta)$ and $N_{\text{gen}}(\Theta)$ distributions. Several binning combinations have been tested, evaluating both the computational time of the procedure and accuracy. The optimal configuration is found to be 70 bins for $\cos\theta_T$ and $\cos\psi_T$, and 30 for φ_T, for a total of 147 000 bins.

Next, the efficiency is parametrized, as

$$\varepsilon(\Theta) = \sum_{l,k,m} c_{l,k,m} \cdot b_{l,k,m}(\Theta)\,, \tag{9.29}$$

where $c_{l,k,m}$ are coefficients obtained projecting the efficiency histogram on a orthogonal basis constructed from spherical harmonics and Legendre polynomials defined by the elements $b_{l,k,m}(\Theta)$. The basis elements and the projection coefficients are respectively defined as:

$$b_{l,k,m}(\Theta) = P_l^m(\cos\theta_T) \cdot P_k^m(\cos\psi_T) \cdot \begin{cases} \sin(m\varphi_T) & \text{if } m < 0 \\ \cos(m\varphi_T) & \text{if } m > 0 \\ 1/2 & \text{if } m = 0 \end{cases} \tag{9.30}$$

and

$$c_{l,k,m} = \frac{1}{\left\|P_l^m\right\|^2 \left\|P_k^m\right\|^2 \pi} \sum_b^{N_{\text{bin}}} \frac{h_b b_{l,k,m}(x_b)}{V_b}\,, \tag{9.31}$$

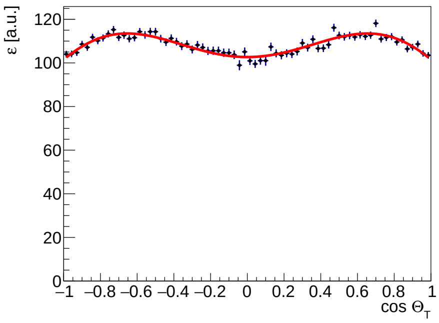
where

- P_i^m are the associated Legendre polynomials of degree i and order m[XXVIII],
- b is an index that run over all the 147 000 bins of the efficiency histogram,
- $b_{l,k,m}(x_b)$ is the basis element evaluated in the center x of the bin b,
- h_b and V_b are the *height* and the *volume* of the bin b, respectively,
- $\left\|P_i^m\right\|^2$ is L^2 norm of the associated Legendre polynomial P_i^m.

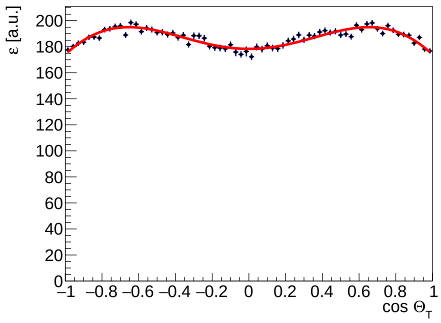
This basis is chosen because it is the simplest orthogonal basis for functions of three angular variables. The optimal maximum value of the degree of the polynomials i is found out to be 6, after testing all the values between 3 and 12[XXIX]. Higher degree polynomials are found to only increase the computational time without improving the efficiency description. This procedure is repeated separately for the 2017 and 2018 samples.

The projections of the acceptance histograms and the respective efficiency functions are shown in Fig. 9.8. The shape of the efficiency functions is very similar between the two samples. A similar observation can be made on the ct efficiency functions shown in Fig. 9.7, indicating no major acceptance difference between the two years of data-taking.

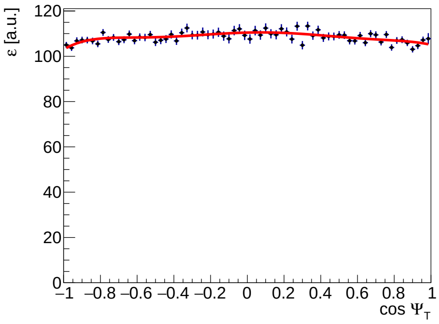
XXVIII In Equations 9.29, 9.30 and 9.31 the polynomials degree i is indicated with l and k.
XXIX For a polynomial of degree i, the order m can only assume integer values $0 \leqslant m \leqslant i$.



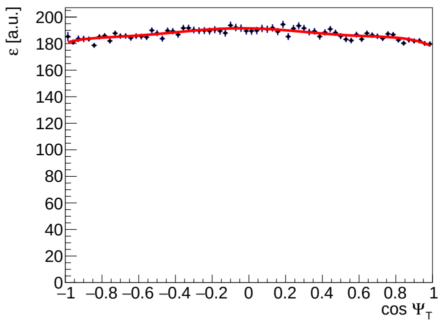
(a) $\varepsilon(\Theta)$ projection on $\cos\theta_T$ (2017).



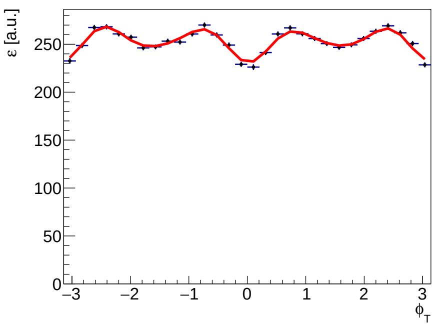
(b) $\varepsilon(\Theta)$ projection on $\cos\theta_T$ (2018).



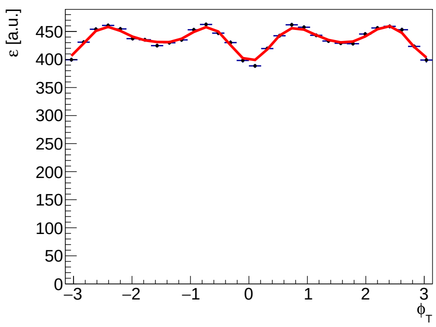
(c) $\varepsilon(\Theta)$ projection on $\cos\psi_T$ (2017).



(d) $\varepsilon(\Theta)$ projection on $\cos\psi_T$ (2018).



(e) $\varepsilon(\Theta)$ projection on φ_T (2017).



(f) $\varepsilon(\Theta)$ projection on φ_T (2018).

Figure 9.8: Projection of the angular efficiency histograms as evaluated in 2017 (left) and 2018 (right) $B_s^0 \to J/\psi\,\phi$ simulated events. The red line represents the respective projection of the efficiency function $\varepsilon(\Theta)$. The same binning used in the efficiency estimation procedure is used (70 bins for $\cos\theta_T$ and $\cos\psi_T$, and 30 for φ_T).

9.3 FIT MODEL

The fit model is composed of probability density functions describing the signal (P_{sig}) and background (P_{bkg}) components. The signal and background PDFs are formed as the product of functions that model the invariant mass distribution and the time-dependent decay rates of the reconstructed candidates. In addition, the time and angular efficiency functions are included in the signal model.

The event PDF is defined as:

$$P = \frac{N_{sig}}{N_{tot}} P_{sig} + \frac{N_{bkg}}{N_{tot}} P_{bkg} \,, \tag{9.32}$$

where

$$P_{sig} = \varepsilon(ct)\,\varepsilon(\Theta)\left[\widetilde{f}(\Theta, ct \,|\, \alpha) \otimes G(ct, \sigma_{ct})\right] P_{sig}(m_{B_s^0})\, P_{sig}(\sigma_{ct})\, P_{sig}(\xi_{tag})\,, \tag{9.33}$$

and[30]

$$P_{bkg} = P_{bkg}(\cos\theta_T, \varphi_T)\, P_{bkg}(\cos\psi_T)\, P_{bkg}(ct)\, P_{bkg}(m_{B_s^0})\, P_{bkg}(\sigma_{ct}) P_{bkg}(\xi_{tag})\,. \tag{9.34}$$

The corresponding negative log-likelihood is equal to[31]:

$$-\ln\mathcal{L} = -\sum_{i=0}^{N_{evt}} \ln P_i + N_{tot} - N_{evt}\ln N_{tot}. \tag{9.35}$$

The yields of signal and background events are free to float parameters indicated with N_{sig} and N_{bkg}, respectively, with their sum being N_{tot}. The number of candidates selected in data is indicated with $N_{evt} = 65\,470$. The second and third terms of Eq. 9.35 are related to the extended term described in Section 9.1.

The PDF $\widetilde{f}(\Theta, ct \,|\, \alpha)$ is the differential decay rate function $f(\Theta, ct \,|\, \alpha)$ defined in Eq. 3.19, modified to include the tagging information, as explained in Section 9.2.1[32]. All assumptions listed in Tab. 9.1 are used.

The efficiency functions $\varepsilon(ct)$ and $\varepsilon(\Theta)$, the ct resolution function $G(ct, \sigma_{ct})$ and the signal σ_{ct} distribution have been already defined in Sections 9.2.3, 9.2.4 and 9.2.2.1, respectively. The other PDFs in Eqs. 9.33 and 9.34 are described in the following sections and summarized in Tab. 9.3.

In the next section, fits performed in simulated samples and data sidebands are presented for validation and visualization purposes. Unless stated otherwise, they should not be intended as *pre-fits*. The fit to data procedure, completed of a list of pre-fitted and fixed components, is described in detail in Section 9.4.

---

30 The angular background is factorized in $P_{bkg}(\cos\theta_T, \varphi_T)\, P_{bkg}(\cos\psi_T)$, as it is observed in simulations that the correlation between $\cos\psi_T$ and the other two angular observables is small.

31 The constant term $\sum_i^{N_{evt}} \ln i$ is omitted.

32 Here, $\alpha = (\phi_s, \Delta\Gamma_s, \Gamma_s, \Delta m_s, |\lambda|, |A_0|, |A_\perp|, \left|A_\parallel\right|, |A_S|, \delta_0, \delta_\parallel, \delta_\perp, \delta_S)$ is the set of physics parameters.

Table 9.3: Summary of the distribution models used to parametrize the signal and background distributions in Eq. 9.32. The sum of two distributions is indicated with "2×". "Decay rate model" indicates that the variable is described by the $B_s^0 \to J/\psi\,\phi$ differential decay rate defined in Eq. 3.19. The tagging decision distribution is not explicitly reported and it is defined in Eq. 9.37. Unless stated otherwise all the distributions are independent and do not share any parameter. The definition of combinatorial and peaking background is given later in the text in Section 9.3.2.

Variable	Signal	Combinatorial bkg.	Peaking bkg.
$m_{B_s^0}$	Johnson's S_U	Exponential	Johnson's S_U
ct	Decay rate model	2×Exponential	Exponential
Θ	Decay rate model	Legendre polynomials	Legendre polynomials
σ_{ct}	2×Gamma	2×Gamma	Same as signal

9.3.1 *Signal model*

The signal model has been developed in $B_s^0 \to J/\psi\,\phi$ simulated samples. Several choices of the PDFs have been tested for all the variables. The chosen functions are the ones found out to best model each individual variable in the simulated events.

INVARIANT MASS The signal mass is modeled using a Johnson's S_U distribution [171]. This PDF is obtained from transforming a normally distributed variables x to

$$x \to z = \gamma + \delta \sinh^{-1}\left(\frac{x-\mu}{\lambda}\right), \tag{9.36}$$

where γ, λ and μ are parameters left to float. With respect to the Gaussian distribution, the Johnson's S_U distribution can be asymmetric if the γ is different from zero. The B_s^0 invariant mass spectrum in the simulated samples, and corresponding fit, is shown in Fig. 9.9.

TAGGING DECISION The signal tag decision[33] PDF is constructed so that the probability for an untagged event is $1-\varepsilon_{tag}$ and the probability of an event tagged as B_s^0 ($\overline{B}_s^0$) is $\varepsilon_{tag} N_{B_s^0}/N_{tag}$ ($\varepsilon_{tag} N_{\overline{B}_s^0}/N_{tag}$), where $N_{tag} = N_{B_s^0} + N_{\overline{B}_s^0}$ is the total number of tagged events. Mathematically it is defined as:

$$P_{sig}(\xi_{tag}) = \begin{cases} 1-\varepsilon_{tag} & \text{if } \xi_{tag} = 0 \\ \frac{1}{2}\varepsilon_{tag}(1 \pm A_{tag}) & \text{if } \xi_{tag} = \pm 1\,, \end{cases} \tag{9.37}$$

33 The tagging decision ξ_{tag} is defined in Eq. 8.1.

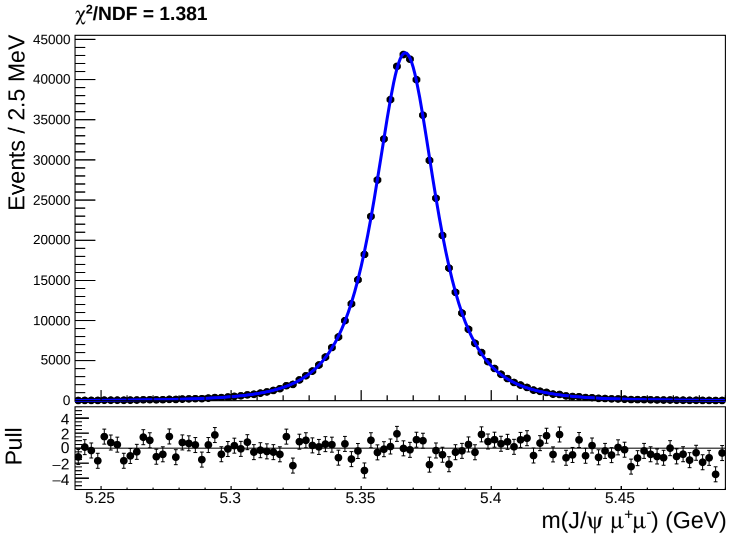


Figure 9.9: Invariant mass distribution in simulated $B_s^0 \to J/\psi\, \phi$ samples. The solid line represents the $P_{sig}(m_{B_s^0})$ fit results.

where A_{tag} is the tagging asymmetry computed as:

$$A_{tag} = \frac{N(\xi_{tag} = +1) - N(\xi_{tag} = -1)}{N(\xi_{tag} = +1) + N(\xi_{tag} = -1)} .$$

Both ε_{tag} and A_{tag} are free parameters of the fit model. The tagging asymmetry is expected to be very small, as there is no reason for observing a measurable flavor asymmetry in the selected data sample. The measured values in the final fit to data are $A_{tag}^{2017} = (-0.8 \pm 1.1) \cdot 10^{-2}$ and $A_{tag}^{2018} = (-1.50 \pm 0.84) \cdot 10^{-2}$ for the 2017 and 2018 data samples, respectively, both compatible with zero. The background tagging decision PDF is defined in the same way[34], as such, it will not be described in Section 9.3.2. The tagging asymmetry is found to be compatible with zero also for the background component.

9.3.2 *Background model*

The background PDF contains two terms, one related to the dominating combinatorial background and another to model the *peaking*[35] background from $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$, where the pion is mistaken as a kaon candidate. Another contribution to the peaking component of the background is the $\Lambda_b^0 \to J/\psi\, K^- p \to \mu^+\mu^- K^- p$ channel where the proton is mistaken as a kaon candidate. However, this particular contribution is not included in the

34 $P_{sig}(\xi_{tag})$ and $P_{bkg}(\xi_{tag})$ have independent parameters.
35 This component is referred to as *peaking* since its invariant mass spectrum is not flat or monotonic.

background model, as it is estimated to have a negligible effect on the fit results compared to the systematic uncertainties discussed in Chapter 10.[36,37]

The combinatorial background model has been defined in the data mass sidebands 5.24–5.28 GeV and 5.45–5.49 GeV, and it is mostly left free to float in the final fit to data. On the other hand, the peaking background model has been defined using MC samples of generated $B^0 \to J/\psi\, K^*(892)^0$ reconstructed as $B_s^0 \to J/\psi\, \phi$ and it is fixed in the final fit to data. The final fit procedure, completed of a list of pre-fitted and fixed components, is described in detail in Section 9.4.

INVARIANT MASS The background invariant mass PDF $P_{bkg}(m_{B_s^0})$ is modeled with an exponential distribution and a Johnson's S_U distribution for the combinatorial and peaking contributions, respectively. The peaking background invariant mass distribution is shown in Fig. 9.10a; the peaking structure is clearly visible. Figure 9.11 shows the invariant mass distribution as measured in the $B_s^0 \to J/\psi\, \phi$ data sample, alongside the fit of the full mass model[38]. Overall, the invariant mass spectrum is found to be well described by the mass PDF. The yields of the signal and combinatorial background are left free to float, while the yield of the peaking background is fixed to the value estimated in Eq. 9.39.

PROPER DECAY LENGTH The background decay length PDF $P_{bkg}(ct)$ is described by the sum of two exponential distributions for the combinatorial background, while a single exponential distribution is used for the peaking components. The ct distribution as measured in the data sidebands is shown in Fig. 9.12, while the one estimated for the peaking background is shown in Fig. 9.10b.

PROPER DECAY LENGTH UNCERTAINTY The background ct uncertainty for the combinatorial component is described with the same model used for the signal, that is a sum of two gamma distributions[39]. The σ_{ct} distribution measured in the data sideband is shown in Fig. 9.13. With respect to the distribution measured in simulation[40], which resembles the signal component, a larger right tail is observed.

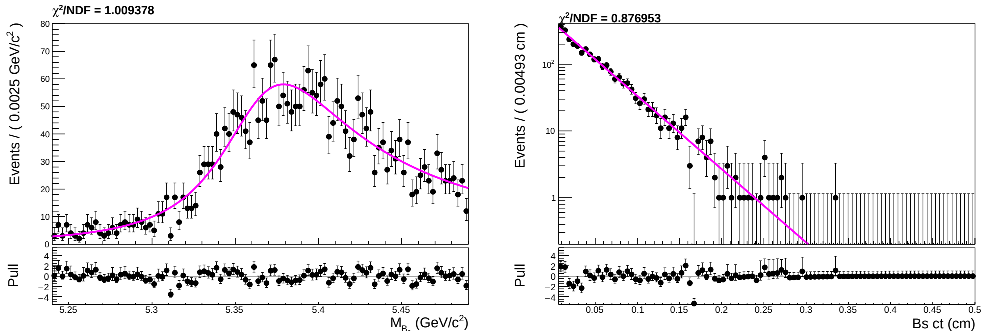
36 The expected number of $\Lambda_b^0 \to J/\psi\, K^- p$ events in the selected $B_s^0 \to J/\psi\, \phi$ data sample is estimated to be around ≈200 from simulations, depending on assumptions. A direct search of these decays in the $B_s^0 \to J/\psi\, \phi$ data sample has been performed to no success, leading to set an upper limit of 31 events at 68% CL.

37 Other contributions, such as the one from resonances decaying into pions (e. g. $B_s^0 \to J/\psi\, f_0(980) \to \mu^+\mu^-\pi^+\pi^-$), are also found to be negligible using simulations.

38 Note that this is a monodimensional fit to the invariant mass distribution and *not* a projection of the full fit.

39 The gamma distribution is defined in Section 9.2.2.1.

40 Shown in Fig. 9.6.



(a) Invariant mass. (b) Proper decay length.

Figure 9.10: Invariant mass (left) and proper decay length (right) distributions for $B_s^0 \to J/\psi\,\phi$ candidates reconstructed in $B^0 \to J/\psi\,K^*(892)^0$ simulated events, representing the peaking background from $B^0 \to J/\psi\,K^*(892)^0$. The solid lines represent the fits of the peaking background components of $P_{bkg}(m_{B_s^0})$ (left) and $P_{bkg}(ct)$ (right). The shape of these distributions are fixed in the final fit.

