Università
degli Studi
di Padova

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"

SCUOLA DI DOTTORATO DI RICERCA IN FISICA
CICLO XXXV

FLAVOR TAGGING ALGORITHMS FOR CP VIOLATION MEASUREMENTS IN CMS

ENRICO LUSIANI

SUPERVISORE:
PROF. PAOLO RONCHESE

CO-SUPERVISORI:
PROF. FRANCO SIMONETTO
PROF. MARTINO MARGONI

ANNO ACCADEMICO 2022–2023

Enrico Lusiani: *Algoritmi di flavor tagging per misure di violazione CP in CMS* ©
March 2023

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. Paolo Ronchese
CO-SUPERVISORI:
Prof. Franco Simonetto Prof. Martino Margoni

ABSTRACT

This thesis presents the development of four flavor tagging algorithms to be used in CP violation measurements in CMS. Flavor tagging is an essential tool for CP violation measurement, which often rely on the knowledge of both the initial and final flavor of a meson. Three of the algorithm presented use the production of b quarks as $b\overline{b}$ pairs and analyze the b-hadron unrelated to the signal, while the last one examines the hadronization of the signal meson to infer its flavor. The last algorithms is specifically novel in its kind, as it does not make use of particle identification for the inference or the selection of the input features. All taggers examined are based on Deep Neural Networks and are developed and calibrated using reconstructed $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi\,K^+$ decays.

The performance of the taggers is evaluated in a dataset corresponding to $96.48\,\mathrm{fb}^{-1}$ collected in proton-proton collisions at $\sqrt{s} = 13$ TeV during the years 2017 and 2018 by the CMS experiment. After combination of the various algorithms, the total tagging power P_{tag} is estimated to be $P_{\mathrm{tag}} = 5.59 \pm 0.02\%$.

With this combination, a $\sim 30\%$ subset of the data sample is used in a time-dependent angular analysis for a measurement of the CP violating phase ϕ_s, obtaining a result of $\phi_s = -21 \pm 42(\text{stat.})$ mrad. A suite of tagging related systematic uncertainties is evaluated on the measurements, showing them to be negligible compared to the statistical uncertainty and to the intrinsic model bias of the fitting procedure.

SINTESI

Questa tesi presenta lo sviluppo di quattro algoritmi di tag da usare in misure di violazione CP in CMS. Il flavor tagging è uno strumento essenziale nelle misure di violazione di CP, che spesso sfruttano la conoscenza del flavor sia iniziale che finale. Tre degli algoritmi presentati fanno uso della produzione di quark b come coppie $b\bar{b}$ e analizzano il l'adrone B scorrelato con il segnale, mentre il quarto esamina l'adronizzazione del mesone di segnale per dedurre il suo flavor. Quest'ultimo algoritmo in particolare è nuovo nel suo genere, in quanto non fa uso di informazioni sull'identità delle particelle per l'inferenza o per la selezione delle feature di input. Tutti i tagger esaminati sono basati su Deep Neural Networks e sono sviluppati e calibrati usando decadimenti ricostruiti di $B_s^0 \to J/\psi\,\phi(1020)$ e $B^+ \to J/\psi\,K^+$.

La performance dei tagger è valutata in un dataset corrispondente ad una luminosità integrata di $96.48\,\mathrm{fb}^{-1}$, raccolto in collisioni protone-protone a $\sqrt{s} = 13$ TeV negli anni 2017 e 2018 dall'esperimento CMS. Dopo la combinazione dei vari algoritmi, il tagging power totale è stimato essere $P_{\mathrm{tag}} = 5.59 \pm 0.02\%$.

Usando la combinazione dei tagger, $\sim 30\%$ del sample di dati è usato in una analisi angolare dipendente dal tempo per una misura della fase di violazione CP ϕ_s, ottenendo un risultato di $\phi_s = -21 \pm 42(\mathrm{stat.})$ mrad. Una serie di incertezze sistematiche legate al tagging sono valutate sulla misura, mostrando che sono trascurabili rispetto all'incertezza statistica e al bias intrinseco della procedura di fit.

CONTENTS

I Theoretical background

II Experimental setup

III Flavor tagging

MOST USED ACRONYMS

BSM	Beyond the Standar Model
CDF	Cumulative Density Function
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
LHC	Large Hadron Collider
MC	Monte Carlo
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
PDL	Proper Decay Length
PF	Particle Flow
PID	Particle ID
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

The development of Quantum Mechanics and Quantum Field Theory in the past century culminated in what is called the Standard Model of particle physics, an extensive theory incorporating all known fundamental interactions save for gravity. Since its inception, the Standard Model has been extensively tested and repeatedly confirmed by experimental observations, the latest great success being the discovery of the Higgs boson in 2012, a particle predicted over fifty years before.

However, despite the great success, the SM leaves a lot unexplained: gravity, as mentioned before, is not included in the model, despite being the interaction most of us are more familiar with. For many other observations the SM provides partial or incorrect answers. Neutrinos are predicted in the model, but only as particles with no mass, while the observation of neutrino oscillation proves that they have one, albeit small. Cosmological and astronomical observations suggest the existence of Dark Matter and Dark Energy but all the candidate explanations lead to predictions wrong sometimes by several orders of magnitude.

The origin of the asymmetry between matter and antimatter is another unsolved problem in the theory: our present day universe consists predominantly of matter with almost no antimatter particles present, but cosmological models predict that an equal amount of matter and antimatter was created in the early phases of the universe. As identified by Sakharov in 1967 [1], there are three essential conditions to explain the different abundance of matter and antimatter: a process violating the conservation of baryon number, an Universe out of thermal equilibrium, and difference in behaviour between matter and antimatter. More precisely, the last condition requires the simultaneous violation of the charge (C) and the parity (P) symmetries, combined in what is called violation of CP symmetry.