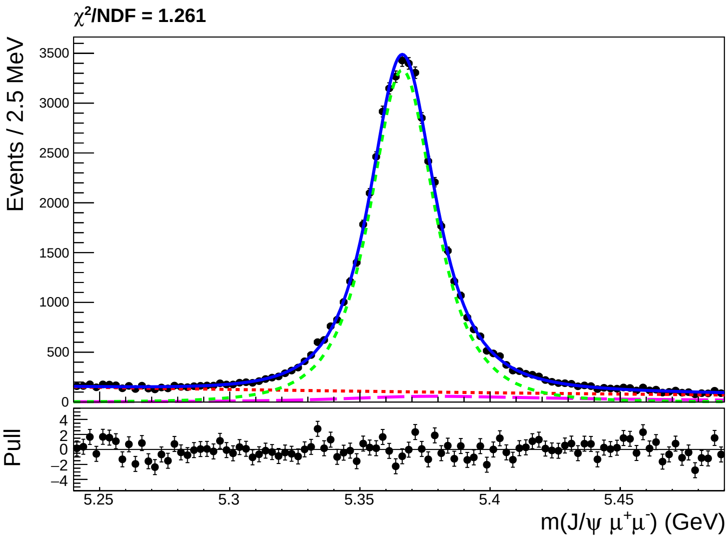


Figure 9.11: Invariant mass distribution for $B_s^0 \to J/\psi\,\phi$ candidates in the data samples. The solid line represents the $P(m_{B_s^0})$ fit, the dashed green line corresponds to the signal, the dotted red line to the combinatorial background, and the long-dashed magenta line to the peaking background from $B^0 \to J/\psi\,K^*(892)^0$.

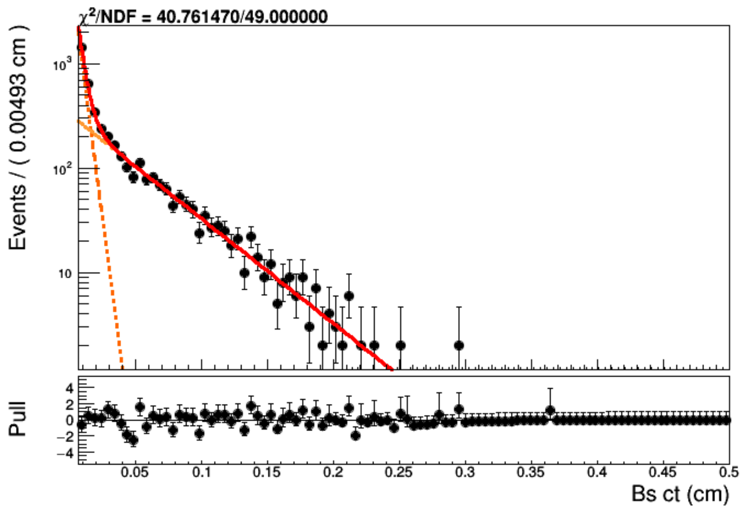


Figure 9.12: Proper decay length distribution as measured in the data mass sidebands 5.24–5.28 GeV and 5.45–5.49 GeV. The red line represents the fitted combinatorial component of $P_{bkg}(ct)$, which is defined as the sum of two exponential distributions (dotted lines).

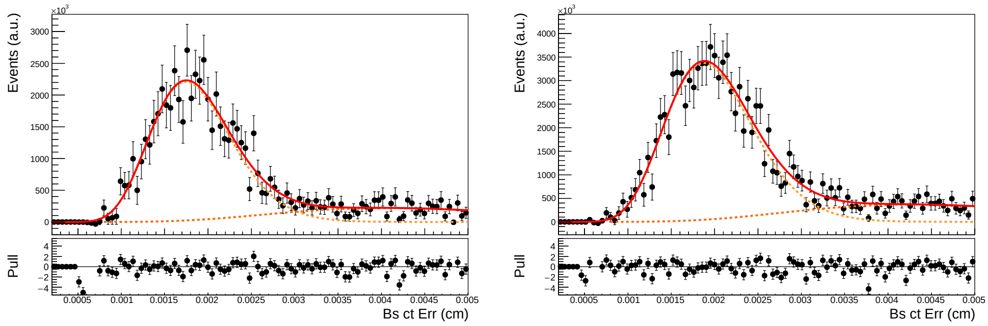
The peaking background component of $P_{bkg}(\sigma_{ct})$ is set equal to the distribution observed in data for the signal[41]. This choice is motivated by the fact that the proper decay length uncertainty is found to be modeled too optimistically in simulated samples, that is a lower value of σ_{ct} on average is found, with respect to what observed in data. The signal distribution is chosen as the model for the peaking background component of $P_{bkg}(\sigma_{ct})$ since the final-state topologies of $B^0_s \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ and $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^-\,K^+\pi^-$ decays are very similar, presenting two muons and two charged tracks belonging to the same displaced vertex.

ANGULAR BACKGROUND MODEL The background model for the angular variables is factorized as:

$$P_{bkg}(\Theta) = P_{bkg}(\cos\theta_T, \varphi_T)P_{bkg}(\cos\psi_T)\,, \tag{9.38}$$

where the $P_{bkg}(\cos\theta_T, \varphi_T)$ and $P_{bkg}(\cos\psi_T)$ are constructed as linear combinations of Legendre polynomials of order $i = 2$, 4 and 6. The combinatorial background component of $P_{bkg}(\Theta)$ as measured in the data mass sidebands is shown in Fig. 9.14, while the peaking background component, as measured in $B^0 \to J/\psi\,K^*(892)^0$ simulated events, is shown in Fig. 9.15.

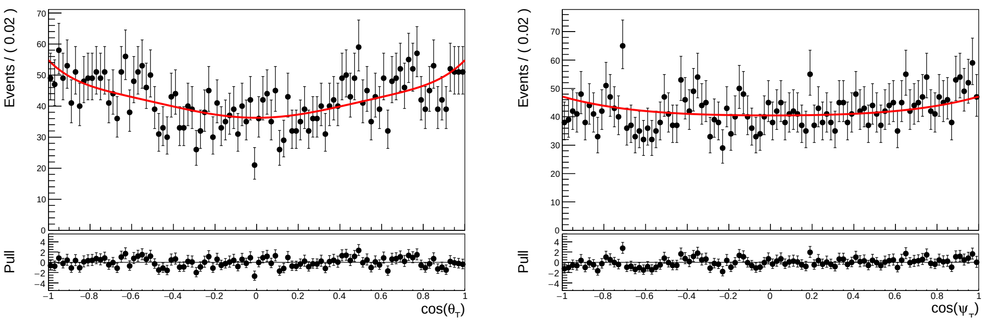
41 I.e., $P^{peak}_{bkg}(\sigma_{ct}) = P_{sig}(\sigma_{ct})$.



(a) 2017 data sample.

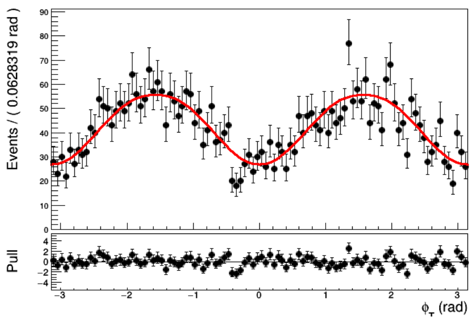
(b) 2018 data sample.

Figure 9.13: Proper decay length uncertainty distribution as measured in the data mass sidebands 5.24–5.28 GeV and 5.45–5.49 GeV. The red line represents the fitted combinatorial component of $P_{bkg}(\sigma_{ct})$, which is defined as the sum of two gamma distribution (dotted lines). *(fix)*



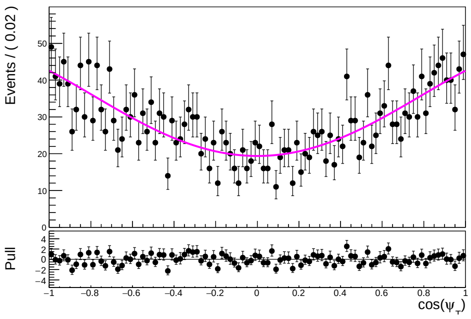
(a) $\cos\theta_T$ (combinatorial background).

(b) $\cos\psi_T$ (combinatorial background).

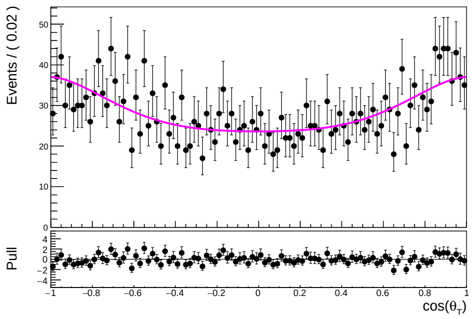


(c) φ_T (combinatorial background).

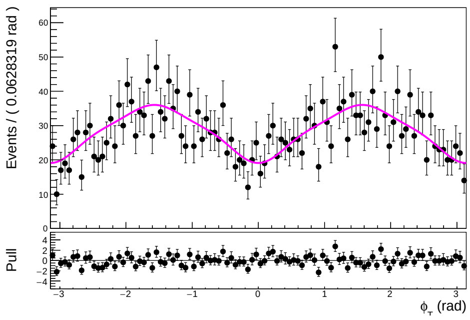
Figure 9.14: Angular distributions for $B_s^0 \to J/\psi\,\phi$ candidates reconstructed as measured in the data mass sidebands 5.24–5.28 GeV and 5.45–5.49 GeV. The solid line represents the projection of the combinatorial background component of the $P_{bkg}(\Theta)$ fit.



(a) $\cos\theta_T$ (peaking background).



(b) $\cos\psi_T$ (peaking background).



(c) φ_T (peaking background).

Figure 9.15: Angular distributions for $B_s^0 \to J/\psi\,\phi$ candidates reconstructed in $B^0 \to J/\psi\,K^*(892)^0$ simulated events, representing the peaking background from $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^-\,K^+\pi^-$. The solid line represents the projection of the peaking background component of the $P_{bkg}(\Theta)$ fit.

PEAKING BACKGROUND YIELD ESTIMATION The expected event yield for the peaking background is estimated, separately for the 2017 and 2018 data samples, using simulated samples of known integrated luminosity, as:

$$N_{\text{peak}}^{\text{data}} = N_{\text{peak}}^{\text{MC}} \frac{\mathcal{L}_{\text{int}}^{\text{data}}}{\mathcal{L}_{\text{int}}^{\text{MC}}}, \tag{9.39}$$

where $N_{\text{peak}}^{\text{MC}}$ is the number of reconstructed $B_s^0 \to J/\psi\,\phi$ candidates in the $B^0 \to J/\psi\,K^*(892)^0$ simulated samples and $\mathcal{L}_{\text{int}}$ are the integrated luminosity values of the respective samples. Using the values of $\mathcal{L}_{\text{int}}^{\text{data}}$ reported in Tab. 7.4, the peaking background yield is estimated to be 1150 and 1640 events for the 2017 and 2018 data sample, respectively. No statistical uncertainty is given, as these yields are fixed in the final fit. A systematic uncertainty related to the precision of this estimate is described in Section 10.3.

9.3.3 *Fit model validation*

The fit model is validated by performing several UML fits to $B_s^0 \to J/\psi\,\phi$ simulated samples.

The decay rate model defined in Eq. 3.19 is validated by fitting the generation information of the large[42] $B_s^0 \to J/\psi\,\phi$ simulated sample where no acceptance filters have been applied. Since the fit is not performed on reconstructed events no resolution or acceptance effects are taken into account. The assumption and conventions reported in Tab. 9.1 are used. The extracted physics parameters are found to be in excellent agreement with the simulation values.

The signal fit model, defined in Eq. 9.33, is validated by fitting the regular $B_s^0 \to J/\psi\,\phi$ simulated samples, as they represent a simulation of the expected signal in data. The results for all the physics parameters are found to be in good agreement with the simulation values. These fits also provide a cross-check for the efficiency functions evaluated in the $B_s^0 \to J/\psi\,\phi$ ($\Delta\Gamma_s = 0$) simulated samples.

The full fit model is validated with multiple fits to *pseudo-experiments*. Since this validation is also used to assess one of the leading systematic uncertainties, the related description is postponed in Chapter 10.

CORRELATIONS BETWEEN FIT COMPONENTS The correlations between the different fit components are studied in both data and simulation and found to be overall negligible. The only significant correlation is found to be in data between the B_s^0 mass and its proper decay length, which is measured to be $\approx$1% and found to have no impact on the final results[43]. The other correlations are found

42 The number of fitted events is roughly 8.5 millions.

43 This has been studied with pseudo-experiments with a procedure similar to the one used to evaluate the *model assumptions* systematic uncertainty described in Section 10.2.1.

to be of the order of $\approx 1‰$, the same order as expected for random correlations in a sample of the same statistics as data.

9.4 FIT PROCEDURE

The physics parameters are extracted with an unbinned extended maximum-likelihood fit performed simultaneously in the 2017 and 2018 data samples. The joint likelihood function shares the decay rate model, the invariant mass PDFs, the peaking background model, and the lifetime and angular components of the combinatorial background model between the two samples. The yields of the signal and background components are measured separately in each data sample, as is the tagging efficiency. The efficiency functions, $P(\sigma_{ct})$ PDFs, tagging decision PDFs, and κ_{ct} factors are also specific to each data sample.

Due to the complexity of the model, the fit cannot be performed in a single step and several cycles of *pre-fits* are used to determine the starting values of the various parameters describing the individual distributions. In what follows, a distribution is indicated as "*pre-fitted*" if the starting values of its parameters have been determined with pre-fits and then left free to float in the final fit. On the other hand, if the parameters are not free to float in the final fit, but instead fixed to what obtained in the pre-fits, the PDF is referred to as *fixed*.

The combinatorial component of the background PDF is pre-fitted to the data mass sidebands, while the peaking component is fixed, except for the σ_{ct} PDF, to the results of the UML multidimensional fit to the $B^0 \to J/\psi\, K^*(892)^0$ simulated samples[44]. The peaking background event yield is also fixed to the value estimated with Eq. 9.39. Due to their instability in the final fit, the signal and background components of the decay length uncertainty PDF are fixed to the ones obtained from a two-dimensional fit together with the invariant mass PDF. The only distributions that are not pre-fitted are the differential decay rate $\widetilde{f}(\Theta, ct \,|\, \alpha)$ and the tag decision PDFs $P(\xi_{tag})$. The pre-fitted, fixed, and shared PDFs are summarized in Tab. 9.4.

After all the cycles of pre-fits, the UML fit to the combined 2017+2018 data samples is finally performed.

9.5 FIT RESULTS

The results of the fit with their statistical uncertainties are given in Tab. 9.5. Statistical uncertainties are obtained with the MINOS method from the increase in $-\log\mathcal{L}$ by 0.5 . The statistical correlation matrix obtained with the HESSE method is reported in Tab. 9.6. The CP-violating phase ϕ_s is found to be weakly correlated to the other physics parameters, with the largest correlation being 22% with $|\lambda|$. The measured number of $B^0_s \to J/\psi\, \phi \to \mu^+\mu^- K^+K^-$ signal events

44 This fit is performed on MC samples of generated $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ events reconstructed as $B^0_s \to J/\psi\, \phi \to \mu^+\mu^- K^+K^-$.

Table 9.4: List of pre-fitted, fixed and shared parameters, functions and distributions. The "pre-fitted" column indicates PDFs for which the parameter starting values have been determined with pre-fits and then left free to float in the final fit. On the other hand, the "fixed" column indicates distributions for which the parameters are fixed to what obtained in the pre-fits. Finally, the "shared" column indicates if PDF is shared between the 2017 and 2018 data samples in the joint likelihood. The "comb" and "peak" subscripts indicate quantities related to the combinatorial and peaking components of the background, respectively. The combinatorial background PDFs are pre-fitted in the data mass sidebands. The peaking background PDFs are pre-fitted in the $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ simulated samples. The dashes (—) indicates inapplicable conditions (e. g. $P_{\text{peak}}(\sigma_{ct})$ is set to be equal to $P_{\text{sig}}(\sigma_{ct})$).

Object	Pre-fitted	Fixed	Shared
$\widetilde{f}(\Theta, ct \mid \alpha)$			✓
$\varepsilon(ct)\varepsilon(\Theta)$	—	✓	
κ_{ct}	—	✓	
N_{sig}	✓		
$P_{\text{sig}}(m_{B^0_s})$	✓		✓
$P_{\text{sig}}(\sigma_{ct})$	✓	✓	
$P_{\text{sig}}(\xi_{\text{tag}})$			
N_{comb}	✓		
$P_{\text{comb}}(m_{B^0_s})$	✓		✓
$P_{\text{comb}}(ct)$		✓	✓
$P_{\text{comb}}(\sigma_{ct})$	✓	✓	
$P_{\text{comb}}(\Theta)$	✓		✓
$P_{\text{comb}}(\xi_{\text{tag}})$			
N_{peak}	✓	✓	
$P_{\text{peak}}(m_{B^0_s})$	✓	✓	✓
$P_{\text{peak}}(ct)$	✓	✓	✓
$P_{\text{peak}}(\sigma_{ct})$	—	—	—
$P_{\text{peak}}(\Theta)$	✓	✓	✓
$P_{\text{peak}}(\xi_{\text{tag}})$	✓	✓	✓

from the fit is $48\,500 \pm 250$. The distributions of the input observables and the corresponding fit projections are shown in Fig. 9.16.

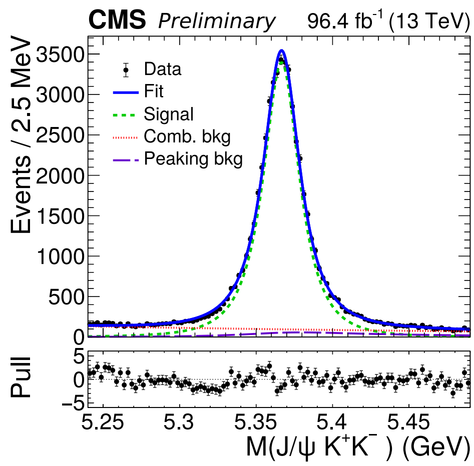
Overall, the fitted model is found to describe well the distribution of the input observables, although some minor discrepancies are observed. The fit projection to the proper decay length (Fig. 9.16b) shows a small discrepancy in the low ct region, while the one to $\cos\theta_T$ (Fig. 9.16d) exhibits a slight left-right asymmetry. The impact on the final results of these effects is estimated, and taken into account, in the so-called "model assumption systematic uncertainty", described later in Section 10.2.1. In both cases the effect is minimal[45]. The fit projection to the invariant mass (Fig. 9.16a) shows a mismodeling in the left tail around 5.33 GeV, which however is not observed in the one-dimensional mass fit (Fig. 9.11). Upon further investigation, it has been observed that the signal and background yields obtained in the one-dimensional fit are slightly different from those obtained from the final fit, which is the cause of the distortions seen in the $P(m_{B^0_s})$ component of the fit model. The impact on the final results of these effects is taken into account in both the "model assumption" and the "peaking background" systematic uncertainties, described in Sections 10.2.1 and 10.3 respectively, which estimate the possible biases in the assumptions made in the mass and peaking background fit models. The impact on the total uncertainty on the final results is found to be negligible[45].

The CP-violating phase ϕ_s is measured with a statistical uncertainty of 50 mrad, half the one obtained in the previous CMS measurement at $\sqrt{s} = 8$ TeV [64], which was based on a similar amount of B^0_s candidates. This reflects the new trigger strategy which drastically improves the tagging efficiency at the cost of a reduced number of signal events. Notably, the improvement in the precision on ϕ_s roughly reflects the sensitivity improvement estimated with Eq. 9.21. At the same time, the statistical precision on parameters that do not benefit from the tagging information, such as $\Delta\Gamma_s$ and Γ_s, is comparable to that of the previous measurement. It is worth mentioning that the number of B^0_s candidates selected in this work is similar only accidentally to the one of the previous CMS measurement. With respect to the previous measurement, the $B^0_s \to J/\psi\,\phi$ data sample in this work is collected with five times the integrated luminosity at a center-of-mass energy that roughly doubles the $pp \to b\bar{b}$ production cross section. These improvements are almost perfectly counterbalanced by the fact that the HLT used in selecting the data samples requires a third muon, which cost a factor of ten due to the branching fraction $\mathcal{B}(b \to \mu) \sim 10\%$.

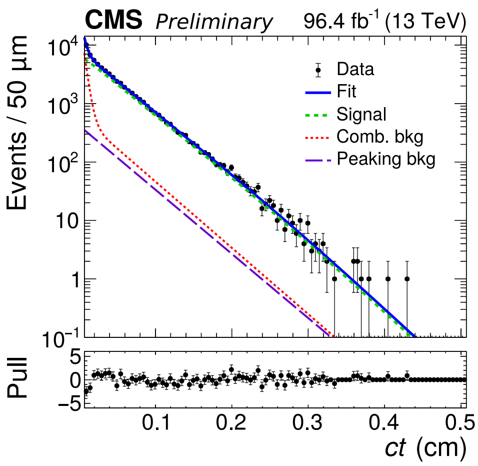
This analysis also represents the first measurement with the CMS experiment of the mass difference Δm_s between the heavy and light B^0_s mass eigenstates and of the direct CP observable $|\lambda|$.

The fit results are discussed more in detail in Chapter 11, where the combination with the earlier CMS result at a center-of-mass energy of 8 TeV is also presented.

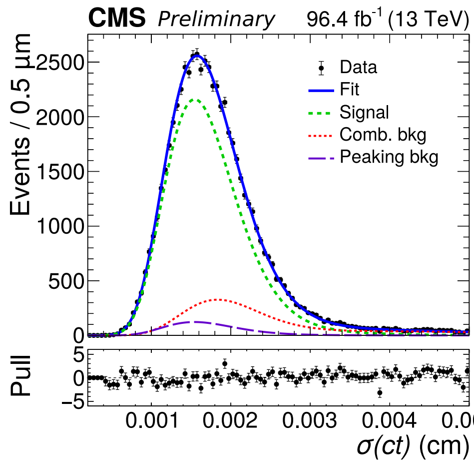
45 See Tab. 10.3 and Tab. 10.9.



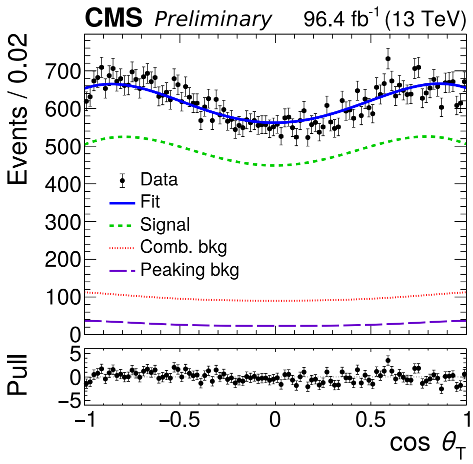
(a) Invariant mass distribution.



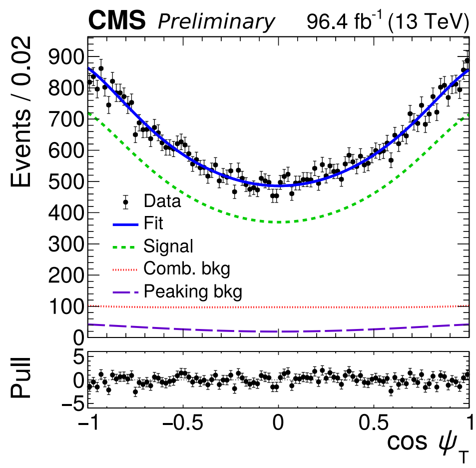
(b) Proper decay length distribution.



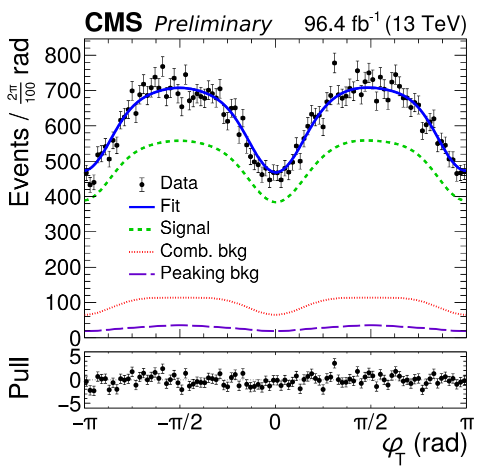
(c) Proper decay length uncertainty distribution.



(d) $\cos\theta_T$ distribution.



(e) $\cos\psi_T$ distribution.



(f) φ_T distribution.

Figure 9.16: Distribution of the input observables and the corresponding fit projections. The vertical bars on the points represent the statistical uncertainties. The solid line represents a projection of the fit to data (solid markers), the dashed line corresponds to the signal, the dotted line to the combinatorial background, and the long-dashed line to the peaking background from $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$, as obtained from the fit. The distribution of the differences between the data and the fit, divided by the combined uncertainty in the data and the best fit function for each bin (pulls) is displayed in the lower panel. Figure from Ref. [1] as of 30 November 2020.

Table 9.5: Results of the fit to data for the physics parameters of most interest. The quoted statistical uncertainties are obtained from the increase in $-\log\mathcal{L}$ by 0.5. The results are converted to the units of measure used by the Particle Data Group [34].

Parameter	Fit value	Stat. uncer.
ϕ_s [mrad]	-11	± 50
$\Delta\Gamma_s$ [ps^{-1}]	0.114	± 0.014
Δm_s [$\hbar\text{ps}^{-1}$]	17.51	$^{+0.10}_{-0.09}$
$\lvert\lambda\rvert$	0.972	± 0.026
Γ_s [ps^{-1}]	0.6531	± 0.0042
$\lvert A_0\rvert^2$	0.5350	± 0.0047
$\lvert A_\perp\rvert^2$	0.2337	± 0.0063
$\lvert A_S\rvert^2$	0.022	$^{+0.008}_{-0.007}$
$\delta_\parallel$ [rad]	3.18	± 0.12
$\delta_\perp$ [rad]	2.77	± 0.16
$\delta_{S\perp}$ [rad]	0.221	$^{+0.083}_{-0.070}$

Table 9.6: Statistical correlations between the main physics parameters of interest as evaluated with the HESSE method. The lower half of the matrix is not reported as it is symmetrical. Correlations larger than 0.5 are underlined.

	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$	$\delta_\perp$	$\delta_{S\perp}$	Γ_s	$\Delta\Gamma_s$	Δm_s	$\lvert\lambda\rvert$	ϕ_s
$\lvert A_0\rvert^2$	+1.00	$\underline{-0.61}$	+0.15	−0.01	−0.01	−0.09	−0.32	$\underline{+0.63}$	−0.01	0.00	−0.01
$\lvert A_\perp\rvert^2$		+1.00	−0.11	−0.06	0.00	+0.06	+0.42	$\underline{-0.71}$	+0.02	−0.01	+0.01
$\lvert A_S\rvert^2$			+1.00	−0.07	−0.02	−0.44	+0.15	0.00	+0.03	−0.03	−0.02
$\delta_\parallel$				+1.00	+0.27	−0.02	−0.03	+0.01	−0.01	−0.06	−0.02
$\delta_\perp$					+1.00	−0.10	+0.02	−0.01	$\underline{+0.68}$	−0.18	−0.09
$\delta_{S\perp}$						+1.00	−0.05	−0.04	−0.05	+0.05	+0.03
Γ_s							+1.00	−0.48	+0.03	−0.02	0.00
$\Delta\Gamma_s$								+1.00	−0.02	0.00	−0.02
Δm_s									+1.00	−0.14	−0.19
$\lvert\lambda\rvert$										+1.00	+0.22
ϕ_s											+1.00

10 SYSTEMATIC UNCERTAINTIES

This chapter describes the various sources of systematic uncertainties induced by the assumptions made in the fit model and in the analysis methods. In this work, the general approach in regards to biases is to treat them as systematic uncertainties, instead of using the information to correct the results or the fit model. The numerical values of the systematic uncertainties on the physics parameters are summarized at the end of the chapter in Tab. 10.9.

PSEUDO-EXPERIMENTS Most of the systematic uncertainties are evaluated using Monte Carlo *pseudo-experiments*, that is sets of *pseudo-data* (“toys”) generated following a given distribution and then fitted with the “nominal” fit model defined in Eq. 9.32. It is important to note that the model used to generate the toys does not have to be the same used to fit them. Typically, the extracted physics parameters are compared with the input values of the toys to construct *pull* distributions (i. e., the difference divided by the combined uncertainty). In absence of biases or other systematic effects, the pull distribution for each parameter should follow a Gaussian distribution with a central value equal to zero and unity width. Deviations from this expectation are typically used to assess the magnitude of systematic effects. Unless stated otherwise, the pseudo-experiments mentioned in this chapter are always generated to be statistically equivalent to the combined $B_s^0 \to J/\psi\,\phi$ data sample and always fitted with the nominal model defined in Eq. 9.32. Each study is performed with one thousand pseudo-experiments.

10.1 FITTING PROCEDURE

10.1.1 *Fit bias*

Possible biases arising from the fitting procedure are evaluated generating pseudo experiments from the *fitted* model in data (hereafter “nominal-model pseudo-experiments”), that is with the parameters fixed to the final fit results, and fitting them with the nominal model defined in Eq. 9.32. The obtained pull distributions of the fitted parameters are shown in Fig. 10.1. In absence of biases, that is if the fit reproduces on a statistical basis the set values, the pull distributions are expected to be centered in zero. A Gaussian function is used to fit the pull distribution of each parameter and the estimated central value is taken as the corresponding systematic uncertainty, if different from zero by more than its error[1]. The central values are reported in Tab. 10.1. Most of the parameters show a fit bias around 10-20% of the respective statistical uncertainty. When considering the total systematic uncertainty, the fit bias is found to be the leading source of systematic uncertainty in the measurement of the CP-violating phase ϕ_s. At the current level of precision, the fit bias is still negligible with respect to the total uncertainty on the final results. However, this may not be the case in future measurements, where the magnitude of statistical and systematic uncertainties could be comparable.

This study also represents a validation of the fit model and was performed before the actual final fit to the data samples.

To avoid double-counting this uncertainty, whenever pseudo-experiments are used to evaluate other systematic uncertainties, the model bias is always subtracted. In these cases, the procedure used is to compare the corresponding pull distributions to those obtained with the nominal-model pseudo-experiments. If the central values of the two pull distributions differ from each other by more than their combined uncertainty, the difference is taken as the corresponding systematic uncertainty. In other words, systematic uncertainties are evaluated only if

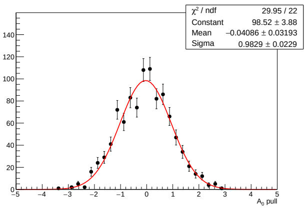
$$\frac{|\mu_{\text{ref}} - \mu|}{\sqrt{\sigma^2_{\mu_{\text{ref}}} + \sigma^2_{\mu}}} > 1\,, \tag{10.1}$$

where μ_{ref} is the central value of the nominal-model pseudo-experiments pull distribution and μ the central value of the pull distribution of the study in question. This condition is applied separately to each parameter.

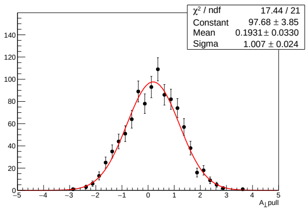
10.1.2 $P(\sigma_{ct})$ *uncertainty*

As described in Section 9.4, in the fit to data the proper decay length uncertainty PDF is fixed to the one obtained from a pre-fit. This means that the same data sample is used twice and the final statistical uncertainties may be underestimated. The uncertainty on the estimation of $P(\sigma_{ct})$ is propagated to the final results by sampling this distribution one thousand times, using the parameter uncertainties obtained from the pre-fit accounting for correlations. With each sampled $P(\sigma_{ct})$ the fit is repeated and the standard deviation of the obtained physics parameter distributions is taken as the systematic uncertainty. The $P(\sigma_{ct})$ uncertainty is found to be the leading source of systematic uncertainty for the mass difference Δm_s.

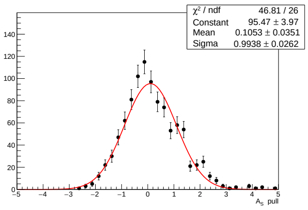
1 This systematic uncertainty is assigned if $|\mu|/\sigma_\mu > 1$, where μ and σ_μ are the central value estimate and its uncertainty.



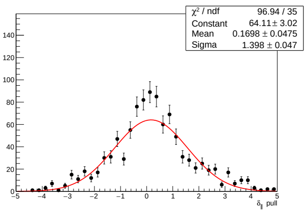
(a) $|A_0|^2$ pulls.



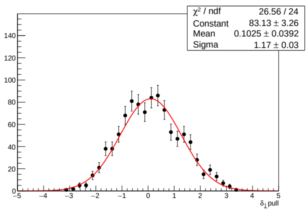
(b) $|A_\perp|^2$ pulls.



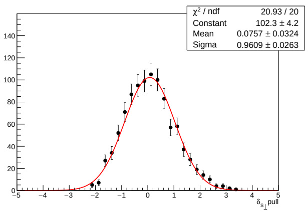
(c) $|A_S|^2$ pulls.



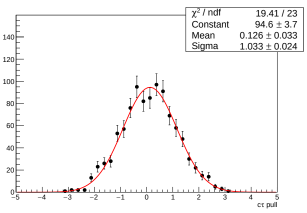
(d) $\delta_{||}$ pulls.



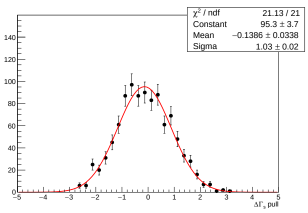
(e) $\delta_\perp$ pulls.



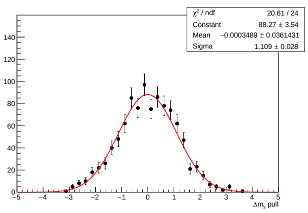
(f) $\delta_{S\perp}$ pulls.



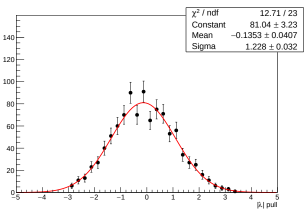
(g) Γ_s pulls.



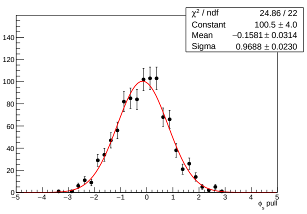
(h) $\Delta\Gamma_s$ pulls.



(i) Δm_s pulls.



(j) $|\lambda|$ pulls.



(k) ϕ_s pulls.

Figure 10.1: Pull distributions of the physics parameters as obtained from the one thousand pseudo-experiments generated to test the fit bias. Each pull is computed as the difference between the fit result and the simulation set value, normalized by the statistical uncertainty on the extracted parameter. The solid lines represent a Gaussian distribution fit. The central values as obtained from the fits are reported in Tab. 10.1 and used to assess the systematic uncertainty.

Table 10.1: Central values of the pull distribution obtained from the one thousand pseudo-experiments generated to test the fit bias. The estimated values and their uncertainties are obtained from a Gaussian distribution fit. The respective plots are shown in Fig. 10.1.

Parameter	Pulls central value
ϕ_s	-0.158 ± 0.031
$\Delta\Gamma_s$	-0.139 ± 0.034
Δm_s	0.000 ± 0.036
$\|\lambda\|$	-0.135 ± 0.041
Γ_s	0.123 ± 0.033
$\|A_0\|^2$	-0.041 ± 0.032
$\|A_\perp\|^2$	0.193 ± 0.033
$\|A_S\|^2$	0.105 ± 0.035
$\delta_\parallel$	0.170 ± 0.048
$\delta_\perp$	0.103 ± 0.039
$\delta_{S\perp}$	0.076 ± 0.032

10.2 FIT MODEL

10.2.1 *Model assumptions*

Several assumptions made in defining the fit PDF are tested as a possible source of bias by generating pseudo-experiments with different PDF hypotheses and fitting them with the nominal model. The pull distributions with respect to the input values are used to evaluate the systematic uncertainty, as described in the "model bias" paragraph[2]. The following assumptions are tested: signal and background invariant mass models, background proper decay length model, and background angular model. The tested hypotheses are reported in Tab. 10.2 while the resulting systematic uncertainty are reported in Tab. 10.3. The assumptions made on the background model are found to be one of the leading sources of systematic uncertainty on the direct CPV observable $|\lambda|$, while the ones made on the mass PDF are found to affect significantly the phase $\delta_\parallel$.

10.2.2 *Different mistag probability distribution in signal and background*

The fit model does not contain a dedicated term to describe the mistag probability (ω_{evt}) distribution, under the assumption that it is the same for both the signal and background component. The impact on the results of this assumption

2 Section 10.1.1.

Table 10.2: List of the alternative hypotheses used in evaluating the model assumptions systematic uncertainty. The distributions used in the nominal model are indicated with "nominal". The sum of two or three distributions of the same type is indicated with 2× and 3×, respectively.

PDF	Nominal distribution	Alternative distribution
$P_{sig}(m_{B_s^0})$	Johnson's S_U	3×Gaussian
$P_{bkg}(m_{B_s^0})$	Exponential	Chebyshev polynomials
$P_{bkg}(ct)$	2×Exponential	Nominal+Gaussian with $\mu = 0$
$P_{bkg}(\Theta)$	Legendre polynomials	Nominal+asymmetric terms

Table 10.3: Summary of the systematic uncertainties related to the model assumptions. The dashes (—) mean that the corresponding uncertainty is either not applicable or not evaluated (see the text for details). The total systematic uncertainty is obtained as the quadratic sum of the individual contributions.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Signal mass PDF	—	—	—	—	—	—	0.0008	—	0.017	0.019	0.011
Background mass PDF	—	—	—	—	0.0003	—	0.0005	0.001	—	—	—
Background ct PDF	—	—	—	0.0023	—	—	—	0.001	—	—	—
Background angular PDF	—	—	—	0.0040	—	—	0.0008	0.001	—	—	—
Total uncertainty	—	—	—	0.0046	0.0003	—	0.0013	0.001	0.017	0.019	0.011

is evaluated with pseudo-experiments generated with a PDF where different signal and background $P(\omega_{evt})$ distributions are used. These distributions are measured in data, by using the B_s^0 candidate invariant mass signal and side-bands regions, and separately modeled using the Kernel Density Estimation method [172, 173]. The pull distributions are used to evaluate the systematic uncertainty, as described in the "model bias" paragraph. This systematic uncertainty is found to be relatively small for all the physics parameters.

10.3 PEAKING BACKGROUND

The peaking background model is fixed in the final fit to that measured in $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^-\, K^+\pi^-$ simulated samples.