CP violation was first observed directly in 1964 in the decay of neutral kaons [2]. A mechanism for describing the violation was introduced in the SM in 1973, in the form of the Cabibbo–Kobayashi–Maskawa (CKM) quark mixing matrix [3] The researchers generalized the existing two dimensional matrix to include a third generation of matter, which at the time had not yet been observed. The third dimension allowed the inclusion of a non-trivial irreducible complex phase, which can introduce differences in the weak transitions of particles and antiparticles. However, the estimated violation of CP in the SM framework is still not sufficient to explain the observed matter-antimatter asymmetry.

All of these observations point to the existence of New Physics beyond the Standard Model. For this reason, the focus of many analyses in these days is the precise probing of the model, to find inconsistencies and processes where theory breaks down. The Large Hadron Collider (LHC) at CERN is currently the most powerful particle accelerator and thus a prime candidate for all of these searches.

Many New Physics models predict the existence of new, heavy particles and two main methods can be used to detect them: the first is direct observation, which was used in 2012 for the discovery of the Higgs boson, the second is indirect detection based on the effect they have when appearing as virtual particles in quantum loops. The particles involved in quantum loops are not limited by the available energy in the interaction, which means that the indirect detection can probe energy scales unreachable in direct searches. The indirect strategy relies both on a precise measurements of such SM processes and on the availability of predictions or constraints within the SM at a similar or better level.

The decay of the B_s^0 meson into $J/\psi\,\phi(1020)$ is a particular interesting probe in this regard. The final state is accessible to both B_s^0 and $\overline{B}_s^0$, meaning that a CP violation measurement in this system would be sensitive to the B_s^0-$\overline{B}_s^0$ mixing, where non-SM particles can contribute. The weak phase ϕ_s, specifically, parametrizes the CP violation in the interference between direct B_s^0 meson decays to a CP eigenstate via $b \rightarrow c\bar{c}s$ transitions and decays through mixing to the same final state. In the SM this phase is related to the elements of the CKM matrix, which are constrained precisely by global fits to experimental data.

As B_s^0 mesons oscillate, the analysis relies on the knowledge of the production flavor of the mesons. This information is extracted from the data by means of flavor tagging algorithms. This dissertation presents the development of four flavor tagging algorithms, two based on semileptonic decays, one based on charged particles around a b-tagged jet and one based on the inclusive analysis of the products of the B_s^0 meson fragmentation.

The thesis also contains a measurement of the ϕ_s phase performed on a limited sub-sample of CMS data. The analysis is used as a benchmark for the performance of the algorithms, as well as an estimation of the impact of their introduction to the systematic uncertainties.

THESIS OUTLINE

The thesis is structured as follows. Part I gives a brief overview of the theoretical background and on the importance of CP violation analyses, where flavor tagging is an essential ingredient. Specifically, Chapter 1 is dedicated to a general introduction to the Standard Model, while Chapter 2 gives a review of the physics behind the CP violation in the SM.

Part II shows the experimental setup in which the flavor tagging algorithms were developed. Chapter 3 describes the Large Hadron Collider and the CMS experiment. Chapter 4 describes the algorithm used in CMS for the reconstruction of high level observables from the raw detector readout, the concepts behind the online selection performed by the CMS trigger system, and the basics of Monte Carlo simulations of a high energy physics experiment Finally, Chapter 5 describes the procedures and selections used to prepare the data for use in the tagger and the CP violation analysis.

Part III, the focus of this thesis, describes in detail the flavor tagging process. Chapter 6 explains the basics behind flavor tagging, the different categories of algorithms, and the role of neural networks for tagging probability estimation. Chapter 7 to 10 explain the different tagging algorithm available in CMS, with a special focus to the Opposite Side Jet and Same Side taggers. Chapter 11 summarizes the performances of the different taggers and describes a method for the combination of tagging from different independent sources.

Finally, Part IV provides a benchmark for the performance of the taggers in the form of a CP violation measurement in the $B_s^0 \to J/\psi\,\phi(1020)$ decay, one of the most important channels for CP violation in CMS. Chapter 12 is an overview of the analysis, specifically of the theoretical motivations. In Chapter 13, the fitting procedure used to extract the values of the parameters and their uncertainties is explained. The thesis concludes with an evaluation, in Chapter 14, of the possible tagging related sources of systematic uncertainty in the measurement, and compares them with the statistical uncertainty and the intrinsic bias of the fitting procedure.

PUBLICATIONS AND BIBLIOGRAPHIC REFERENCES

The analysis presented in this dissertation is part of the official CMS measurement of the CP-violating phase ϕ_s performed at a center-of-mass energy of 13 TeV. It expands on the results from Ref. [4] with the inclusion of a new trigger strategy and three new flavor taggers. This work is also documented in the internal analysis notes AN-22-084 and AN-22-078 [I].

OWN CONTRIBUTIONS

The taggers and the measurement presented in this thesis were developed with contributions from a team of several people, each contributing to various aspects of the analysis. In the context of the tagging algorithms, my focus within the group was the development of the Opposite Side Jet and Same Side taggers, of the calibration procedure for the Opposite Side taggers, and all the various studies for the complex Same Side tagger calibration.

In the CPV analysis, I developed and optimized the whole fitting procedure, as well as the studies for the evaluation of the tagging-based systematic uncertainties.

UNIT CONVENTION

Unless stated otherwise, through this dissertation, the "natural units" convention is implied 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.

[I] 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 and give raise to both matter and the fundamental forces. The Standard Model incorporates all known fundamental interactions save for gravity into a precise, consistent and very successful theory. However, several phenomena are left unexplained in the SM, a prominent example being the matter-antimatter asymmetry in the Universe. 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. CP violation was observed in the second half of the past century, and incorporated into the SM, but not in an amount sufficient to explain the observed asymmetry. For this reasons, measures targeting CP-violating phenomena are of high interest in modeling high energy physics: any variation from the predicted value would help identifying possible extensions on the Standard Model, and help shedding light on this currently unsolved problem. 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.