The systematic uncertainty related to the fixed event yield is evaluated by repeating the fit using a different yield value obtained from a $B^0 \to J/\psi\, K^*(892)^0$ control sample in data. The alternative yield is evaluated as:

$$N_{peak}^{data} = N_{peak}^{MC} \frac{N_{B^0}^{data}}{N_{B^0}^{MC}}\,, \tag{10.2}$$

where N^{MC}_{peak} and $N^{MC}_{B^0}$ are the numbers of $B^0_s \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ and $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^-\,K^+\pi^-$ candidates reconstructed in the $B^0 \to J/\psi\,K^*(892)^0$ simulated sample, respectively, and $N^{data}_{B^0}$ is the number of $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^-\,K^+\pi^-$ decays reconstructed in the data samples. Using Eq. 10.2 a peaking background yield of 2100 events is estimated for the combined data sample. This yield is found to be ≈25% lower than the one estimated on simulations and used in the nominal fit[3]. The fit is repeated using this value and the difference with respect to the nominal result is taken as the systematic uncertainty.

A systematic uncertainty is also assigned to the proper decay length modeling by forcing the B^0 lifetime to match the world-average value, repeating the fit, and taking the difference from the nominal result as the systematic uncertainty.

The peaking background systematic uncertainties are reported in Tab. 10.4 and found to be relatively small for all the physics parameters

Table 10.4: Summary of the systematic uncertainties related to the peaking background. The total systematic uncertainty is obtained as the quadratic sum of the individual contributions.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Event yield	0.3	0.0008	0.011	$<10^{-4}$	0.0001	0.0005	0.0002	0.003	0.005	0.007	0.011
Lifetime constant	$<10^{-1}$	$<10^{-4}$	0.001	$<10^{-4}$	0.0001	$<10^{-4}$	$<10^{-4}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
Total uncertainty	0.3	0.0008	0.011	$<10^{-4}$	0.0002	0.0005	0.0002	0.003	0.005	0.007	0.011

10.4 S-P WAVE INTERFERENCE

The fit model does not take into account the difference in the m_{KK} invariant mass dependence between the P-wave from the $B^0_s \to J/\psi\,\phi$ decay and the S-wave from non-resonant $B^0_s \to J/\psi\,K^+K^-$ and $B^0_s \to J/\psi\,f_0(980) \to \mu^+\mu^-\,K^+K^-$ decays. Since m_{KK} is not fitted, this may cause biases in the physics results.

The amplitude dependence on m_{KK} can be expressed as:

$$\begin{aligned} A_i(m_{KK}) &= A_i \pi(m_{KK})\,, \\ A_S(m_{KK}) &= A_S \sigma(m_{KK})\,, \end{aligned} \tag{10.3}$$

where $i = (0, \parallel, \perp)$, and $\pi(m_{KK})$ and $\sigma(m_{KK})$ is the mass dependence parametrization of the P- and S-wave amplitudes, respectively, and are assumed normalized to unity in the selected m_{KK} range. To properly model this dependence, the invariant mass of the kaon system should be parametrized for the signal and background components in the fit model. If m_{KK} is not fitted,

3 The peaking background yield in the nominal fit is estimated from the luminosity information of $B^0 \to J/\psi\,K^*(892)^0$ simulations to be ≈2800 events.

as in this analysis, the decay rate terms related to the interference between the P and S wave, i. e., the terms in Tab. 3.1 with $i = 8, 9, 10$, are modified by an effective coupling factor k_{SP}. The value of k_{SP} can be evaluated by integrating the amplitudes over the selected m_{KK} range $[m_1, m_2]$[4], as:

$$\int_{m_1}^{m_2} A_i^* A_S \, \pi(m_{KK})^* \sigma(m_{KK}) \, dm_{KK} = k_{SP} |A_i A_S| \,. \tag{10.4}$$

Since $\pi(m_{KK})$ and $\sigma(m_{KK})$ are normalized in $[m_1, m_2]$, the $|A_i|^2$, $|A_i A_j|$ and $|A_S|^2$ terms do not require any corrections:

$$\begin{aligned} \int_{m_1}^{m_2} A_i^* A_j \, \pi(m_{KK})^* \pi(m_{KK}) \, dm_{KK} &= |A_i A_j| \,, \\ \int_{m_1}^{m_2} A_S^* A_S \, \sigma(m_{KK})^* \sigma(m_{KK}) \, dm_{KK} &= |A_S|^2 \,. \end{aligned} \tag{10.5}$$

The effective coupling factor is computed to be $k_{SP} = 0.54$, assuming that the P-wave is described by a relativistic Breit–Wigner distribution and the S-wave amplitude by a constant. Different S-wave lineshapes are found leading to very similar values of k_{SP}, with a variation no larger than ≈2%. In general k_{SP} is observed to be mostly dependent on the mass range and not on the specific lineshapes used.

A systematic uncertainty is evaluated by generating pseudo-experiments with $k_{SP} = 0.54$ applied to the $i = 8, 9, 10$ terms in Tab. 3.1 related to the S- and P-wave interference. Pull distributions are used to evaluate the systematic uncertainty, as described in the "model bias" paragraph. The only parameters whose total uncertainty are affected significantly by this approximation are $|A_S|^2$ and Δm_s.

10.5 EFFICIENCY PARAMETRIZATION

10.5.1 *Angular efficiency*

The precision on the angular efficiency parametrization is limited by the number of MC events used to estimate the efficiency functions. A related systematic uncertainty is estimated by generating one thousand efficiency histograms by randomly *sampling* the reference one. Each time the fit to data is repeated after reestimating the efficiency function and the RMS of the obtained physics parameter distributions is taken as the systematic uncertainty[5]. The CPV observables ϕ_s and $|\lambda|$ are particularly affected by this systematic uncertainty.

4 In this analysis: $m_{1,2} = m^{\mathrm{w.a.}}_{\phi(1020)} \mp 10$ MeV.

5 The RMS is used, instead of the standard deviation obtained by a Gaussian fit, to take into consideration more conservatively outliers in the distribution.

The angular efficiency is also heavily affected by differences between data and simulation. A dedicated systematic uncertainty is studied in Section 10.7.

10.5.2 *Proper decay length*

The finite MC statistics used in parametrizing the ct efficiency is propagated to the final results by varying the parameters of the efficiency function within their statistical uncertainties accounting for correlations. The RMS of the distribution of each extracted physics parameter is taken as the corresponding systematic uncertainty.

The choice of the efficiency model is tested by repeating the fit using the efficiency histogram, instead of the smooth efficiency function, and taking the difference from the nominal result as the uncertainty.

The uncertainty values are reported in Tab. 10.5. The proper decay length efficiency is found to be the major source of systematic uncertainty for the two observables most relying on its precise parametrization, that is $\Delta\Gamma_s$ and Γ_s.

Table 10.5: Summary of the systematic uncertainties related to the proper decay length efficiency parametrization. The total systematic uncertainty is obtained as the quadratic sum of the individual contributions.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Simulation statistics	0.2	0.0025	0.001	0.0001	0.0014	0.0006	0.0009	$< 10^{-3}$	$< 10^{-3}$	0.001	0.001
Efficiency model	0.3	0.0057	0.001	0.0001	0.0017	0.0013	0.0022	0.001	$< 10^{-3}$	0.002	0.002
Total uncertainty	0.3	0.0062	0.001	0.0002	0.0022	0.0014	0.0023	0.001	0.001	0.002	0.002

10.6 PROPER DECAY LENGTH RESOLUTION

Two systematic uncertainties related to the proper decay length resolution are evaluated.

First, the impact of possible differences between the scaling factor κ_{ct} in data and simulation are evaluated by varying κ_{ct} by $\pm 10\%$, repeating the fit, and taking the largest difference from the nominal result as the uncertainty. The 10% window is estimated from a data-to-simulation comparison of the observed Gaussian distribution. As discussed in Section 9.2.2.1, the data cannot be reliably used to estimate κ_{ct}, hence only a rough comparison can be performed.

Next, the impact of assuming that κ_{ct} is independent of the proper decay length is tested by parametrizing κ_{ct} as a function of ct. As shown in Fig. 9.5, in fact the scaling κ_{ct} increase at very large values ct. A related systematic uncertainty is assigned with the same methodology used to evaluate the "model assumption" systematic uncertainties[6], using the $\kappa_{ct}(ct)$ parametrization as an alternative hypothesis.

The values of the two uncertainties are reported in Tab. 10.6. The parameters most impacted by these systematic uncertainties are the CP-violating phase ϕ_s and the mass difference Δm_s. However, in both cases, the estimated systematic effects are negligible with respect to the statistical uncertainties on the parameters.

Table 10.6: Summary of the systematic uncertainties related to the proper decay length resolution. The dashes (—) mean that the corresponding uncertainty is either not applicable or it was not evaluated. The total systematic uncertainty is obtained as the quadratic sum of the individual contributions.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Data/simulation difference	2.5	0.0008	0.015	0.0009	0.0005	0.0007	0.0009	0.007	0.006	0.025	0.022
Dependence on ct	2.5	—	0.014	0.0012	0.0002	—	—	0.001	0.001	0.001	—
Total uncertainty	3.5	0.0008	0.021	0.0015	0.0006	0.0007	0.0009	0.007	0.006	0.025	0.022

10.7 DIFFERENCES BETWEEN DATA AND SIMULATIONS

The efficiency parametrization is found to be very sensitive to the muon and kaon p_Ts, and B^0_s meson pseudorapidity distributions. Since the efficiency is parametrized entirely on MC samples a systematic uncertainty is assigned to cover the differences between data and simulation in each of these variables. The effect is evaluated by reweighting the simulated distributions in each variable to agree with the data. These weights are then applied to the simulated samples used to estimate the ct and angular efficiencies, which are then recomputed. The fit is repeated in each case and the differences from the nominal result are taken as the systematic uncertainty. The three individual systematic uncertainties are reported in Tab. 10.7.

The p_T of the muons is observed to be modeled extremely well in simulations and the related uncertainty is found to be negligible. On the other hand, small differences between data and simulations are observed for the kaon p_T and B^0_s meson pseudorapidity distributions. The respective systematic uncertainties are found to be negligible for the physics parameter of more interest, such as ϕ_s and $\Delta\Gamma_s$, but quite large for the squared amplitudes $|A_0|^2$ and $|A_\perp|^2$. For these two parameters, the difference between data and simulations is in fact the leading source of systematic uncertainty.

6 See Section 10.2.1.

Table 10.7: Summary of the systematic uncertainties related to the difference between data and simulations. The rows indicated the reweighted distribution. The total systematic uncertainty is obtained as the quadratic sum of the individual contributions.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
$p_T(K^+, K^-)$	0.6	0.0006	0.003	0.0003	0.0002	0.0032	0.0020	0.005	0.006	0.003	0.018
$p_T(\mu^+, \mu^-)$	0.1	$< 10^{-4}$	$< 10^{-3}$	$< 10^{-4}$	$< 10^{-4}$	0.0001	$< 10^{-4}$	$< 10^{-3}$	0.001	$< 10^{-3}$	$< 10^{-3}$
$\eta(B_s^0)$	0.1	0.0005	0.003	$< 10^{-4}$	0.0002	0.0031	0.0021	0.004	0.002	0.007	0.022
Total uncertainty	0.6	0.0008	0.004	0.0003	0.0003	0.0044	0.0029	0.007	0.007	0.007	0.028

10.8 FLAVOR TAGGING

Several sources of systematic uncertainties related to the flavor tagging techniques are considered.

The statistical uncertainties associated with the ω_{evt} calibration are propagated to the final results by varying the parameters of the mistag probability calibration curves within their statistical uncertainties. For each variation, a new calibration curve is produced and the data are refitted. The RMS of each fitted parameter distribution is then taken as the corresponding systematic uncertainty.

The effect of assumptions on the calibration function shape used to fit the calibration data is tested by repeating the fit using a third-order polynomial[7] and taking the difference with respect to the nominal results as the systematic uncertainty.

Finally, the effect of the assumption that the tagging algorithm applied to $B_s^0 \to J/\psi\,\phi$ and $B^+ \to J/\psi\,K^+$ channels has the same mistag calibration is tested by measuring the differences in calibration parameters between the two channels in simulated samples and propagating it to the final results.

The values of the individual uncertainties are reported in Tab. 10.8. The systematic uncertainties related to the flavor tagging are found to be negligible for all physics parameters. This is a remarkable result when taking into account that more than half of the B_s^0 candidates are flavor tagged.

Table 10.8: Summary of the systematic uncertainties related to flavor tagging. The total systematic uncertainty is obtained as the quadratic sum of the individual contributions.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Calibration statistics	$< 10^{-3}$	$< 10^{-4}$	0.001	0.001	$< 10^{-4}$	0.0003	$< 10^{-4}$	$< 10^{-3}$	$< 10^{-3}$	0.002	0.001
B_s^0/B^+ difference	$< 10^{-4}$	$< 10^{-4}$	0.001	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-3}$	$< 10^{-3}$	0.003	$< 10^{-3}$
Calibration model	0.5	$< 10^{-4}$	0.006	0.0001	$< 10^{-4}$	0.0001	$< 10^{-4}$	$< 10^{-3}$	0.001	0.006	$< 10^{-3}$
Total uncertainty	0.5	$< 10^{-4}$	0.006	0.0002	$< 10^{-4}$	0.0003	$< 10^{-4}$	$< 10^{-3}$	0.001	0.007	0.001

7 In the nominal fit the calibration function is a first-order polynomial.

10.9 SUMMARY OF SYSTEMATIC UNCERTAINTIES

The systematic uncertainties are summarized in Tab. 10.9. The precision on the majority of the physics parameters is found to be statistically limited, with most of the systematic uncertainties being a fraction of the statistical ones. Notable exceptions are the squares of amplitudes $|A_0|^2$ and $|A_\perp|^2$, for which the systematic and statistical uncertainties are comparable, and the squares of the S-wave amplitude $|A_S|^2$, whose uncertainty is dominated by the systematic effect due to the S-P-wave different m_{KK} dependence not being accounted for in the fit model. The final systematic uncertainty on the CP-violating phase ϕ_s is estimated to be 10 mrad, one-fifth of the statistical one.

Table 10.9: Summary of the systematic uncertainties for the physics parameters. The dashes (—) mean that the corresponding uncertainty is either not applicable or it was not evaluated (see text for details). The total systematic uncertainty is obtained as the quadratic sum of the individual contributions. Statistical uncertainties are also presented to ease comparisons. The names of the various sources are defined in the text.

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar\mathrm{ps}^{-1}$]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Statistical uncertainty	50	0.014	0.10	0.026	0.0042	0.0047	0.0063	0.0077	0.12	0.16	0.083
Model bias	7.9	0.0019	—	0.0035	0.0005	0.0002	0.0012	0.001	0.020	0.016	0.006
$P(\sigma_{ct})$ uncertainty	$<10^{-1}$	0.0019	0.028	0.0004	0.0008	0.0006	0.0008	0.001	0.001	0.002	0.005
Model assumptions	—	—	—	0.0046	0.0003	—	0.0013	0.001	0.017	0.019	0.011
Sig./bkg. ω_{evt} difference	3.0	—	—	—	0.0005	—	0.0008	—	—	—	0.006
Peaking background	0.3	0.0008	0.011	$<10^{-4}$	0.0002	0.0005	0.0002	0.003	0.005	0.007	0.011
S-P wave interference	—	0.0010	0.019	—	0.0005	0.0005	—	0.013	—	0.019	0.019
Angular efficiency	3.8	0.0006	0.007	0.0057	0.0002	0.0008	0.0010	0.002	0.006	0.015	0.015
ct efficiency	0.3	0.0062	0.001	0.0002	0.0022	0.0014	0.0023	0.001	0.001	0.002	0.002
ct resolution	3.5	0.0008	0.021	0.0015	0.0006	0.0007	0.0009	0.007	0.006	0.025	0.022
Data/simulation difference	0.6	0.0008	0.004	0.0003	0.0003	0.0044	0.0029	0.007	0.007	0.007	0.028
Flavor tagging	0.5	$<10^{-4}$	0.006	0.0002	$<10^{-4}$	0.0003	$<10^{-4}$	$<10^{-3}$	0.001	0.007	0.001
Total systematic uncertainty	10.0	0.0070	0.032	0.0083	0.0026	0.0049	0.0045	0.016	0.028	0.045	0.048

11 FINAL RESULTS AND COMBINATION

The final results of the CP violation analysis are reported in Tab. 11.1. After adding the systematic uncertainties in quadrature the following value of the CP-violating phase ϕ_s is measured:

$$\phi_s = -11 \pm 50\,(\text{stat}) \pm 10\,(\text{syst})\ \text{mrad},$$

which is consistent with both the SM prediction $\phi_s^{\text{SM}} = -36.96\,^{+0.72}_{-0.84}$ mrad [44] and with no CP in the interference between decay and mixing in $B_s^0 \to J/\psi\,\phi$ decays ($\phi_s = 0$ mrad). The decay width difference between the two B_s^0 mass eigenstates is measured:

$$\Delta\Gamma_s = 0.114 \pm 0.014\,(\text{stat}) \pm 0.007\,(\text{syst})\ \text{ps}^{-1},$$

a value consistent with the latest SM prediction $\Delta\Gamma_s^{\text{SM}} = 0.091 \pm 0.013\ \text{ps}^{-1}$ [75]. These measurements are both consistent with the latest world-average values [35]:

$$\phi_s^{\text{w.a.}} = -51 \pm 23\ \text{mrad},$$
$$\Delta\Gamma_s^{\text{w.a.}} = 0.085 \pm 0.004\ \text{ps}^{-1}.$$

The CP observable $|\lambda|$ is measured to be $|\lambda| = 0.972 \pm 0.026\,(\text{stat}) \pm 0.008\,(\text{syst})$, consistent with no direct CP violation in $B_s^0 \to J/\psi\,\phi$ decays ($|\lambda| = 1$) and the world average value $|\lambda|^{\text{w.a.}} = 1.012 \pm 0.017$ [34]. The average decay width B_s^0 mass is determined to be $\Gamma_s = 0.6531 \pm 0.0042\,(\text{stat}) \pm 0.0026\,(\text{syst})\ \text{ps}^{-1}$, consistent with the current world-average value $\Gamma_s^{\text{w.a.}} = 0.6600 \pm 0.0016\ \text{ps}^{-1}$ [34]. The mass difference between the heavy and light mass eigenstates of the B_s^0 meson is measured to be $\Delta m_s = 17.51\,^{+0.10}_{-0.09}\,(\text{stat}) \pm 0.03\,(\text{syst})\ \hbar\text{ps}^{-1}$, consistent with the theoretical prediction $\Delta m_s^{\text{SM}} = 18.77 \pm 0.86\ \hbar\text{ps}^{-1}$ [75], and in slight tension with the current world-average value $\Delta m_s^{\text{w.a.}} = 17.749 \pm 0.020\ \hbar\text{ps}^{-1}$ [34]. The difference between the measured and world-average values of Δm_s divided by the combined uncertainty is -2.3σ. The uncertainties in the measurement of all these parameters are dominated by the statistical component.

The comparison between the results of this measurement, the world-average values, and the theoretical predictions are summarized in Tab. 11.2. It is worth to mention that this analysis represents the first measurement with CMS data of the mass splitting Δm_s and of the direct CP observable $|\lambda|$.

Table 11.1: Final results of the CPV analysis. The physics parameter values are extracted with the UML fit to the combined data sample as described in Chapter 9. Statistical uncertainties are obtained from the increase in $-\log\mathcal{L}$ by 0.5, whereas systematic uncertainties are described in Chapter 10 and summarized in Tab. 10.9. The values are converted to the units of measure used by the Particle Data Group [34].

Parameter	Fit value	Stat. uncer.	Syst. uncer.
ϕ_s [mrad]	-11	± 50	± 10
$\Delta\Gamma_s$ [ps^{-1}]	0.114	± 0.014	± 0.007
Δm_s [$\hbar$ps^{-1}]	17.51	$^{+0.10}_{-0.09}$	± 0.03
$\lvert\lambda\rvert$	0.972	± 0.026	± 0.008
Γ_s [ps^{-1}]	0.6531	± 0.0042	± 0.0026
$\lvert A_0\rvert^2$	0.5350	± 0.0047	± 0.0049
$\lvert A_\perp\rvert^2$	0.2337	± 0.0063	± 0.0045
$\lvert A_S\rvert^2$	0.022	$^{+0.008}_{-0.007}$	± 0.016
$\delta_\parallel$ [rad]	3.18	± 0.12	± 0.03
$\delta_\perp$ [rad]	2.77	± 0.16	± 0.05
$\delta_{S\perp}$ [rad]	0.221	$^{+0.083}_{-0.070}$	± 0.048

Table 11.2: Comparison of the results of this measurement with the available world-average values [34, 35] and theory predictions [44, 75]. For the measured values the quoted uncertainties are obtained as the quadratic sum of the statistical and systematic components. The dashes (–) represents unavailable quantities.

Parameter	Measured value	World-average	Theory prediction
ϕ_s [mrad]	-11 ± 51	-51 ± 23	$-36.96\,^{+0.72}_{-0.84}$
$\Delta\Gamma_s$ [ps^{-1}]	0.114 ± 0.015	0.085 ± 0.004	0.091 ± 0.013
Δm_s [$\hbar$ps^{-1}]	17.51 ± 0.10	17.749 ± 0.020	18.77 ± 0.86
$\lvert\lambda\rvert$	0.972 ± 0.027	1.012 ± 0.017	1
Γ_s [ps^{-1}]	0.6531 ± 0.0050	0.6600 ± 0.0016	–
$\lvert A_0\rvert^2$	0.5350 ± 0.0068	0.5199 ± 0.0035	–
$\lvert A_\perp\rvert^2$	0.2337 ± 0.0077	0.245 ± 0.004	–
$\lvert A_S\rvert^2$	0.022 ± 0.018	–	–
$\delta_\parallel$ [rad]	3.18 ± 0.12	3.10 ± 0.06	–
$\delta_\perp$ [rad]	2.77 ± 0.17	2.9 ± 0.4	–
$\delta_{S\perp}$ [rad]	0.221 ± 0.089	–	–

Table 11.3: Summary of the results obtained from CMS at 8 TeV [64].

Parameter	Value	Stat. uncer.	Syst. uncer.
ϕ_s [mrad]	-75	± 97	± 31
$\Delta\Gamma_s$ [ps^{-1}]	0.095	± 0.013	± 0.007
Γ_s [ps^{-1}]	0.6704	± 0.0043	± 0.0055
$\lvert A_0\rvert^2$	0.510	± 0.005	± 0.011
$\lvert A_\perp\rvert^2$	0.243	± 0.008	± 0.012
$\lvert A_S\rvert^2$	0.012	± 0.009	± 0.022
$\delta_\parallel$ [rad]	3.48	± 0.09	± 0.68
$\delta_\perp$ [rad]	2.98	± 0.36	± 0.66
$\delta_{S\perp}$ [rad]	0.37	± 0.12	± 0.18

11.1 COMBINATION WITH PREVIOUS CMS RESULTS

The results presented in this work are in agreement with the earlier CMS result at $\sqrt{s} = 8$ TeV, which can be found in Ref. [64] and are reported in Tab. 11.3. As mentioned before, both measurements are performed with a similar number of events, with the one at $\sqrt{s} = 13$ TeV having a larger fraction of flavor tagged events due to the trigger strategy. This leads to an improvement in the uncertainty in quantities that require tagging, such as ϕ_s, while those whose precision depends on the raw number of events, such as $\Delta\Gamma_s$ and Γ_s, are not improved relative to the 8 TeV result. This conclusion applies only to statistical uncertainties, while the improvement on the systematic ones depends on the parameter.

The results presented in this work and those obtained by CMS at 8 TeV are finally combined. The combination is performed using *Best Linear Unbiased Estimator* (*BLUE*) method [174, 175] as implemented [176, 177] in the ROOT analysis package. BLUE is one of the most widely used combination methods and it is based on the Gauss–Markov theorem [178]. Provided that the estimates are unbiased, the BLUE method guarantees that the combined result has the smallest variance and that is unbiased, i. e., when the procedure is repeated several times the mean of all combined results equals the true value of the observables[1].

The following observables are used in the combination: ϕ_s, $\Delta\Gamma_s$, Γ_s, $|A_0|^2$, $|A_\perp|^2$, $|A_S|^2$, $\delta_\parallel$, $\delta_\perp$ and $\delta_{S\perp}$. Notable absences are the mass splitting Δm_s and direct CP observable $|\lambda|$, which were not measured in the previous CMS work[2]. The two sets of results are combined using their respective statistical and systematic correlation between matrices. The statistical correlations between the parameters obtained in each measurement are taken into account as well as the

1 This holds only if all the inputs are consistent with the underlying multidimensional PDF.
2 In the previous CMS result the mass splitting Δm_s was constrained to the world-average value, while the direct CP observables $|\lambda|$ was fixed to 1.

correlations of the systematic uncertainties discussed in Section 10. Different sources of systematic uncertainties are assumed to be uncorrelated. The systematic uncertainty correlation between the parameters of the 8 TeV result is assumed to be zero. This assumption has been found to not impact the results in a noticeable way. Since the muon tagging, the efficiency evaluation, and a large portion of the fit model are different in the two measurements, the respective systematic uncertainties are treated as uncorrelated between the two sets of result[3]. The statistical correlation between the two measurements is zero as the data samples are independent. In the case of asymmetric uncertainties, the larger one is used.

The combination results and their uncertainties are reported in Tab. 11.4. The combined results for the CP-violating phase and lifetime difference between the two B_s^0 mass eigenstates are:

$$\begin{aligned} \phi_s &= -21 \pm 44\,(\text{stat}) \pm 10\,(\text{syst})\ \text{mrad}, \\ \Delta\Gamma_s &= 0.1032 \pm 0.0095\,(\text{stat}) \pm 0.0048\,(\text{syst})\ \text{ps}^{-1}. \end{aligned}$$

The two-dimensional ϕ_s-$\Delta\Gamma_s$ likelihood contours at 68% confidence level (CL) for the individual and combined results, as well as the SM prediction, are shown in Fig. 11.1. The combined results are in agreement both with the SM predictions and with no CP violation in the interference between decay and mixing in $B_s^0 \to J/\psi\,\phi$ ($\phi_s = 0$ mrad). The other parameters show good agreement with the theoretical predictions and with the world-average values.

Table 11.4: Result of the combination between the 8 and 13 TeV results. The combination is performed with BLUE, assuming no correlation between the two sets of results.

Parameter	Fit value	Stat. uncer.	Syst. uncer.
ϕ_s [mrad]	-21	± 44	± 10
$\Delta\Gamma_s$ [ps^{-1}]	0.1032	± 0.0095	± 0.0048
Γ_s [ps^{-1}]	0.6590	± 0.0032	± 0.0023
$\|A_0\|^2$	0.5289	± 0.0038	± 0.0041
$\|A_\perp\|^2$	0.2393	± 0.0050	± 0.0037
$\|A_S\|^2$	0.016	± 0.006	± 0.013
$\delta_\parallel$ [rad]	3.19	± 0.12	± 0.04
$\delta_\perp$ [rad]	2.78	± 0.15	± 0.06
$\delta_{S\perp}$ [rad]	0.238	± 0.078	± 0.046

3 This assumption is tested by repeating the combination assuming that the "model bias" systematic uncertainty (the leading systematic uncertainty on ϕ_s) is fully correlated between the two sets of measurements. The impact on the combined results is found to be negligible.

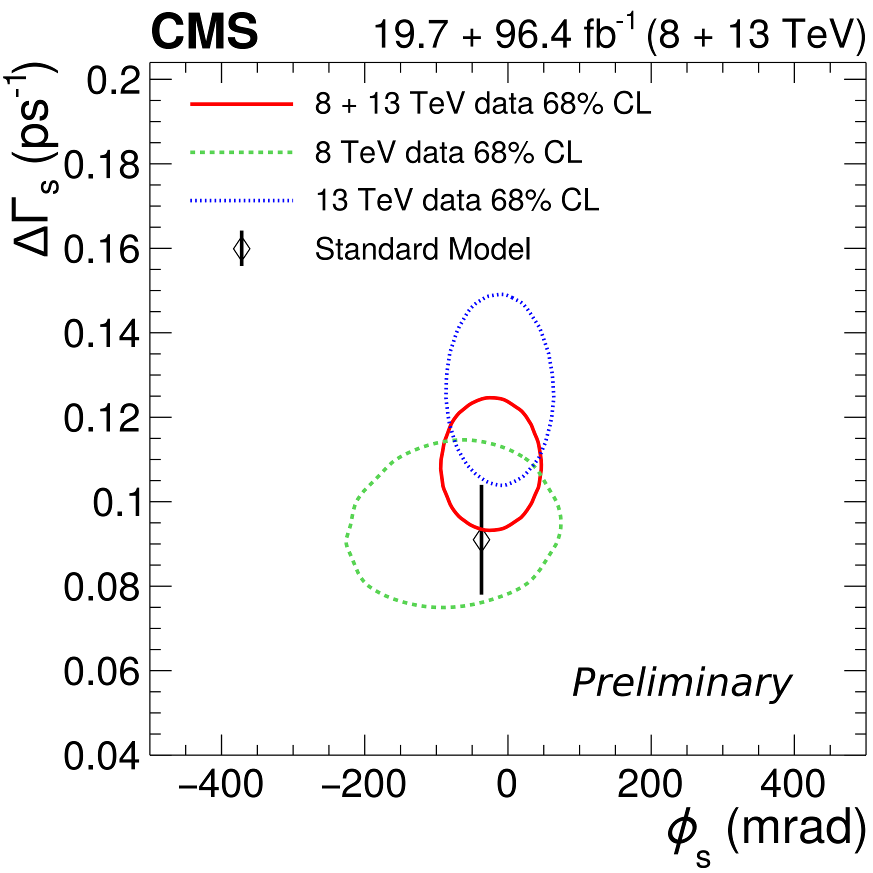


Figure 11.1: The two-dimensional likelihood contours at 68% CL in the $(\phi_s, \Delta\Gamma_s)$ plane, for the CMS 8 TeV (dashed line) [64], 13 TeV (dotted line), and combined (solid line) results. The SM prediction is shown with the diamond marker [44, 75]. The contours are obtained with a likelihood scan, where only statistical uncertainties are taken into account. The contours for the individual results are obtained with likelihood scans, which are used to obtain the combined contour. The contours only account for the statistical uncertainty and the correlation between the two scanned variables. Figure from Ref. [1] as of 30 November 2020.

12 SUMMARY AND OUTLOOK

This dissertation presents a time-dependent and flavor-tagged angular analysis of the $B^0_s \to J/\psi\,\phi$ decay rate performed with the CMS detector at the LHC using a sample of 48 500 reconstructed $B^0_s \to J/\psi\,\phi \to \mu^+\mu^-\,K^+K^-$ signal events collected in proton-proton collisions at $\sqrt{s} = 13$ TeV and corresponding to an integrated luminosity of $\mathcal{L}_{\text{int}} = 96.4\,\text{fb}^{-1}$. Events are selected with a trigger that requires a $J/\psi \to \mu^+\mu^-$ candidate plus an additional third muon, which is exploited to infer the flavor of the B^0_s meson at production time. An original opposite-side muon tagger based on calibrated deep neural networks has been developed to maximize the analysis sensitivity and to directly estimate the per-event mistag probability. A high tagging power of $\approx$10%, corresponding to a tagging efficiency of $\approx$50%, is achieved, aided by the requirement of an additional muon in the signal sample imposed at the trigger level. The systematic uncertainties related to the flavor tagging algorithm are found to be negligible for all the physics parameters. The detailed tagging performance is reported in Tab. 8.8, while the respective systematic uncertainties are presented in Tab. 10.8.

The CP-violating weak phase ϕ_s, responsible for CP violation in the interference between decay and mixing of $B^0_s \to J/\psi\,\phi$, is measured to be:

$$\phi_s = -11 \pm 50\,(\text{stat}) \pm 10\,(\text{syst})\ \text{mrad},$$

consistent both with the Standard Model predictions $\phi_s = -36.96\,^{+0.72}_{-0.84}$ mrad, the world average-value $\phi_s = -51 \pm 23$ mrad, and the absence of CP violation in the decay-mixing interference ($\phi_s = 0$ mrad).

The decay width (mass) difference $\Delta\Gamma_s$ (Δm_s) between the B^0_s meson mass eigenstates, the average decay width Γ_s, and the $|\lambda|$ parameter, describing the amount of direct CP violation, are measured to be:

$$\begin{aligned}
\Delta\Gamma_s &= 0.114 \pm 0.014\,(\text{stat}) \pm 0.007\,(\text{syst})\ \text{ps}^{-1},\\
\Delta m_s &= 17.51\,^{+\,0.10}_{-\,0.09}\,(\text{stat}) \pm 0.03\,(\text{syst})\ \hbar\text{ps}^{-1},\\
\Gamma_s &= 0.6531 \pm 0.0042\,(\text{stat}) \pm 0.0026\,(\text{syst})\ \text{ps}^{-1},\\
|\lambda| &= 0.972 \pm 0.026\,(\text{stat}) \pm 0.008\,(\text{syst}).
\end{aligned}$$

The results are in agreement with the Standard Model and can be used to further constrain possible new-physics effects in the decay and mixing of B^0_s mesons. A complete list of measured physics parameters, as well as comparisons with the world-average values and theory predictions, is given in Tabs. 11.1 and 11.2.

The results presented in this dissertation are further combined with those obtained by CMS at $\sqrt{s} = 8$ TeV, yielding a value for ϕ_s of:

$$\phi_s = -21 \pm 44\,(\text{stat}) \pm 10\,(\text{syst})\ \text{mrad}\,,$$

significantly more precise than the one from the previous CMS measurement. A complete list of the combined results is given in Tab. 11.4.

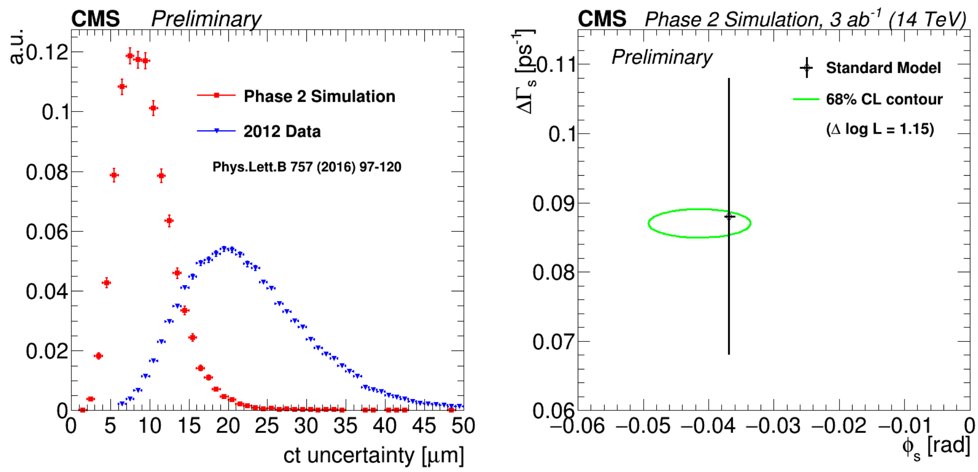
12.1 OUTLOOK

Although both the measured and combined value of the CP-violating phase ϕ_s are consistent with the SM prediction, the interest in more precise measurements remains high. The uncertainty in the measurements of the most interesting physics parameters is largely dominated by the statistical component, meaning that further studies with larger data samples will be able to undoubtedly improve the precision on ϕ_s and the other parameters of the $B^0_s \to J/\psi\,\phi$ decay. Moreover, the tagging strategy used in this analysis prioritizes flavor-sensitive parameters, such as ϕ_s, $|\lambda|$, and Δm_s, while being less optimal for measuring $\Delta\Gamma_s$ and Γ_s with high precision.

For these reasons, the CMS Collaboration is currently working on extending the current work by exploiting a data sample collected with a *non-flavor-dedicated* trigger in the full Run-2 collision data. This trigger sets no requirements on the opposite-side b-hadron, resulting in a high number of selected events which can then be flavor tagged with multiple techniques, such as lepton and jet taggers. This strategy is complementary to the one used in this work and the two are planned to be used in conjunction to maximize the analysis sensitivity. This study is expected to improve significantly the precision of the measurement of all physics parameters and provide more stringent constraints on new-physics contributions.

Combinations of results from multiple experimental collaborations are also used to perform world averages and further probe physics beyond the SM. An uncertainty of $\sigma_{\phi_s} \approx 20$ mrad on the CP-violating weak phase ϕ_s is expected from the combination of the Run-1 and Run-2 results from the ATLAS, CMS, and LHCb collaborations.

The analysis of the $B^0_s \to J/\psi\,\phi$ decay is expected to be statistically limited in all future projections. As such, significant improvements are expected for the planned HL-LHC and Phase-2 CMS upgrades. A projection of sensitivity on ϕ_s using the Phase-2 CMS detector in a $\mathcal{L}_{\text{int}} = 3\ \text{ab}^{-1}$ data sample is available in Refs. [179, 180]. Figure 12.1 shows the simulated proper decay length uncertainty of the Phase-2 CMS detector and the ϕ-$\Delta\Gamma_s$ sensitivity at the end of Phase-2. The sensitivity on ϕ_s is expected to be in the 5–6 mrad range, which improves the current world average uncertainty by a factor of five and will either establish new-physics effects or put very stringent constraints on them.



(a) Proper decay length uncertainty distribution in Phase-0 and Phase-2 simulations.

(b) Expected ϕ_s sensitivity in Phase-2.

Figure 12.1: *a)*: proper decay length uncertainty distribution in 2012 data (blue) and Phase-2 MC (red) samples. The better performance of Phase-2 with respect to 2012 data is due to the upgraded Phase-2 tracker. Note that the performance in 2012 data does not represent that of this work, as the upgraded Phase-1 pixel tracker was installed in the meanwhile (see Section 4.2.2). *b)*: ϕ_s and $\Delta\Gamma_s$ sensitivity in Phase-2, as obtained from MC pseudo-experiments. The contour combines statistical and systematic uncertainties. The black cross represents the SM expectation (the prediction on $\Delta\Gamma_s$ is not up-to-date). Figure from Ref. [179].

Part V

APPENDIX

A MUON DISCRIMINATOR AGAINST LIGHT HADRONS

Long-lived light hadrons, like kaons and pions, can either decay in flight into muons or survive the calorimeter system to leave a signal in the muon spectrometers, and therefore being reconstructed as muons by the Particle Flow algorithm. If selected in a lepton flavor tagging algorithm, these *fake* muons degrade the tagging performance, by carrying no charge correlation with the signal b-meson.

The "DNN against light hadrons" (DALIHA) discriminator is an MVA method that combines variables related to the muon kinematics and reconstruction quality and is specifically trained to discriminate soft muons from b-hadrons from fake muons from light hadron.

A.1 TRAINING CANDIDATES

The DNN is trained with simulated muon candidates reconstructed as *global muons* by the Particle Flow algorithm. Training candidates are extracted from all the MC samples listed in Tab. 7.5, with the exception of the for $\Lambda_b^0 \to J/\psi\, K^- p$ data sets. Different simulated processes are used to ensure a training data set as unbiased as possible, with respect to the specific signal b-meson decay.