1

THE STANDARD MODEL OF PARTICLE PHYSICS

The Standard Model of Particle Physics (SM) [5–11] describes our knowledge of the fundamental particles and interactions up to the present day. The SM is a Quantum Field Theory which unifies all previous attempts at unifying Quantum Mechanics and Special Relativity. 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, the world is described through the interaction of several particles, 12 elementary fermions with spin 1/2, which describe visible matter, and 13 bosons, which describe fundamental interactions. The particles can be further classified on the basis of how they interact with each other.

The quarks are a group of 6 fermions that are affected by all fundamental interactions. In particular, they can interact through the strong interaction, since they are the only particles which carry color charge. Quarks can only be found in color-neutral bound states due to the phenomenon called *color confinement*. The most common states are groups of three quarks (baryons) or a quark and an anti-quark (mesons), but more recently there have also been various experimental evidences for bounded states of four and five quarks [12–14]. There are two types of quarks: the up-type (u, c, t), with electric charge $+\frac{2}{3}e$ and the down-type (d, s, b), with charge $-\frac{1}{3}e$.

Leptons are particles which 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 electrically neutral and can then interact only through weak interaction.

Both groups of fermions can be divided into three generations each containing an up-and down-type quark, a charged lepton and a neutrino, with increasing mass[I].

Gauge bosons in the SM are the particles that mediate the fundamental interactions. All gauge bosons have spin 1 and they are classified according to the force they mediate.

- The *strong* interaction affects color charged particles (i. e., quarks) and is mediated by eight massless gluons (g). Gluons also interact among themselves as they have an effective color charge.

[I] Neutrinos, however, are massless in the SM. Experimental results have shown that they have mass, but it has not yet been determined.

- The *electromagnetic* force affects electrically charged particles and is mediated by the massless neutral photon γ.
- The *weak* interaction affects all elementary fermions (i. e., quarks and leptons) and is mediated by the W^+, W^- and Z massive bosons. The Z is electrically neutral, while the W bosons have an electric charge of $+1e$ and $-1e$, respectively.

1.2 FIELD CONTENT AND THE STANDARD MODEL LAGRANGIAN

1.2.1 *The Standard Model field content*

The Standard Model is a quantum field theory that exploits the concept of local gauge symmetries. This means that the theory implements a set of gauge transformations under which the actual physics processes are invariant although the mathematical formulation of the theory changes. Those gauge transformations are local, i.e. they do depend on a specific point $\mathrm{X} = (ct, \vec{x})$ in space time in contrast to global gauge transformations which do not depend on X. The invariance group of the SM Lagrangian $\mathcal{L}_{\mathrm{SM}}$ is

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

The mathematical formulation of the gauge transformations dictates the bosonic field content of the Standard Model and the interactions between the particles, which are mediated by the bosonic force carriers. It is however important to understand that the choice of the gauge transformations in the Standard Model is not arbitrary but justified by symmetries observed in nature and the conservation of quantum numbers connected to these symmetries. The octet of gluons G^a_μ for the $\mathrm{SU}(3)_C$ group established the interactions of quantum chromodynamics (QCD) by acting on $\mathrm{SU}(3)_C$ triplets of quark. Electroweak gauge bosons (EW) are represented by the $\mathrm{SU}(2)_L$ triplet W^i_μ and the singlet B_μ of $\mathrm{U}(1)_Y$.

Elementary fermions are described by right- and left-handed spinors. Left-handed fermions form $\mathrm{SU}(2)_L$ doublets of up-/down-quarks or a charged lepton and a neutrino in the same generation. Right-handed fermions form $\mathrm{SU}(2)_L$ singlets, except for right-handed neutrinos, which in the SM are considered sterile (non-interacting). All elementary fermions are charged under $\mathrm{U}(1)_Y$.

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

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

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

1.2.2 *The quantum chromodynamics sector*

The Quantum ChromoDynamics (QCD) Lagrangian can be described as

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

where

- ψ is a generic *massless* quark spinor field;
- γ^μ are the Dirac matrices [15];
- 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 [16] 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

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

the f_{abc} are the structure constants of $\mathrm{SU}(3)_C$; the last term of the field tensor is responsible for the self-interaction of the gluons.

The color confinement of color charged particles is believed to be caused by the energy dependance of the QCD coupling constant, which decreases with increasing energy, leading to an interaction that becomes stronger as the quarks move apart. At the same time, this implies that a perturbative description of QCD is only possible at high energy.

1.2.3 *The electroweak sector*

The electromagnetic and weak interactions are described as effective low energy views of a more general electroweak interaction. The Lagrangian of the electroweak sector of the SM includes the $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ simmetries, and is written as

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

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

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

where

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

  $$F^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk}W^j_\mu W^k_\nu\,, \tag{1.8}$$

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

  where W^i_μ is the $\mathrm{SU}(2)_L$-gauge field triplet, B_μ is the $\mathrm{U}(1)_Y$-gauge field, and ϵ^{ijk} is the three-dimensional Levi-Civita symbol;

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

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

  where g and g' are the coupling constants of the $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ interaction, respectively, and T^i and Y are 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]

The particles involved in Eq. 1.5 are massless. The mass could be added with terms in the form of $m_f\overline{\psi}\psi$, $m_B^2 B_\mu B^\mu$ or $m_W^2 W_\mu W^\mu$, but such terms would break

---

[II] In the SM, there are no right-handed neutrinos, nor left-handed antineutrinos.

the gauge simmetry. In the SM the masses of charged elementary fermions and gauge bosons are instead generated through electroweak spontaneous symmetry breaking (SSB), in the so-called "Higgs mechanism".

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 simmetry breaking* (SSB) refer to a situation in which the symmetry is not broken explicitly by the interaction, but the asymmetry is found instead in the state of lowest energy, referred to as the vacuum state. According to the Goldstone theorem, spontaneous symmetry breaking (SSB) creates massless bosons, the so-called "Goldstone bosons" [17], but in case of a broken gauge symmetry, 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 [18, 19], Robert Brout and François Englert [20], and by Gerald Guralnik, Carl Hagen and Tom Kibble [21][III].