To avoid biases due to generation filters, all tracks associated with the specific simulated process are explicitly excluded by the training data set[1]. All training muons are therefore also *opposite side* muons. The specific *simulated process*[2] is identified using generation information and selected with requirements on the generated momenta and pseudorapidity values. Muons from $J/\psi \to \mu^+\mu^-$ decays are required to be generated with $p_T > 3.5$ GeV and $|\eta| < 2.4$, while the signal b-meson is required having $p_T > 10$ GeV and $|\eta| < 2.5$. This selection, summarized in Tab. A.1, ensures signal b-mesons in the same kinematic region of the B_s^0 candidates used in the CPV analysis.

Training candidates are required to have $p_T > 2$ GeV and $|\eta| < 2.4$. All selected candidates are required to be matched to a gen particle with the following selection:

$$\sqrt{\Delta\eta^2_{\mu,\text{gen}} + \Delta\phi^2_{\mu,\text{gen}}} < 0.12\,, \tag{A.1}$$

$$\frac{p_T^{\text{gen}} - p_T^{\mu}}{p_T^{\text{gen}}} < 0.3\,. \tag{A.2}$$

1 For example in $B_s^0 \to J/\psi\,\phi$ events the two muons from the $J/\psi \to \mu^+\mu^-$ decay and the two tracks from the $\phi \to K^+K^-$ decay are excluded.

2 The simulated process refers to the specific decay forced in each simulated event (see Tab. 7.5).

In case of multiple gen candidates the one that minimize $\Delta R_{\eta,\phi}$ is chosen. Muon candidates matched to genuine gen muons are denominated *Signal*, while candidates matched to other species (mostly kaons and pions) are denominated *Background* or *Fakes*. To improve the performance in flavor tagging algorithms, *Signal* muons are also required to originate from a b-hadron ancestor. The selection requirements used to construct the training data set are summarized in Tab. A.2.

The total number of genuine and fake muon candidates selected across all data samples is 6 626 306, and 451 973, respectively.

Table A.1: Signal b-meson selection for the DALIHA discriminator training data set. All requirements are applied to gen quantities.

Variable	Requirement
Gen matched	✓
$p_T(\mu_{gen})$	> 3.5 GeV
$\lvert\eta(\mu_{gen})\rvert$	< 2.4
$p_T(B_{gen})$	> 10 GeV
$\lvert\eta(B_{gen})\rvert$	< 2.5

Table A.2: Selection requirements for the DALIHA discriminator training data set.

Object	Requirement
Muon reconstruction	*Global muon*
Muon track quality	*High Purity*
p_T	> 2 GeV
$\lvert\eta\rvert$	< 2.4
Gen matching	$\Delta R_{\eta,\phi} < 0.12$ $\Delta p_T^{\mu,gen}/p_T < 0.3$
Signal definition	matched to a gen muon has a b-hadron ancestor
Background definition	not matched to a gen muon

TRAINING DATA SET The data set is split into *training*, *validation* and *testing* subsets with respective fractions of 50%, 25% and 25%. The training data set is used to train the DNN model, the validation data set is used to evaluate the performance at each training epoch and trigger early stopping, while the testing data set is used to evaluate the performance and overfitting of the trained model. In order to balance the *Signal*:*Background* ratio in the training data set, all *Signal* muons have been assigned a weight of 0.1.

A.2 INPUT FEATURES

The DNN model uses several variables related to the muon kinematics and reconstruction quality as input features. These are[3]:

- p_T: the muon transverse momentum,
- η: the muon pseudorapidity
- *PFiso*: the muon relative isolation[4],
- *muoSegmComp*: a measure of the track's compatibility with the muon hypothesis
- *muoChi2LM*: the similarity measure of *Local Momentum*[5],
- *muoChi2LP*: the similarity measure of *Local Position*[6],
- *muoGlbTrackTailProb*: the probability of the χ^2 of the muon's global track fit being larger than the one observed,
- *muoTimeAtIpInOut*: time of arrival at the interaction point for muons moving inside-out assuming $\beta = 1$,
- *muoTimeAtIpInOutErr*: uncertainty in the time of arrival at the interaction point for muons moving inside-out assuming $\beta = 1$,
- *muoGlbDeltaEtaPhi*: squared difference in η and ϕ of the stand alone muon track and the tracker track on the common surface during track matching,
- *muoIValFrac*: number of valid hits of the inner track divided by the total number of hits and missing hits of the inner track,
- *muoLWH*: amount of tracker layers with hits,

3 As these are very technical variables, only a brief description is provided.
4 The isolation is evaluated as in Eq. 8.35.
5 A scalar measuring the differences between the momentum vectors of the stand-alone muon track and of the tracker track.
6 A scalar measuring the differences between the position of the last hit of the tracker track and the first hit of the stand-alone muon track.

- *muoValPixHits*: amount of valid hits in the pixel detector,
- *muoNTrkVHits*: amount of valid hits in the tracker,
- *muoVMuHits*: amount of valid hits in the muon system,
- *muoNumMatches*: number of muon stations containing matched segments,
- *muoVMuonHitComb* $= \nu DT/2 + \nu CSC_{max} + \nu RPC$, where νDT is the number of valid hits in the drift tubes, νCSC_{max} max is the number of valid hits in the cathode strip chambers limited to a maximum of 6 hits per station and νRPC is the number of valid hits in the resistive plate chambers,
- *muoTrkKink*: output of the kink algorithm applied to the global muon's inner track,
- *muoGlbKinkFinderLOG* $= \log(2 + glbKinkFinder)$ where *glbKinkFinder* is the output of the kink algorithm applied to the global muon's track,
- *muoOuterChi2*: χ^2 of the muon outer track divided by the number of degrees of freedom of the fit,
- *muoInnerChi2*: χ^2 of the inner track divided by the number of degrees of freedom of the fit,
- *muoTrkRelChi2*: sum of χ^2 estimates of the hits in the tracker with respect to the global muon track,
- *muoStaRelChi2*: sum of χ^2 estimates of the hits in the muon system with respect to the global muon track,
- *muoGNchi2*: χ^2 of the global muon track divided by the number of degrees of freedom of the fit,
- *muoQprod*: product of the tracker track's charge and the standalone muon track's charge.

The input features distributions for genuine muons and fakes are shown in Figures A.1 and A.2.

FEATURE PREPROCESSING The input features are preprocessed in order to increase the performance of the DNN method. The variables are first normalized in the range $[-1, 1]$, then their distribution is transformed into a Gaussian ("*Gaussianisation*") and finally normalized again. The normalization is helpful to the training process because it ensures that both positive and negative values are used as inputs, which makes learning more flexible, and that the network's regards all features to a similar extent without favoritism. Gaussianisation is found to speed up learning and to improve the convergence properties

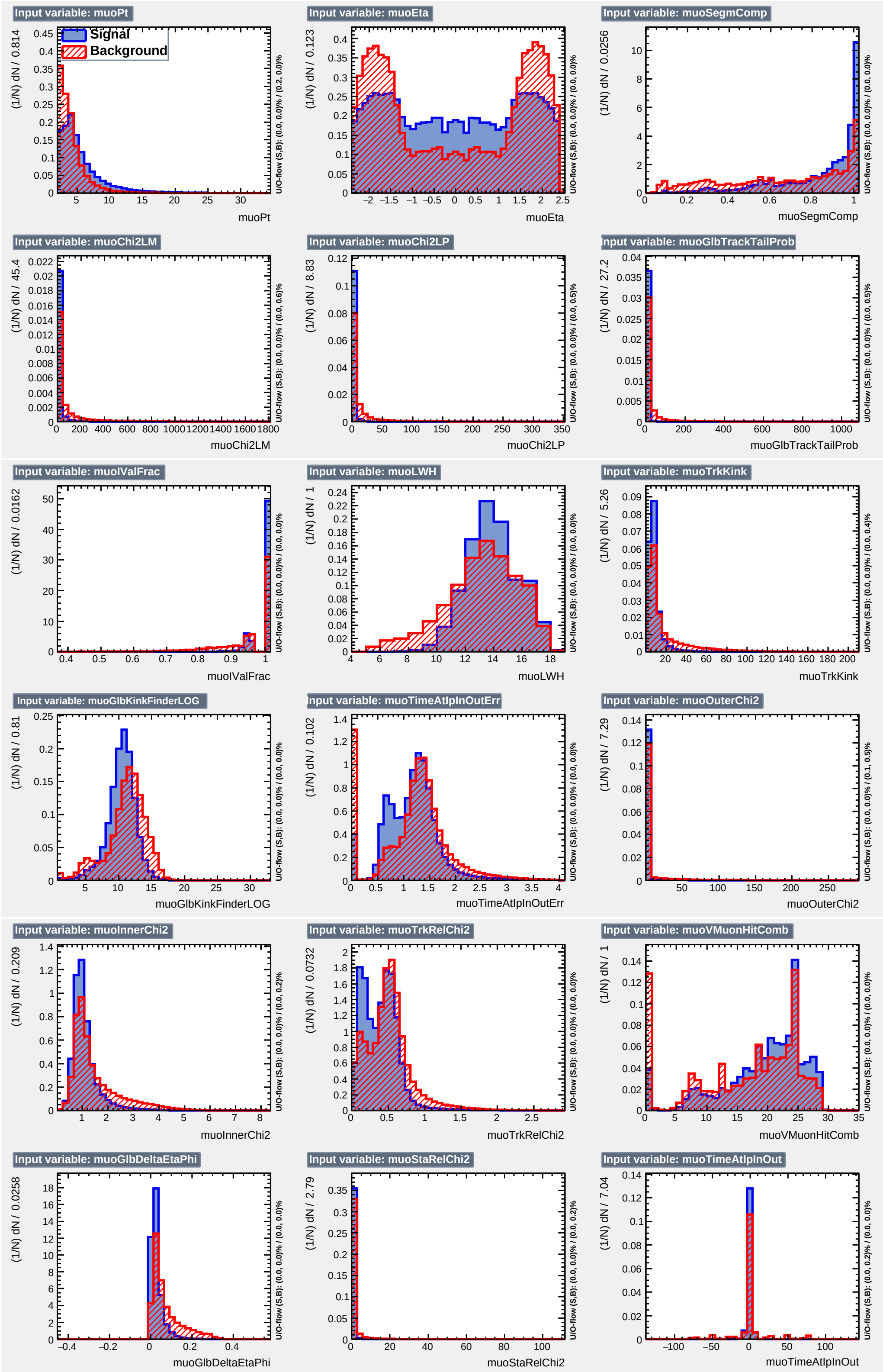


Figure A.1: DALIHA discriminator input features distribution for genuine muons (blue) and fakes (red). Part 1.

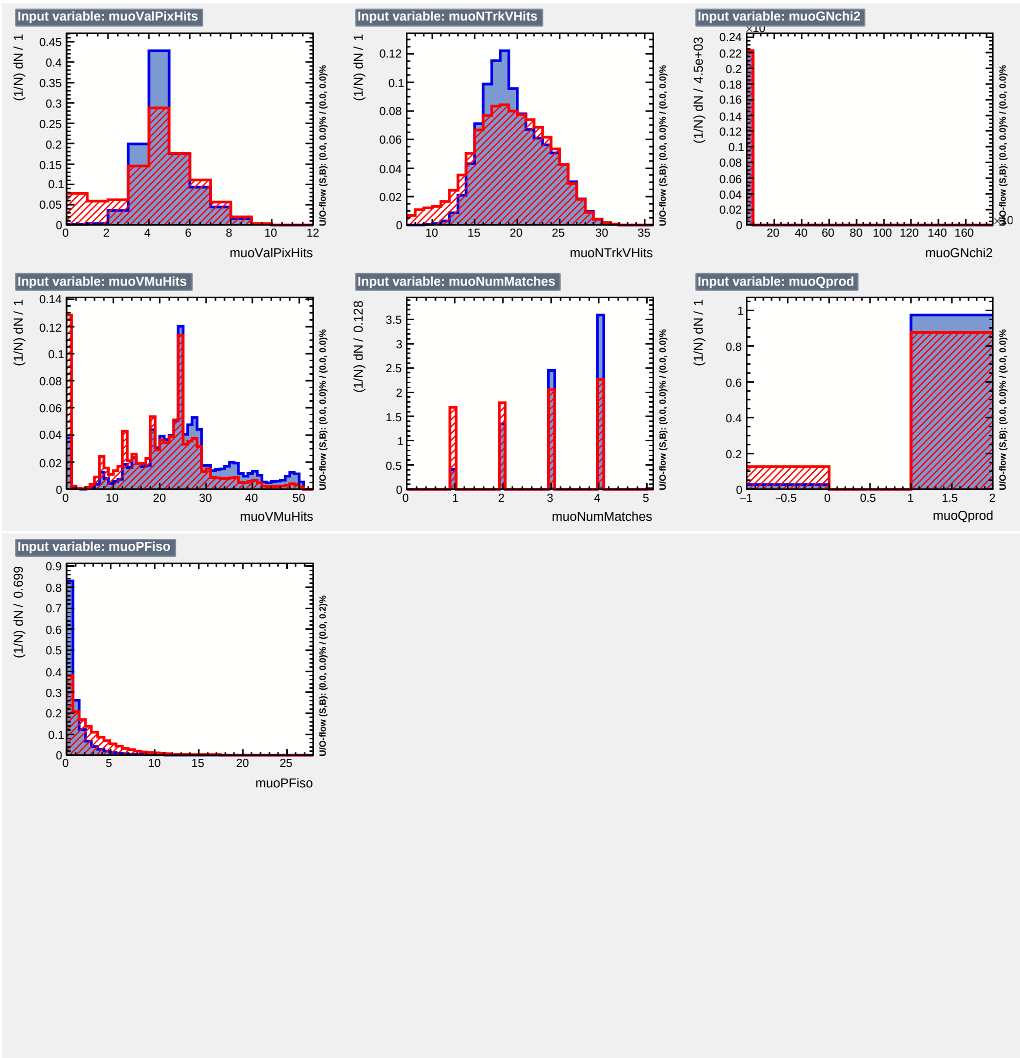


Figure A.2: DALIHA discriminator input features distribution for genuine muons (blue) and fakes (red). Part 2.

when combined with a Gaussian random initialization of the model weights. All preprocessing is handled with the TMVA toolkit and all input features are preprocessed except for *muoQprod* since this variable can assume only two values (± 1).

A.3 DEEP NEURAL NETWORK MODEL AND TRAINING

The DNN consists of a KERAS model trained within the TMVA framework. In order to define the model with the best *Signal* vs *Background* separation, the DNN hyperparameters are optimized with a grid scan. The parameter configuration that maximizes the area under the *Receiver Operating Characteristic* (*ROC*) curve, with no observable overfitting, is chosen for the definitive model. Between equally performing models, the simpler one is chosen. The tested values in the grid scan and the best configuration are reported in Tab. A.3.

The optimal model is found to be three *Dense* layers of 200 neurons, each followed by a *Dropout* layers with 20% dropout probability. The activation function of choice is *ReLU*, while the loss function is the *cross-entropy*.

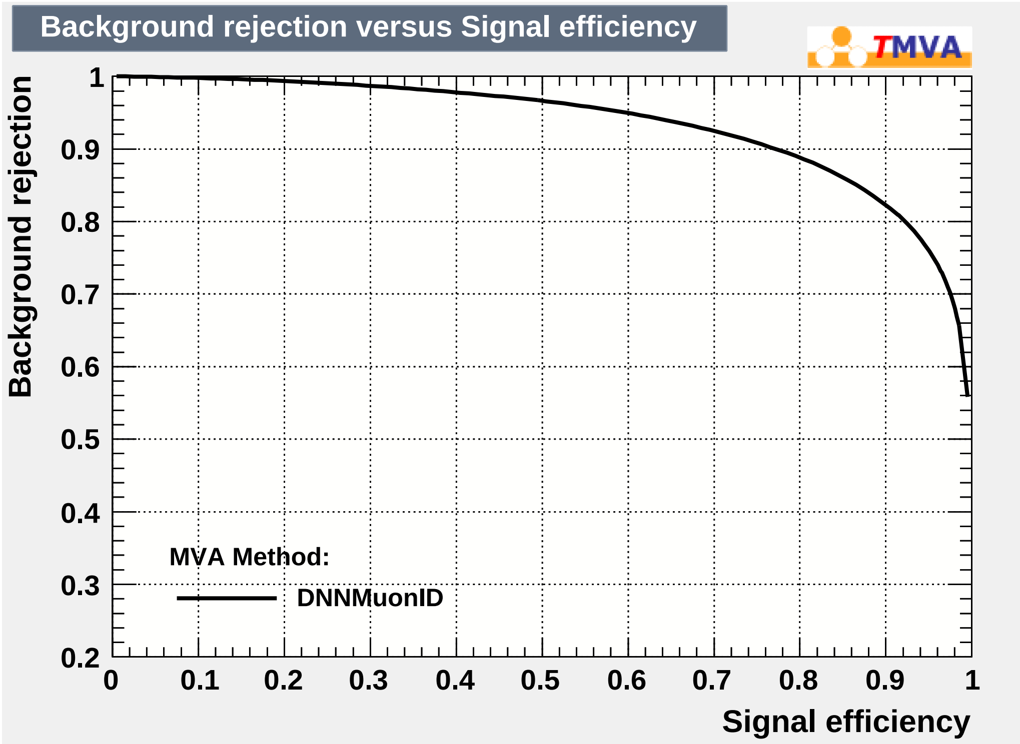
The model is trained using the *Adam* optimizer and a batch size of 1024 events. The maximum amount of epochs is 40, however, to avoid overfitting the training is stopped after 10 epochs with no accuracy improvement, and the model with the smaller loss is kept. The full list of training hyperparameters is reported in Tab. A.4.

Table A.3: DALIHA discriminator model hyperparameters optimized through grid scan and their tested values. The values related to the optimal configuration are underlined.

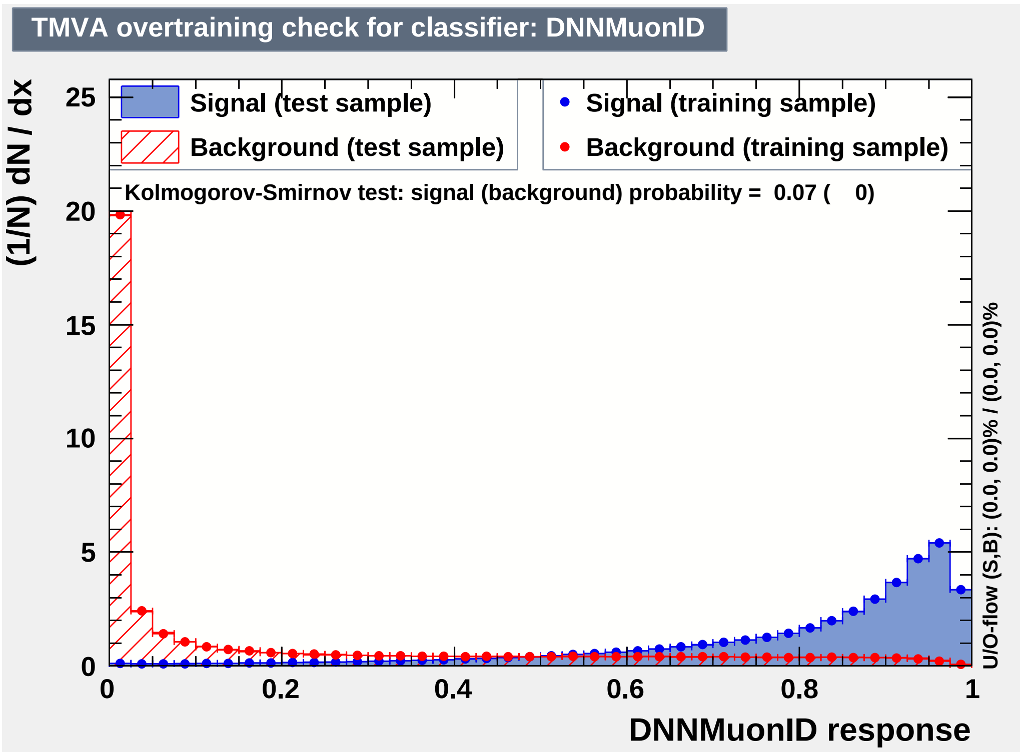
Hyperparameter	Tested values
Number of *Dense* layers	[2, <u>3</u>, 4, 5]
Neurons for each layer	[100, 150, <u>200</u>, 300]
Dropout probability	[0.1, <u>0.2</u>, 0.3, 0.4, 0.5]
Optimizer	[<u>Adam</u>, SDG]

A.4 RESULTS AND PERFORMANCE

The trained model ROC curve and the output score are shown in Fig. A.3. No signs of overfitting are observed at the chosen hyperparameters configuration when comparing the output distributions from the testing and training samples. Three working points are defines as a function of the *Signal* efficiency: *Loose*, *Medium* and *Tight*. The working points efficiency and score cut values are reported in Tab. A.5.