In its simplest implementation, the Higgs mechanism requires the introduction of a weakly coupled spin-0 particle with a potential that is minimized at a non-zero field value. 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

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

describing the interaction with the gauge bosons, and $\mathcal{L}_{\text{Yukawa}}$ is the Yukawa Lagrangian, shown in Section 1.3.2, which describes the generation of charged elementary fermion masses. SSB requires a non zero vacuum expectation value. The only renormalizable model where a complex scalar field acquires a nonzero value is the *Mexican-hat* model, with potential:

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

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

Due to unitarity the constants λ and μ^2 are required to be real, while for a stable minimum to exist, λ is required to be positive. The parameter μ^2 is chosen to be negative, in order to realize the spontaneous symmetry breaking

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

where $\mathrm{U}(1)_{\mathrm{QED}}$ 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 $v \equiv \sqrt{-\frac{\mu^2}{\lambda}}$. Since the vacuum must be electrically neutral, 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 represents *radial* excitations of the Higgs VEV.

1.3.1 *Boson masses*

The bosonic mass Lagrangian $\mathcal{L}_m$ can be obtained by rewriting Eq. 1.11 using the unitarity gauge and ignoring for now the Yukawa terms:

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

With a rotation of the bosonic fields, we can obtain the physical states

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

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

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

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^d_{Ri}, and q^u_{Ri} with $i = 1, 2, 3$ are the right-handed lepton, down-type quark, and up-type quark singlets, respectively;
- $L_{Li} = (\nu_{Li}, \ell_{Li})^\top$ and $Q_{Li} = (q^u_{Li}, q^d_{Li})^\top$ 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[IV];
- 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 Rewritten using the unitarity gauge, the lepton part of Eq. 1.28
In the unitary gauge[V] the lepton sector of the Yukawa Lagrangian is given by:[VI]

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

or in matrix form

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

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

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

This behaves as a mass term, but 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.

In order to construct fields with a definite mass, both the fields and the Yukawa matrix Y^ℓ needs to be rotated through a biunitary transformation obtaining a diagonal matrix

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

$$\ell'_L = V^{\ell\dagger}_L\ell_L \equiv \begin{pmatrix} e'_L \\ \mu'_L \\ \tau'_L \end{pmatrix}, \quad \ell'_R = V^{\ell\dagger}_R\ell_R \equiv \begin{pmatrix} e'_R \\ \mu'_R \\ \tau'_R \end{pmatrix} \tag{1.33}$$

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.

[IV] σ_2 is the second Pauli matrix.

[V] See Eq. 1.17.

[VI] 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:

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}}^{\text{lepton}} &= -\left(\frac{v+H}{\sqrt{2}}\right)\overline{\ell'_L}Y'^{\ell}\ell'_R + \text{h.c.}\,, \\ &= -\sum_{i=e,\mu,\tau}\frac{y_i^{\ell}v}{\sqrt{2}}\overline{\ell'_i}\ell'_i - \sum_{i=e,\mu,\tau}\frac{y_i^{\ell}}{\sqrt{2}}\overline{\ell'_i}\ell'_i H\,, \end{aligned} \tag{1.34}$$

where $\ell'_i \equiv \ell'_{Li} + \ell'_{Ri}$ $(i = e, \mu, \tau)$ have definite masses

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

In the following sections, by convention, the fermionic fields with definite mass, hereafter "mass eigenstates", will be indicated with a prime symbol to separate them from the ones with well-defined transformations under the electroweak gauge group, hereafter "flavor eigenstates".

The couplings y^{ℓ} are not constrained by theory and need to be measured experimentally.

In the SM, neutrino fields are massless, and can be used to reabsorb the flavor rotation in the charged weak current j_W^{μ}[VII], that would otherwise be modified.

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

where the massless neutrino fields $\nu_L = (\nu_{eL},\, \nu_{\mu L},\, \nu_{\tau L})^{\top}$ 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.

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

$$\begin{aligned} Y'^d &= V_L^d Y^d V_R^{d\dagger} \qquad \text{with} \quad Y'^d_{ij} = y_i^d\delta_{ij} \quad (i,j = d,s,b)\,, \\ Y'^u &= V_L^u Y^u V_R^{u\dagger} \qquad \text{with} \quad Y'^u_{ij} = y_i^u\delta_{ij} \quad (i,j = u,c,t)\,, \end{aligned} \tag{1.38}$$

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

where $V^u_{L,R}$ and $V^d_{L,R}$ are four appropriate 3×3 complex unitary matrices. To write the Lagrangian in matrix form, the quark flavor eigenstate vectors can be defined as in the lepton case

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

and transformed into the mass eigenstantes as

$$\begin{aligned} q'^d_L &= V^{d\dagger}_L q^d_L \equiv \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix}, \quad q'^d_R = V^{d\dagger}_R q^d_R \equiv \begin{pmatrix} d'_R \\ s'_R \\ b'_R \end{pmatrix}, \\ q'^u_L &= V^{u\dagger}_L q^u_L \equiv \begin{pmatrix} u'_L \\ c'_L \\ t'_L \end{pmatrix}, \quad q'^u_R = V^{u\dagger}_R q^u_R \equiv \begin{pmatrix} u'_R \\ c'_R \\ t'_R \end{pmatrix}, \end{aligned} \tag{1.40}$$