(a) DALIHA discriminator ROC curve



(b) DALIHA discriminator normalized output score

Figure A.3: DALIHA discriminator performance plots: ROC curve (top) and normalized DNN output score (bottom) for genuine muon (blue) and fakes (red) in the training sample (markers) and in the test sample (histograms).

Table A.4: DALIHA discriminator training hyperparameters.

Hyperparameter	Value
Signal weight	0.1
Background weight	1.0
Optimizer	Adam
Optimizer learning rate	0.001
Batch size	1024
Number of training epochs	40
Early stopping epochs	10

Table A.5: Working point for the DALIHA discriminator. The efficiency for genuine muons and fakes are evaluated on global muon candidates in the test sample.

Working Point	Score	Genuine muons efficiency	Fakes efficiency
Loose	$s_{\text{muonID}} > 0.21$	98%	33%
Medium	$s_{\text{muonID}} > 0.40$	95%	23%
Tight	$s_{\text{muonID}} > 0.45$	90%	19%

BIBLIOGRAPHY

[1] CMS Collaboration. "Measurement of the CP-violating phase ϕ_s in the $B_s^0 \rightarrow J/\psi\,\phi(1020) \rightarrow \mu^+\mu^-\,K^+K^-$ channel in proton-proton collisions at $\sqrt{s} = 13$ TeV." July 2020. arXiv: 2007.02434 [hep-ex]. Submitted to Phys. Lett. B.

[2] A. Bragagnolo. "Measurement of the CP-violating phase ϕ_s in the $B_s^0 \rightarrow J/\psi\,\phi(1020) \rightarrow \mu^+\mu^-\,K^+K^-$ channel in collisions at 13 TeV by CMS." In: *CERN EP Seminars*. CERN, Mar. 2020. URL: https://indico.cern.ch/event/868251/.

[3] A. Bragagnolo. "Measurement of the CP-violating phase ϕ_s in the $B_s^0 \rightarrow J/\psi\,\phi$ channel at 13 TeV by CMS." In: *LHCP2020*. Virtual, May 2020. URL: https://indico.cern.ch/event/856696/contributions/3742167/.

[4] G. Fedi, A. Bragagnolo, E. Lusiani, M. Alibordi, P. Behera, A. Boletti, F. Fiori, M. Galanti, F. Ligabue, M. Margoni, F. Palla, P. Ronchese, F. Simonetto, and H. Yarar. "Measurement of the CP-violating phase $\phi_s^{c\bar{c}s}$ in the $B_s^0 \rightarrow J/\psi K^+K^-$ channel using Run-2 data." CMS AN-2019/255.

[5] A. Bragagnolo, M. Galanti, E. Lusiani, G. Fedi, R. Demina, A. Garcia-Bellido, F. Palla, P. Ronchese, F. Simonetto, and H. Yarar. "Flavour tagging for $B_s^0 \rightarrow J/\psi\,\phi$ in Run-2." CMS AN-2019/135.

[6] J. Goldstone, A. Salam, and S. Weinberg. "Broken symmetries." In: *Phys. Rev.* 127 (1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965.

[7] S. Glashow. "Partial symmetries of weak interactions." In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[8] A. Salam and J. C. Ward. "Electromagnetic and weak interactions." In: *Phys. Lett.* 13 (1964), pp. 168–171. DOI: 10.1016/0031-9163(64)90711-5.

[9] M. Gell-Mann. "A schematic model of baryons and mesons." In: *Phys. Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3.

[10] S. Weinberg. "A model of leptons." In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[11] S. Glashow, J. Iliopoulos, and L. Maiani. "Weak interactions with lepton-hadron symmetry." In: *Phys. Rev. D* 2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[12] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet gluon picture." In: *Phys. Lett. B* 47 (1973), pp. 365–368. DOI: 10.1016/0370-2693(73)90625-4.

[13] L. Maiani, A. D. Polosa, and V. Riquer. "$X(3872)$ tetraquarks in B and B_s^0 decays." In: *Phys. Rev. D* 102.3 (2020), p. 034017. DOI: 10.1103/PhysRevD.102.034017. arXiv: 2005.08764 [hep-ph].

[14] LHCb Collaboration. "Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays." In: *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex].

[15] LHCb Collaboration. "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$." In: *Phys. Rev. Lett.* 122.22 (2019), p. 222001. DOI: 10.1103/PhysRevLett.122.222001. arXiv: 1904.03947 [hep-ex].

[16] W. Pauli. "Mathematical contributions to the theory of Dirac matrices." (French). In: *Annals of the Henri Poincaré Institute* 6.2 (1936), pp. 109–136. URL: www.numdam.org/item/AIHP_1936__6_2_109_0/.

[17] M. Gell-Mann. "Symmetries of baryons and mesons." In: *Phys. Rev.* 125 (3 Feb. 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067.

[18] J Goldstone. "Field theories with "superconductor" solutions." In: *Nuovo Cimento* 19.CERN-TH-118 (Aug. 1960), pp. 154–164. DOI: 10.1007/BF02812722.

[19] P. W. Higgs. "Broken symmetries, massless particles and gauge fields." In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[20] P. W. Higgs. "Broken symmetries and the masses of gauge bosons." In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[21] F. Englert and R. Brout. "Broken symmetry and the mass of gauge vector mesons." In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[22] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global conservation laws and massless particles." In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[23] N. Cabibbo. "Unitary symmetry and leptonic decays." In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[24] M. Kobayashi and T. Maskawa. "CP violation in the renormalizable theory of weak interaction." In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[25] B. Pontecorvo. "Inverse beta processes and nonconservation of lepton charge." In: *Sov. Phys. JETP* 7 (1958), pp. 172–173.

[26] Z. Maki, M. Nakagawa, and S. Sakata. "Remarks on the unified model of elementary particles." In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880. DOI: 10.1143/PTP.28.870.

[27] T2K Collaboration. "Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations." In: *Nature* 580.7803 (Apr. 2020), pp. 339–344. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2177-0.

[28] E. Noether. "Invariante variationsprobleme." (German). In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918), pp. 235–257.

[29] A. Sakharov. "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe." In: *Sov. Phys. Usp.* 34.5 (1991), pp. 392–393. DOI: 10.1070/PU1991v034n05ABEH002497.

[30] J. Christenson, J. Cronin, V. Fitch, and R. Turlay. "Evidence for the 2π decay of the K_2^0 meson." In: *Phys. Rev. Lett.* 13 (1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138.

[31] BABAR Collaboration. "Measurement of CP-violating asymmetries in B^0 decays to CP eigenstates." In: *Phys. Rev. Lett.* 86 (12 Mar. 2001), pp. 2515–2522. DOI: 10.1103/PhysRevLett.86.2515.

[32] Belle Collaboration. "Observation of large CP violation in the neutral B meson system." In: *Phys. Rev. Lett.* 87 (9 Aug. 2001), p. 091802. DOI: 10.1103/PhysRevLett.87.091802.

[33] LHCb Collaboration. "Observation of CP violation in Charm decays." In: *Phys. Rev. Lett.* 122.21 (2019), p. 211803. DOI: 10.1103/PhysRevLett.122.211803. arXiv: 1903.08726 [hep-ex].

[34] Particle Data Group. "Review of particle physics." In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[35] Heavy Flavour Averaging Group. "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018." 2019. arXiv: 1909.12524 [hep-ex]. Updated results and plots available at https://hflav.web.cern.ch/.

[36] G. Luders. "On the equivalence of invariance under time reversal and under particle-antiparticle conjugation for relativistic field theories." In: *Kong. Dan. Vid. Sel. Mat. Fys. Med.* 28N5.5 (1954), pp. 1–17.

[37] W. Pauli. "Exclusion principle, Lorentz group and reflection of space-time and charge." In: *Niels Bohr and the Development of Physics*. Ed. by W. Pauli. Pergamon Press (New York), 1955, pp. 30–51.

[38] R. Jost. "Eine bemerkung zum CPT-Theorem." (German). In: *Hel. Phys. Acta* 30 (1957), pp. 409–416.

[39] T. Lee and C.-N. Yang. "Question of parity conservation in weak interactions." In: *Phys. Rev.* 104 (1956), pp. 254–258. DOI: 10.1103/PhysRev.104.254.

[40] C. Wu, E. Ambler, R. Hayward, D. Hoppes, and R. Hudson. "Experimental test of Parity conservation in β decay." In: *Phys. Rev.* 105 (1957), pp. 1413–1414. DOI: 10.1103/PhysRev.105.1413.

[41] J. Lees et al. "Observation of time reversal violation in the B^0 meson system." In: *Phys. Rev. Lett.* 109 (2012), p. 211801. DOI: 10.1103/PhysRevLett.109.211801. arXiv: 1207.5832 [hep-ex].

[42] L. Wolfenstein. "Parametrization of the Kobayashi-Maskawa matrix." In: *Phys. Rev. Lett.* 51 (1983), p. 1945. DOI: 10.1103/PhysRevLett.51.1945.

[43] C. Jarlskog. "Commutator of the quark mass matrices in the standard Electroweak Model and a measure of maximal CP violation." In: *Phys. Rev. Lett.* 55 (1985), p. 1039. DOI: 10.1103/PhysRevLett.55.1039.

[44] The CKMfitter Group. "Predictions of selected flavour observables within the Standard Model." In: *Phys. Rev. D* 84 (2011), p. 033005. DOI: 10.1103/PhysRevD.84.033005. arXiv: 1106.4041. Updated with Summer 2019 results: http://ckmfitter.in2p3.fr/www/results/plots_summer19/num/ckmEval_results_summer19.pdf.

[45] A. Buras, W. Slominski, and H. Steger. "B^0–$\overline{B}^0$ mixing, CP violation and the B meson decay." In: *Nucl. Phys. B* 245 (1984), pp. 369–398. DOI: 10.1016/0550-3213(84)90437-1.

[46] NA31 Collaboration. "First evidence for direct CP violation." In: *Phys. Lett. B* 206 (1988), pp. 169–176. DOI: 10.1016/0370-2693(88)91282-8.

[47] KTeV Collaboration. "Observation of direct CP violation in $K_{S,L} \to \pi\pi$ decays." In: *Phys. Rev. Lett.* 83 (1999), pp. 22–27. DOI: 10.1103/PhysRevLett.83.22. arXiv: hep-ex/9905060.

[48] NA48 Collaboration. "A new measurement of direct CP violation in two pion decays of the neutral kaon." In: *Phys. Lett. B* 465 (1999), pp. 335–348. DOI: 10.1016/S0370-2693(99)01030-8. arXiv: hep-ex/9909022.

[49] BaBar Collaboration. "Measurement of time-dependent CP asymmetry in $B^0 \to c\bar{c}K^{(*)0}$ decays." In: *Phys. Rev. D* 79 (2009), p. 072009. DOI: 10.1103/PhysRevD.79.072009. arXiv: 0902.1708 [hep-ex].

[50] X. Liu, W. Wang, and Y. Xie. "Penguin pollution in $B \to J/\psi V$ decays and impact on the extraction of the B_s^0–$\overline{B}_s^0$ mixing phase." In: *Phys. Rev. D* 89.9 (2014), p. 094010. DOI: 10.1103/PhysRevD.89.094010. arXiv: 1309.0313 [hep-ph].

[51] UTfit Collaboration. "The unitarity triangle fit in the Standard Model and hadronic parameters from lattice QCD: a reappraisal after the measurements of Δm_s and $BR(B \to \tau\nu_\tau)$." In: *JHEP* 10 (2006), p. 081. DOI: 10.1088/1126-6708/2006/10/081. arXiv: hep-ph/0606167.

[52] LHCb Collaboration. "Determination of the sign of the decay width difference in the B_s^0 system." In: *Phys. Rev. Lett.* 108 (2012), p. 241801. DOI: 10.1103/PhysRevLett.108.241801. arXiv: 1202.4717 [hep-ex].

[53] A. S. Dighe, I. Dunietz, and R. Fleischer. "Extracting CKM phases and B_s^0–$\overline{B}_s^0$ mixing parameters from angular distributions of nonleptonic B decays." In: *Eur. Phys. J. C* 6 (1999), pp. 647–662. DOI: 10.1007/s100520050372. arXiv: hep-ph/9804253.

[54] A. S. Dighe, I. Dunietz, H. J. Lipkin, and J. L. Rosner. "Angular distributions and lifetime differences in $B_s^0 \to J/\psi\phi$ decays." In: *Phys. Lett. B* 369 (1996), pp. 144–150. DOI: 10.1016/0370-2693(95)01523-X. arXiv: hep-ph/9511363.

[55] D0 Collaboration. "Measurement of B_s^0 mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$." In: *Phys. Rev. Lett.* 101 (2008), p. 241801. DOI: 10.1103/PhysRevLett.101.241801. arXiv: 0802.2255 [hep-ex].

[56] D0 Collaboration. "Measurement of the CP-violating phase $\phi_s^{J/\psi\phi}$ using the flavor-tagged decay $B_s^0 \to J/\psi\phi$ in 8fb^{-1} of $p\overline{p}$ collisions." In: *Phys. Rev. D* 85 (2012), p. 032006. DOI: 10.1103/PhysRevD.85.032006. arXiv: 1109.3166 [hep-ex].

[57] CDF Collaboration. "First flavor-tagged determination of bounds on mixing-induced CP violation in $B_s^0 \to J/\psi\phi$ decays." In: *Phys. Rev. Lett.* 100 (2008), p. 161802. DOI: 10.1103/PhysRevLett.100.161802. arXiv: 0712.2397 [hep-ex].

[58] CDF Collaboration. "Measurement of the CP-violating phase $\beta_s^{J/\psi\phi}$ in $B_s^0 \to J/\psi\phi$ decays with the CDF II detector." In: *Phys. Rev. D* 85 (2012), p. 072002. DOI: 10.1103/PhysRevD.85.072002. arXiv: 1112.1726 [hep-ex].

[59] CDF Collaboration. "Measurement of the bottom-strange meson mixing phase in the full CDF data set." In: *Phys. Rev. Lett.* 109 (2012), p. 171802. DOI: 10.1103/PhysRevLett.109.171802. arXiv: 1208.2967 [hep-ex].

[60] ATLAS Collaboration. "Time-dependent angular analysis of the decay $B_s^0 \to J/\psi\phi$ and extraction of $\Delta\Gamma_s$ and the CP-violating weak phase ϕ_s by ATLAS." In: *JHEP* 12 (2012), p. 072. DOI: 10.1007/JHEP12(2012)072. arXiv: 1208.0572 [hep-ex].

[61] ATLAS Collaboration. "Flavor tagged time-dependent angular analysis of the $B_s^0 \to J/\psi\phi$ decay and extraction of $\Delta\Gamma_s$ and the weak phase ϕ_s in ATLAS." In: *Phys. Rev. D* 90 (2014), p. 052007. DOI: 10.1103/PhysRevD.90.052007. arXiv: 1407.1796 [hep-ex].

[62] ATLAS Collaboration. "Measurement of the CP-violating phase ϕ_s and the B_s^0 meson decay width difference with $B_s^0 \to J/\psi\phi$ decays in ATLAS." In: *JHEP* 12 (2016), p. 072. DOI: 10.1007/JHEP08(2016)147. arXiv: 1601.03297 [hep-ex].

[63] ATLAS Collaboration. "Measurement of the CP-violating phase ϕ_s in $B_s^0 \to J/\psi\phi$ decays in ATLAS at 13 TeV." 2020. arXiv: 2001.07115 [hep-ex]. Submitted to EPJC.

[64] CMS Collaboration. "Measurement of the CP-violating weak phase ϕ_s and the decay width difference $\Delta\Gamma_s$ using the $B_s^0 \to J/\psi\phi$ decay channel in pp collisions at $\sqrt{s} = 8$ TeV." In: *Phys. Lett. B* 757 (2016), p. 97. DOI: 10.1016/j.physletb.2016.03.046. arXiv: 1507.07527 [hep-ex].

[65] LHCb Collaboration. "Measurement of the CP-violating phase ϕ_s in $\overline{B}_s^0 \to J/\psi\pi^+\pi^-$ decays." In: *Phys. Lett. B* 713 (2012), p. 378. DOI: 10.1016/j.physletb.2012.06.032. arXiv: 1204.5675 [hep-ex].

[66] LHCb Collaboration. "Measurement of CP violation and the B_s^0 meson decay width difference with $B_s^0 \to J/\psi K^+K^-$ and $B_s^0 \to J/\psi\pi^+\pi^-$ decays." In: *Phys. Rev. D* 87 (2013), p. 112010. DOI: 10.1103/PhysRevD.87.112010. arXiv: 1304.2600 [hep-ex].

[67] LHCb Collaboration. "Measurement of the CP-violating phase ϕ_s in $\overline{B}_s^0 \to J/\psi\pi^+\pi^-$ decays." In: *Phys. Lett. B* 736 (2014), p. 186. DOI: 10.1016/j.physletb.2014.06.079. arXiv: 1405.4140 [hep-ex].

[68] LHCb Collaboration. "Precision measurement of CP violation in $B_s^0 \to J/\psi K^+K^-$ decays." In: *Phys. Rev. Lett.* 114 (2015), p. 041801. DOI: 10.1103/PhysRevLett.114.041801. arXiv: 1411.3104 [hep-ex].

[69] LHCb Collaboration. "Updated measurement of time-dependent CP-violating observables in $B_s^0 \to J/\psi K^+K^-$ decays." In: *Eur. Phys. J. C* 79.8 (2019), p. 706. DOI: 10.1140/epjc/s10052-019-7159-8. arXiv: 1906.08356 [hep-ex]. [Erratum: Eur.Phys.J.C 80, 601 (2020)].

[70] LHCb Collaboration. "First study of the CP-violating phase and decay-width difference in $B_s^0 \to \psi(2S)\phi$." In: *Phys. Lett. B* 762 (2016), p. 253. DOI: 10.1016/j.physletb.2016.09.028. arXiv: 1608.04855 [hep-ex].

[71] LHCb Collaboration. "Measurement of the CP-violating phase ϕ_s in $\overline{B}_s^0 \to D_s^+D_s^-$ decays." In: *Phys. Rev. Lett.* 113 (2014), p. 211801. DOI: 10.1103/PhysRevLett.113.211801. arXiv: 1409.4619 [hep-ex].

[72] UTfit Collaboration. "The UTfit collaboration report on the status of the unitarity triangle beyond the standard model. I. Model-independent analysis and minimal flavor violation." In: *JHEP* 03 (2006), p. 080. DOI: 10.1088/1126-6708/2006/03/080. arXiv: hep-ph/0509219.