After the change of basis and a rewrite in the unitary gauge, the quark Yukawa terms become

$$\begin{aligned} \mathcal{L}^{\text{quark}}_{\text{Yukawa}} &= -\left(\frac{v+H}{\sqrt{2}}\right)\left[\overline{q'^d_L} Y'^d q'^d_R + \overline{q'^u_L} Y^u q'^u_R\right] + \text{h.c.} \\ &= -\left(\frac{v+H}{\sqrt{2}}\right)\left[\sum_{i=d,s,b} y^d_i \overline{q'^d_{Li}} q'^d_{Ri} + \sum_{i=u,c,t} y^u_i \overline{q'^u_{Li}} q'^u_{Ri}\right] + \text{h.c.}\,, \end{aligned} \tag{1.41}$$

and as in the lepton case, the rotated fields have a definite mass

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

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

Unlike in leptons, however, the transformations in Eq. 1.40 cannot be reabsorbed. The charged weak current for quarks can then be derived from the electroweak Lagrangian in Eq. 1.5 and written in matrix form as:

$$j^\mu_{W,q} = \underbrace{\overline{q^u_L}\gamma^\mu q^d_L}_{\text{flavor eigenstates}} = \underbrace{\overline{q'^u_L} V^{u\dagger}_L \gamma^\mu V^d_L q'^d_L}_{\text{mass eigenstates}} = \overline{q'^u_L}\gamma^\mu \underbrace{V^{u\dagger}_L V^d_L}_{\equiv \boldsymbol{V}_{\text{CKM}}} q'^d_L\,. \tag{1.44}$$

Equation 1.44 shows that the charged weak current in the quark sector does not depend separately on the matrices V_L^u and V_L^d, but only on their product

$$\boldsymbol{V}_{\mathrm{CKM}} = V_L^u V_L^{d\dagger} \,. \tag{1.45}$$

The matrix $\boldsymbol{V}_{\mathrm{CKM}}$ is the Cabibbo–Kobayashi–Maskawa (CKM) matrix [3, 22] and contains information relative to the physical effects of quark mixing.

2

FLAVOR PHYSICS AND CP VIOLATION

Symmetries play an important role in the fundamental description of nature. Continuous symmetries like translations or rotations in the space-time lead to conservation laws as for Noether's Theorem [23] and discrete ones impose strong constraints on the possible behaviour of interactions.

Three discrete symmetries, and combination of them, are relevant in the context of the Standard Model: parity (P), charge conjugation (C), and time reversal (T). C and CP violations were identified by Sakharov as one of the basic conditions for the matter-antimatter asymmetry which is observed in the Universe. While all three symmetries are conserved by the electromagnetic and strong interactions, the weak interaction was observed to violate all of them [24, 25]. Only the product of the three symmetries, CPT, is conserved in the SM, and in fact, the CPT theorem states that any Lorentz invariant local quantum field theory with an Hermitian Hamiltonian must have CPT symmetry [26–28].

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 strong and electromagnetic interactions. However, it is violated maximally in weak interactions [29, 30].

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

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

where the charge conjugation operator C is unitary. This is maximally violated in weak interaction: right-handed neutrinos for example, unlike left-handed ones, 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:

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

This is a near-symmetry of the SM Lagrangian[I]. Unlike P and C, the time reversal operator T is antiunitary. Since the CP symmetry is violated by the weak interaction, necessarily the T symmetry must be too, as otherwise it would be impossible to conserve CPT. T violation has been observed directly in decays of B^0 meson [31].

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

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

where

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

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

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

The fields in Eq. 2.5 represent the flavor eigenstates, which as stated in Section 1.3.2 do not coincide to the mass eigenstates in quarks.

[I] 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 $V^*_{qq'}$.

The quark term in Eq. 2.4 can be rewritted in terms of the mass eigenstates using Eq. 1.44

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

It is easy to see in Eq. 2.6 that the up- and down-type quarks, from different generations i and j, couple with each other with a 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

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 quark flavor eigenstates[II] 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 $\left|V_{ij}\right|^2$.

The current best determination of the magnitudes of the CKM elements is [24]:

$$|\mathbf{V}_{\text{CKM}}| = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361\,^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053\,^{+0.00083}_{-0.00061} \\ 0.00854\,^{+0.00023}_{-0.00016} & 0.03978\,^{+0.00082}_{-0.00060} & 0.999172\,^{+0.000024}_{-0.000035} \end{pmatrix}. \tag{2.8}$$

It is easy to observe that the CKM matrix is nearly diagonal, with off-diagonal terms that are very small and diagonal terms that are very close to one. This is reflected in the suppression of transitions between different quark families. The

---

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

hierarchy in the matrix elements can be used to define approximate parametrizations expanding the small differences from the diagonal form.

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 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 [32].

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

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

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 $\boldsymbol{V}_{\text{CKM}}$, CP violation is expected in all three generations of quarks.

The Wolfenstein representation, instead, is based on the observed hierarchy among the parameters and is obtained by expanding the CKM terms in powers of $\lambda = s_{12} \simeq 0.23$. In this parametrization, the $\boldsymbol{V}_{\text{CKM}}$ matrix can be approximated to the order λ^3 as

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

where A, ρ and η are real parameters of order unity ($A \simeq 0.79$, $\rho \simeq 0.14$, $\eta \simeq 0.36$ [24]). In the approximation, CP is violated if $\eta \neq 0$. An alternative and widely used convention for the parameters ρ and η is:

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

2.3.2 *The unitary triangles*

Requiring unitarity in the CKM matrix[III] 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 implies 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 constraints can be written as

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

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

The three terms in the sum are vectors in the complex plane, thus each relation can be represented geometrically as a triangle, called "unitary triangle".

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

$$V_{us}V^*_{ub} + V_{cs}V^*_{cb} + V_{ts}V^*_{tb} = 0\,, \tag{2.15}$$

with angles:

$$\alpha_s = \arg\left(-\frac{V_{ts}V^*_{tb}}{V_{us}V^*_{ub}}\right), \quad \beta_s = \arg\left(-\frac{V_{ts}V^*_{tb}}{V_{cs}V^*_{cb}}\right), \quad \gamma_s = \arg\left(-\frac{V_{us}V^*_{ub}}{V_{cs}V^*_{cb}}\right). \tag{2.16}$$

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

[III] $V^\dagger V = VV^\dagger = \mathbb{1}_3$.

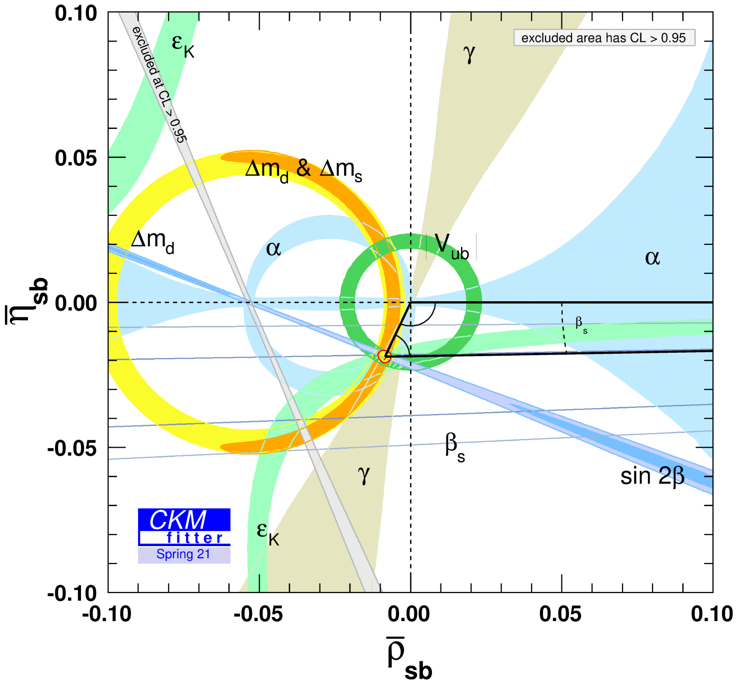


Figure 2.2: Constraints and global fit results for the B_s^0 unitary triangle. The red hashed region of the global combination for the vertex position corresponds to a 68% confidence level (CL). Figure from Ref. [33].

2.4 NEUTRAL B MESONS MIXING

Due to the structure of the charged current of the weak interaction neutral mesons exhibit an oscillation behaviour, i.e. there is a time dependent probability for the transition of a given flavor eigenstate, e.g. B_s^0, into its antiparticle, $\overline{B}_s^0$. An example of the leading order Feynman diagrams contributing to this process is given in Fig. 2.3. The oscillation is a direct consequence of the disparity between the flavor eigenstates and the mass eigenstates of the quarks which leads to a mixing between the quark generations by the CKM mechanism. The following discussion concentrates on the oscillation in the B_s^0 system. The formulation is equivalent for all neutral mesons.

2.4.1 *Mass and flavor*

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

$$|\,\Psi(t)\,\rangle = \left|\,B_s^0(t)\,\right\rangle + \left|\,\overline{B}_s^0(t)\,\right\rangle \tag{2.17}$$

with time evolution governed by the Schrödinger equation:

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \left|\,B_s^0(t)\,\right\rangle \\ \left|\,\overline{B}_s^0(t)\,\right\rangle \end{pmatrix} = \boldsymbol{H}\begin{pmatrix} \left|\,B_s^0(t)\,\right\rangle \\ \left|\,\overline{B}_s^0(t)\,\right\rangle \end{pmatrix}. \tag{2.18}$$

The effective Hamiltonian $\boldsymbol{H}$ can be split using 2×2 Hermitian matrices, as

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

where:

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

Given the structure of Eq. 2.18, it's easy to notice that the diagonal and off-diagonal terms of $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ are associated with the flavor-preserving transitions $B_s^0 \leftrightarrows B_s^0$ ($\overline{B}_s^0 \leftrightarrows \overline{B}_s^0$) and with the flavor-changing transitions $B_s^0 \leftrightarrows \overline{B}_s^0$ ($\overline{B}_s^0 \leftrightarrows B_s^0$), respectively. The CPT theorem also requires the diagonal terms of $\boldsymbol{H}$ to be equal, which leads to

$$\text{CPT} \quad \leftrightarrow \quad \begin{matrix} M_{11} = M_{22} \\ \Gamma_{11} = \Gamma_{22} \end{matrix}\,, \tag{2.19}$$

where M_{ij} and Γ_{ij} are elements of the $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ matrices, respectively.

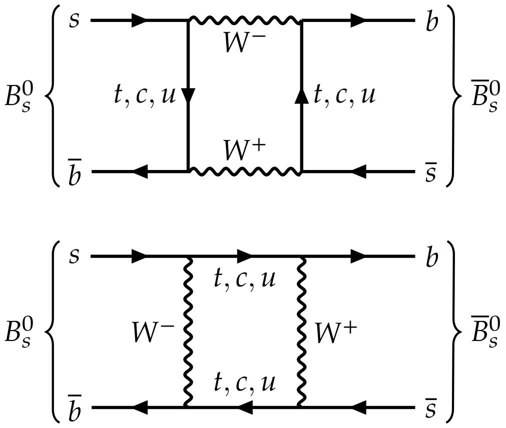


Figure 2.3: Leading order diagrams for B_s^0–$\overline{B}_s^0$ mixing.