[73] M. Artuso, G. Borissov, and A. Lenz. "CP violation in the B_s^0 system." In: *Rev. Mod. Phys.* 88.4 (2016), p. 045002. DOI: 10.1103/RevModPhys.88.045002. arXiv: 1511.09466 [hep-ph]. [Addendum: Rev.Mod.Phys. 91, 049901 (2019)].

[74] J. Charles et al. "Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ New Physics." In: *Phys. Rev. D* 91.7 (2015), p. 073007. DOI: 10.1103/PhysRevD.91.073007. arXiv: 1501.05013 [hep-ph].

[75] A. Lenz and G. Tetlalmatzi-Xolocotzi. "Model-independent bounds on new physics effects in non-leptonic tree-level decays of B-mesons." In: *JHEP* 07 (2020), p. 177. DOI: 10.1007/JHEP07(2020)177. arXiv: 1912.07621 [hep-ph].

[76] Gargamelle Neutrino Collaboration. "Observation of neutrino like interactions without muon or electron in the Gargamelle Neutrino experiment." In: *Phys. Lett. B* 46 (1973), pp. 138–140. DOI: 10.1016/0370-2693(73)90499-1.

[77] UA1 Collaboration. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV." In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[78] UA2 Collaboration. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN anti-pp collider." In: *Phys. Lett. B* 122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2.

[79] OPAL Collaboration. "Photonic events with missing energy in e+ e- collisions at $\sqrt{s} = 189$ GeV." In: *Eur. Phys. J. C* 18 (2000), pp. 253–272. DOI: 10.1007/s100520000522. arXiv: hep-ex/0005002.

[80] ALEPH Collaboration. "Single photon and multiphoton production in e^+e^- collisions at $\sqrt{s}$ up to 209 GeV." In: *Eur. Phys. J. C* 28 (2003), pp. 1–13. DOI: 10.1140/epjc/s2002-01129-7.

[81] L3 Collaboration. "Single photon and multiphoton events with missing energy in e^+e^- collisions at LEP." In: *Phys. Lett. B* 587 (2004), pp. 16–32. DOI: 10.1016/j.physletb.2004.01.010. arXiv: hep-ex/0402002.

[82] DELPHI Collaboration. "Photon events with missing energy in e+ e- collisions at $\sqrt{s} = 130$ GeV to 209 GeV." In: *Eur. Phys. J. C* 38 (2005), pp. 395–411. DOI: 10.1140/epjc/s2004-02051-8. arXiv: hep-ex/0406019.

[83] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[84] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[85] R Scrivens, M Kronberger, D Küchler, J Lettry, C Mastrostefano, O Midttun, M O'Neil, H Pereira, and C Schmitzer. "Overview of the status and developments on primary ion sources at CERN." In: CERN-ATS-2011-172 (Sept. 2011), 4 p.

[86] E. Mobs. *The CERN accelerator complex - 2019*. General Photo. July 2019. URL: https://cds.cern.ch/record/2684277.

[87] A. Team. *Diagram of an LHC dipole magnet*. June 1999. URL: https://cds.cern.ch/record/40524.

[88] ALICE Collaboration. "The ALICE experiment at the CERN LHC." In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[89] ATLAS Collaboration. "The ATLAS experiment at the CERN Large Hadron Collider." In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[90] LHCb Collaboration. "The LHCb detector at the LHC." In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[91] L. R. Evans and P. Bryant. "LHC Machine." In: *JINST* 3 (2008), S08001. 164 p. DOI: 10.1088/1748-0221/3/08/S08001. This report is an abridged version of the LHC Design Report (CERN-2004-003).

[92] C. Pralavorio. *Record luminosity: well done LHC*. Nov. 2017. URL: https://cds.cern.ch/record/2295027.

[93] CMS Collaboration. *CMS luminosity public results*. 2019. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Online; accessed November 2020.

[94] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, and J. Mazzitelli. "Bottom-quark production at hadron colliders: fully differential predictions in NNLO QCD." Oct. 2020. arXiv: 2010.11906 [hep-ph].

[95] W. Sirling. *Private communication*. URL: http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html. Online; accessed November 2020.

[96] CMS Collaboration. "The CMS experiment at the CERN LHC." In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[97] I. Neutelings. *How to draw diagrams in LaTeX with TikZ*. Sept. 2019. URL: https://wiki.physik.uzh.ch/cms/latex:tikz. Online; accessed November 2020.

[98] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Tech. rep. CERN-LHCC-97-010. 1997. URL: https://cds.cern.ch/record/331056.

[99] CMS Collaboration. "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays." In: *JINST* 5 (2010), T03021. DOI: 10.1088/1748-0221/5/03/T03021. arXiv: 0910.5530 [physics.ins-det].

[100] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Tech. rep. CERN-LHCC-98-006. 1997. URL: http://cds.cern.ch/record/368412.

[101] CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report*. Tech. rep. CERN-LHCC-2000-016. 2000. URL: http://cds.cern.ch/record/490194.

[102] A Dominguez et al. *CMS Technical Design Report for the pixel detector upgrade*. Tech. rep. CERN-LHCC-2012-016. Sept. 2012. URL: https://cds.cern.ch/record/1481838.

[103] The Tracker Group of the CMS Collaboration. *The CMS Phase-1 pixel detector upgrade*. Tech. rep. CMS-NOTE-2020-005. Mar. 2020. URL: https://cds.cern.ch/record/2745805.

[104] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-033. 1997. URL: http://cds.cern.ch/record/349375; *changes to CMS ECAL electronics: addendum to the Technical Design Report*, CERN-LHCC-2002-027, URL: http://cdsweb.cern.ch/record/581342.

[105] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-031. 1997. URL: http://cds.cern.ch/record/357153.

[106] CMS Collaboration. *The CMS muon project: Technical Design Report*. Tech. rep. CERN-LHCC-97-032. 1997. URL: http://cds.cern.ch/record/343814.

[107] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector." In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[108] CMS Collaboration. *The Phase-2 upgrade of the CMS tracker*. Tech. rep. CMS-TDR-014. June 2017. URL: https://cds.cern.ch/record/2272264.

[109] CMS Collaboration. *CMS tracker detector performance results*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK. Online; accessed November 2020.

[110] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: detector performance and software*. Tech. rep. CERN-LHCC-2006-001. 2006. URL: https://cds.cern.ch/record/922757.

[111] CMS HCAL/ECAL Collaborations. "The CMS barrel calorimeter response to particle beams from 2 to 350-GeV/c." In: *Eur. Phys. J. C* 60 (2009), pp. 359–373. DOI: 10.1140/epjc/s10052-009-0959-5. [Erratum: Eur.Phys.J.C 61, 353–356 (2009)].

[112] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[113] CMS Collaboration. *CMS computing: Technical Design Report*. Tech. rep. CERN-LHCC-2005-023. 2005. URL: https://cds.cern.ch/record/838359.

[114] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: the trigger systems*. Tech. rep. CERN-LHCC-2000-038. 2000. URL: http://cds.cern.ch/record/706847.

[115] CMS Collaboration. *CMS The TriDAS Project: Technical Design Report, Volume 2: data acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Tech. rep. CERN-LHCC-2002-026. 2002. URL: http://cds.cern.ch/record/578006.

[116] CMS Collaboration, Data Acquisition Group. "The CMS high level trigger." In: *Eur. Phys. J. C* 46 (2006), pp. 605–667. DOI: 10.1140/epjc/s2006-02495-8. arXiv: hep-ex/0512077.

[117] CMS Collaboration. *CMS the TriDAS project: Technical Design Report, Volume 2: data acquisition and high-level trigger.* Tech. rep. CERN-LHCC-2002-026. 2002. URL: https://cds.cern.ch/record/578006.

[118] T. Bawej et al. "The new CMS DAQ system for Run-2 of the LHC." In: *IEEE Trans. Nucl. Sci.* 62.3 (2015), pp. 1099–1103. DOI: 10.1109/RTC.2014.7097437.

[119] CMS Collaboration. "The CMS trigger system." In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[120] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker." In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[121] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting." In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[122] P. Billoir. "Progressive track recognition with a Kalman like fitting procedure." In: *Comput. Phys. Commun.* 57 (1989), pp. 390–394. DOI: 10.1016/0010-4655(89)90249-X.

[123] P. Billoir and S. Qian. "Simultaneous pattern recognition and track fitting by the Kalman filtering method." In: *Nucl. Instrum. Meth. A* 294 (1990), pp. 219–228. DOI: 10.1016/0168-9002(90)91835-Y.

[124] CMS Collaboration. "CMS tracking performance results from early LHC operation." In: *Eur. Phys. J. C* 70 (2010), pp. 1165–1192. DOI: 10.1140/epjc/s10052-010-1491-3. arXiv: 1007.1988 [physics.ins-det].

[125] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems." In: *IEEE Proc.* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788.

[126] R. Frühwirth, W. Waltenberger, and P. Vanlaer. "Adaptive vertex fitting." In: *J. Phys. G* 34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[127] CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV." In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

[128] J. C. Collins, D. E. Soper, and G. F. Sterman. "Factorization of hard processes in QCD." In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[129] D. Klingebiel. "Measurement of the t-channel single top quark production cross section and the CKM matrix element V_{tb} with the CMS experiment." PhD thesis. RWTH Aachen University, 2014. URL: https://publications.rwth-aachen.de/record/444972.

[130] M. Dobbs et al. "Les Houches guidebook to Monte Carlo generators for hadron collider physics." In: *3rd Les Houches Workshop on Physics at TeV Colliders*. Mar. 2004, pp. 411–459. arXiv: hep-ph/0403045.

[131] A. Schaelicke, T. Gleisberg, S. Hoeche, S. Schumann, J. Winter, F. Krauss, and G. Soff. "Event generator for particle production in high-energy collisions." In: *Prog. Part. Nucl. Phys.* 53 (2004). Ed. by A. Faessler, pp. 329–338. DOI: 10.1016/j.ppnp.2004.02.031. arXiv: hep-ph/0311270.

[132] GEANT4 Collaboration. "Geant4 — A simulation toolkit." In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8.

[133] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. "An introduction to PYTHIA 8.2." In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[134] CMS Collaboration. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements." In: *Eur. Phys. J. C* 80 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[135] NNPDF Collaboration. "Parton distributions from high-precision collider data." In: *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[136] E. Barberio, B. van Eijk, and Z. Wąs. "PHOTOS — a universal Monte Carlo for QED radiative corrections in decays." In: *Comput. Phys. Commun.* 66 (1991), p. 115. ISSN: 0010-4655. DOI: 10.1016/0010-4655(91)90012-A.

[137] E. Barberio and Z. Wąs. "PHOTOS — a universal Monte Carlo for QED radiative corrections: version 2.0." In: *Comput. Phys. Commun.* 79 (1994), p. 291. ISSN: 0010-4655. DOI: 10.1016/0010-4655(94)90074-4.

[138] D. J. Lange. "The EvtGen particle decay simulation package." In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. DOI: 10.1016/S0168-9002(01)00089-4.

[139] G. Punzi. "Comments on likelihood fits with variable resolution." In: *eConf* C030908 (2003). Ed. by L. Lyons, R. Mount, and R. Reitmeyer, WELT002. arXiv: physics/0401045.

[140] CMS Collaboration. *CMS Offline Software*. URL: http://cms-sw.github.io/. Online; accessed November 2020.

[141] I. Antcheva et al. "ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization." In: *Comput. Phys. Commun.* 180 (2009), p. 2499. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2009.08.005. eprint: 1508.07749.

[142] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt. "TMVA, the Toolkit for Multivariate Data Analysis with ROOT." In: *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*. 2007, p. 40. DOI: 10.22323/1.050.0040. arXiv: physics/0703039. [PoS(AT)040]CA. Up-to-date website and Users Guide: http://tmva.sourceforge.net/.

[143] W. Verkerke and D. Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an].

[144] F. Faccio, S. Michelis, and G. Ripamonti. *Failure of FEAST2.1-based modules in the CMS pixel detector system (physics run 2017)*. Tech. rep. CERN, 2018. URL: https://espace.cern.ch/project-DCDC-new/_layouts/15/start.aspx#/Reports.

[145] CMS Collaboration. *Crack in the mirror*. 2020. URL: https://cms.cern/news/crack-mirror. Online; accessed November 2020.

[146] M. Pivk and F. R. Le Diberder. "SPlot: a statistical tool to unfold data distributions." In: *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083.

[147] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung. "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit." In: *Nature* 405.6789 (June 2000), pp. 947–951. ISSN: 1476–4687. DOI: 10.1038/35016072.

[148] X. Glorot, A. Bordes, and Y. Bengio. "Deep sparse rectifier neural networks." In: ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: JMLR Workshop and Conference Proceedings, Mar. 2011, pp. 315–323. URL: http://proceedings.mlr.press/v15/glorot11a.html.

[149] A. I. Georgevici and M. Terblanche. "Neural networks and deep learning: a brief introduction." In: *Intensive Care Medicine* 45.5 (May 2019), pp. 712–714. ISSN: 1432–1238. DOI: 10.1007/s00134-019-05537-w.

[150] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. "Improving neural networks by preventing co-adaptation of feature detectors." 2012. arXiv: 1207.0580 [cs.NE].

[151] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[152] J. S. Bridle. "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition." In: *Neurocomputing*. Ed. by F. F. Soulié and J. Hérault. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 227–236. ISBN: 978-3-642-76153-9.

[153] R. D. Luce. *Individual choice behavior: a theoretical analysis*. New York, NY, USA: Wiley, 1959.

[154] A. Niculescu-Mizil and R. Caruana. "Predicting good probabilities with supervised learning." In: Jan. 2005, pp. 625–632. DOI: 10.1145/1102351.1102430.

[155] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. "On calibration of modern neural networks." 2017. arXiv: 1706.04599 [cs.LG].

[156] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors." In: *Nature* 323.6088 (Oct. 1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0.

[157] D. P. Kingma and J. Ba. "Adam: a method for stochastic optimization." 2017. arXiv: 1412.6980 [cs.LG].

[158] J. Wishahi. *Flavour tagging plots for conference*. 2016. URL: https://twiki.cern.ch/twiki/bin/view/LHCb/FlavourTaggingConferencePlots. Online; accessed November 2020.

[159] CDF Collaboration. "Observation of B_s^0–$\overline{B}_s^0$ oscillations." In: *Phys. Rev. Lett.* 97 (2006), p. 242003. DOI: 10.1103/PhysRevLett.97.242003. arXiv: hep-ex/0609040.

[160] D. Fazzini. "Development of "same side" flavour tagging algorithms for measurements of flavour oscillations and CP violation in the B^0 mesons system." MA thesis. Milan Bicocca University, Mar. 2015.

[161] LHCb Collaboration. "Measurement of CP asymmetries in two-body $B^0_{(s)}$-meson decays to charged pions and kaons." In: *Phys. Rev. D* 98.3 (2018), p. 032004. DOI: 10.1103/PhysRevD.98.032004. arXiv: 1805.06759 [hep-ex].

[162] J. Pazzini. "B flavour tagging with leptons and measurement of the CP violation phase ϕ_s in the $B_s^0 \to J/\psi\phi$ decay at the CMS experiment." PhD thesis. Padova University, 2015.

[163] F. Chollet et al. *Keras*. 2015. URL: https://keras.io. Online; accessed November 2020.

[164] A. Boletti. “Classification of muons produced in semileptonic B hadron decays in $pp \to t\bar{t}$ events at LHC.” MA thesis. Padova University, Sept. 2014.

[165] R. A. Fisher. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd, 1925.

[166] F. James and M. Roos. “Minuit — a system for function minimization and analysis of the parameter errors and correlations.” In: *Comput. Phys. Commun.* 10 (1975), p. 343. DOI: 10.1016/0010-4655(75)90039-9.

[167] R. J. Barlow. “Extended maximum likelihood.” In: *Nucl. Instrum. Meth. A* 297 (1990), pp. 496–506. DOI: 10.1016/0168-9002(90)91334-8.

[168] F James and M Goossens. *MINUIT: function minimization and error analysis reference manual*. 1998. CERN Program Library Long Writeups. URL: http://cds.cern.ch/record/2296388.

[169] H. G. Moser and A Roussarie. “Mathematical methods for B^0–$\overline{B}^0$ oscillation analyses.” In: *Nucl. Instrum. Methods Phys. Res., A* 384.CERN-OPEN-99-030. CERN-ALEPH-PUB-96-005. 2-3 (May 1996), 491–505. 27 p. DOI: 10.1016/S0168-9002(96)00887-X.

[170] CMS Collaboration. “Measurement of b hadron lifetimes in pp collisions at $\sqrt{s} = 8$ TeV.” In: *Eur. Phys. J. C* 78.6 (2018), p. 457. DOI: 10.1140/epjc/s10052-018-5929-3. arXiv: 1710.08949 [hep-ex]. [Erratum: Eur.Phys.J.C 78, 561 (2018)].

[171] N. L. Johnson. “Systems of frequency curves generated by methods of translation.” In: *Biometrika* 36 (1949), p. 149. DOI: 10.1093/biomet/36.1-2.149.

[172] M. Rosenblatt. “Remarks on some nonparametric estimates of a density function.” In: *Ann. Math. Statist.* 27 (1956), p. 832. DOI: 10.1214/aoms/1177728190.

[173] E. Parzen. “On estimation of a probability density function and mode.” In: *Ann. Math. Statist.* 33 (1962), p. 1065. DOI: 10.1214/aoms/1177704472.

[174] L. Lyons, D. Gibaut, and P. Clifford. “How to combine correlated estimates of a single physical quantity.” In: *Nucl. Instrum. Meth. A* 270 (1988), p. 110. DOI: 10.1016/0168-9002(88)90018-6.

[175] A. Valassi. “Combining correlated measurements of several different physical quantities.” In: *Nucl. Instrum. Meth. A* 500 (2003), pp. 391–405. DOI: 10.1016/S0168-9002(03)00329-2.

[176] R. Nisius. “On the combination of correlated estimates of a physics observable.” In: *Eur. Phys. J. C* 74.8 (2014), p. 3004. DOI: 10.1140/epjc/s10052-014-3004-2. arXiv: 1402.4016 [physics.data-an].

[177] R. Nisius. "BLUE: combining correlated estimates of physics observables within ROOT using the Best Linear Unbiased Estimate method." In: *SoftwareX* 11 (Jan. 2020), p. 100468. ISSN: 2352-7110. DOI: 10.1016/j.softx.2020.100468. arXiv: 2001.10310 [physics.data-an].

[178] R. L. Plackett. "A historical note on the method of least squares." In: *Biometrika* 36.3/4 (1949), pp. 458–460. ISSN: 00063444. DOI: 10.2307/2332682.

[179] CMS Collaboration. *CP-violation studies at the HL-LHC with CMS using* B_s^0 *decays to* $J/\psi\phi(1020)$. Tech. rep. CMS-PAS-FTR-18-041. 2018. URL: http://cds.cern.ch/record/2650772.

[180] A. A. Cerri, V. V. Gligorov, S. Malvezzi, S. Martin Camalich, J. Zupan, et al. "Report from Working Group 4: Opportunities in flavour physics at the HL-LHC and HE-LHC." In: *Report on the physics at the HL-LHC, and perspectives for the HE-LHC* 7 (Dec. 2019). Ed. by A. Dainese, M. Mangano, A. B. Meyer, A. Nisati, G. Salam, and M. A. Vesterinen, pp. 867–1158. DOI: 10.23731/CYRM-2019-007.867. arXiv: 1812.07638 [hep-ph].