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 as a rotation of the flavor states as:

$$\begin{aligned} |\, B_s^L \,\rangle &= p|\, B_s^0 \,\rangle + q|\, \overline{B}_s^0 \,\rangle\,, \\ |\, B_s^H \,\rangle &= p|\, B_s^0 \,\rangle - q|\, \overline{B}_s^0 \,\rangle\,, \end{aligned} \tag{2.20}$$

where p and q are complex coefficients that obey 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:

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} |\, B_s^L(t) \,\rangle \\ |\, B_s^H(t) \,\rangle \end{pmatrix} = \begin{pmatrix} M_L - \frac{i}{2}\Gamma_L & 0 \\ 0 & M_H - \frac{i}{2}\Gamma_H \end{pmatrix}\begin{pmatrix} |\, B_s^L(t) \,\rangle \\ |\, B_s^H(t) \,\rangle \end{pmatrix}, \tag{2.21}$$

where the 2×2 matrix has been obtained by diagonalizing $\boldsymbol{H}$ in the eignevector basis:

$$\boldsymbol{Q}^{-1}\boldsymbol{H}\boldsymbol{Q} = \begin{pmatrix} M_L - \frac{i}{2}\Gamma_L & 0 \\ 0 & M_H - \frac{i}{2}\Gamma_H \end{pmatrix}, \quad \text{with } \boldsymbol{Q} = \begin{pmatrix} q & p \\ p & -q \end{pmatrix}. \tag{2.22}$$

From Eq. 2.22, the ratio of the complex coefficients p and q is given by:

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}}\,. \tag{2.23}$$

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

$$\left| B_s^{H,L}(t) \right\rangle = e^{-iM_{H,L}t-\frac{1}{2}\Gamma_{H,L}t}\left| B_s^{H,L}(t) \right\rangle . \tag{2.24}$$

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 eigenstates can be defined as:

$$m_s \equiv \frac{M_H + M_L}{2} = M_{11} = M_{22} , \tag{2.25}$$

$$\Gamma_s \equiv \frac{\Gamma_H + \Gamma_L}{2} = \Gamma_{11} = \Gamma_{22} , \tag{2.26}$$

$$\Delta m_s \equiv M_H - M_L , \tag{2.27}$$

$$\Delta\Gamma_s \equiv \Gamma_L - \Gamma_H . \tag{2.28}$$

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

$$\Delta m_s \simeq 2|M_{12}| , \tag{2.29}$$

$$\Delta\Gamma_s \simeq 2|\Gamma_{12}| \cos\phi , \tag{2.30}$$

$$\frac{q}{p} \simeq \sqrt{\frac{M_{12}^*}{M_{12}}} , \tag{2.31}$$

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

$$\left|\frac{q}{p}\right| \simeq 1 , \tag{2.32}$$

which is supported by both experimental results and theoretical predictions [24] [IV]. Since its module is $\simeq 1$, the ratio q/p can be expressed in terms of a complex phase as:

$$\frac{q}{p} \simeq e^{-i\phi_M} , \tag{2.33}$$

where $\phi_M = \arg(M_{12})$. Since the top quark is the dominant contribution in the mixing diagrams (Fig. 2.3) [34], this can further be expressed in terms of CKM elements as

$$\frac{q}{p} \simeq \frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}} = e^{-i\phi_M} , \tag{2.34}$$

where $M_{12} \propto \left(V_{ts}^*V_{tb}\right)^2$ has been used.

[IV] $|q/p|_{B_s^0} = 1.0003 \pm 0.0014$ (experiment), $|q/p|_{B_s^0} = 0.999989 \pm 0.000001$ (theory) [24].

2.4.3 *Time evolution of the flavor eigenstates*

Using Eqs. 2.20 and 2.24, the time evolution of the flavor eigenstates can be described as:

$$\begin{aligned}
|\, B_s^0(t)\, \rangle &= g_+(t)|\, B_s^0\, \rangle + \frac{q}{p} g_-(t)|\, \overline{B}_s^0\, \rangle\,, \\
|\, \overline{B}_s^0(t)\, \rangle &= g_+(t)|\, \overline{B}_s^0\, \rangle + \frac{p}{q} g_-(t)|\, B_s^0\, \rangle\,,
\end{aligned} \tag{2.35}$$

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

$$\begin{aligned}
g_+(t) &= e^{-im_s t} e^{-\frac{\Gamma_s}{2}t}\left[\quad \cosh\left(\frac{\Delta\Gamma_s t}{4}\right)\cos\left(\frac{\Delta m_s t}{2}\right) - i\sinh\left(\frac{\Delta\Gamma_s t}{4}\right)\sin\left(\frac{\Delta m_s t}{2}\right)\right], \\
g_-(t) &= e^{-im_s t} e^{-\frac{\Gamma_s}{2}t}\left[-\sinh\left(\frac{\Delta\Gamma_s t}{4}\right)\sin\left(\frac{\Delta m_s t}{2}\right) + i\cosh\left(\frac{\Delta\Gamma_s t}{4}\right)\cos\left(\frac{\Delta m_s t}{2}\right)\right].
\end{aligned} \tag{2.36}$$

From Eq. 2.35 the probability that a particle produced as B_s^0 ($\overline{B}_s^0$) at $t = 0$ has oscillated to $\overline{B}_s^0$ (B_s^0) after time t can be calculated as

$$\begin{aligned}
\left|\langle\, B_s^0(0)\,|\,\overline{B}_s^0(t)\,\rangle\right|^2 &= \left|\frac{p}{q}\right|^2 |g_-(t)|^2\,, \\
\left|\langle\, \overline{B}_s^0(0)\,|\,B_s^0(t)\,\rangle\right|^2 &= \left|\frac{q}{p}\right|^2 |g_-(t)|^2\,,
\end{aligned} \tag{2.37}$$

with

$$|g_-(t)|^2 = \frac{e^{-\Gamma_s t}}{2}\left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) - \cos(\Delta m_s t)\right]. \tag{2.38}$$

From Eq. 2.38 we can see that the B_s^0 meson flavor eigenstates oscillate with frequency Δm_s, modulated by an exponential decay with time constant $1/\Gamma_s$, itself 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.

While approximately true for B_s^0 mesons, the mixing probabilities in Eq. 2.37 do not have to be equal. In general the ratio $|q/p|$ does not have to be equal to one and 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 with time constant $1/\Gamma_s$, itself 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 [24]	Unit
m_s	5366.92 ± 0.10	MeV
Δm_s	17.765 ± 0.006	$\hbar\mathrm{ps}^{-1}$
Γ_s	0.6578 ± 0.0024	ps^{-1}
$\Delta\Gamma_s$	0.084 ± 0.005	ps^{-1}

2.5 CP VIOLATION IN B MESONS

To better describe CPV, the following amplitudes are defined for a given $B_s^0 \to f$ transition where a neutral B_s^0 meson decays into a final state f:

$$\begin{aligned} A_f &= \left\langle f \middle| \boldsymbol{H} \middle| B_s^0 \right\rangle, & \overline{A}_f &= \left\langle f \middle| \boldsymbol{H} \middle| \overline{B}_s^0 \right\rangle, \\ A_{\bar{f}} &= \left\langle \bar{f} \middle| \boldsymbol{H} \middle| B_s^0 \right\rangle, & \overline{A}_{\bar{f}} &= \left\langle \bar{f} \middle| \boldsymbol{H} \middle| \overline{B}_s^0 \right\rangle. \end{aligned} \tag{2.39}$$

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, so in describing CPV it is useful to use quantities that are invariant under such transformation, such as $\left|A_f\right|$, $\left|\overline{A}_f\right|$ and $|q/p|$. Another quantity invariant under phase redefinition and useful for parameterizing CP violation is the complex parameter λ_f, defined as:

$$\lambda_f \equiv \frac{q}{p}\frac{\overline{A}_f}{A_f}. \tag{2.40}$$

Using the equations defined in Section 2.4.3, the time-dependent decay rate $\Gamma_{B_s^0\to f}(t) \equiv \left|\left\langle f \middle| \boldsymbol{H} \middle| B_s^0(t) \right\rangle\right|^2$ of an initial state B_s^0 to decay into f after a time t, and the corresponding $\Gamma_{\overline{B}_s^0\to f}(t)$ rate, can be expressed as

$$\begin{aligned}\Gamma_{B_s^0\to f}(t) = \left|A_f\right|^2 \quad \frac{1}{1+C_f}e^{-\Gamma_s t}\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right)+D_f\sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ + C_f\cos(\Delta m_s t) - S_f\sin(\Delta m_s t)\Bigg]\,,\end{aligned} \tag{2.41}$$

$$\begin{aligned}\Gamma_{\overline{B}_s^0\to f}(t) = \left|A_f\right|^2\left|\frac{p}{q}\right|^2 \frac{1}{1+C_f}e^{-\Gamma_s t}\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right)+D_f\sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ - C_f\cos(\Delta m_s t) + S_f\sin(\Delta m_s t)\Bigg]\,,\end{aligned} \tag{2.42}$$

with

$$C_f \equiv \frac{1-\left|\lambda_f\right|^2}{1+\left|\lambda_f\right|^2}\,,\quad S_f \equiv \frac{2\,\mathrm{Im}(\lambda_f)}{1+\left|\lambda_f\right|^2} \quad \text{and} \quad D_f \equiv -\frac{2\,\mathrm{Re}(\lambda_f)}{1+\left|\lambda_f\right|^2}\,. \tag{2.43}$$

CP is violated if Eq. 2.41 and 2.42 are different, which is the case if either λ_f or $|q/p|$ are different from one. CP violation effects can be classified into various categories, depending on how they affect these variables.

2.5.1 *CP violation in decay*

Direct CP violation occurs when the decay rate of a B_s^0 to a final state f, represented by A_f and the decay rate of the CP-conjugated process $\overline{B}_s^0 \to \bar{f}$, represented by $\overline{A}_{\bar{f}}$, are not equal in module, meaning that

$$\left|\frac{\overline{A}_{\bar{f}}}{A_f}\right| \neq 1\,. \tag{2.44}$$

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

$$\lambda_f = \eta_{\mathrm{CP}}\frac{q}{p}\frac{\overline{A}_{\bar{f}}}{A_f}\,, \tag{2.45}$$

where the transformation of $|f\rangle$ under CP was used:

$$\left|\bar{f}\right\rangle = CP|f\rangle = \eta_{\mathrm{CP}}|f\rangle\,, \tag{2.46}$$

where $\eta_{\mathrm{CP}} = \pm 1$ is the CP eigenvalue of f. Eq. 2.45 shows that, assuming $|q/p| = 1$ (i. e. no CP violation in mixing), CP violation in decay is equivalent to a deviation from unity of $\left|\lambda_f\right|$.

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

$$\begin{aligned} A_f &= |A_f|\, e^{i(\theta+\phi)} \\ \overline{A}_{\overline{f}} &= |A_f|\, e^{i(\theta-\phi)}\,, \end{aligned} \tag{2.47}$$

where the amplitudes and strong phases are invariant under CP[V]. Physical observables are proportional to $\left|A_f\right|^2$, which in case of single amplitude returns the same number for A_f and $\overline{A}_{\overline{f}}$. To produce measurable CPV at least two interfering amplitudes are needed:

$$\begin{aligned} A_f &= |A_1|\, e^{i(\theta_1+\phi_1)} + |A_2|\, e^{i(\theta_2+\phi_2)}\,, \\ \overline{A}_{\overline{f}} &= |A_1|\, e^{i(\theta_1-\phi_1)} + |A_2|\, e^{i(\theta_2-\phi_2)}\,. \end{aligned} \tag{2.48}$$

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

$$\left|A_f\right|^2 - \left|\overline{A}_{\overline{f}}\right|^2 = -4|A_1||A_2|\sin(\theta_1-\theta_2)\sin(\phi_1-\phi_2)\,. \tag{2.49}$$

Equation 2.49 shows that to generate an observable direct CP violation, 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 $\sin(\theta_1-\theta_2) \neq 0$.

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, since they are not subject to mixing.

Direct CP violation has been discovered in kaon decay [35–37], and later observed in B [38, 39] and D [40] mesons decays.

2.5.2 *CP violation in mixing*

If the probability of transition $B_s^0 \to \overline{B}{}_s^0$ is different from the one of $\overline{B}{}_s^0 \to B_s^0$, the mixing process can produce a CP violation:

$$\left|\left\langle B_s^0(0)\middle|\overline{B}{}_s^0(t)\right\rangle\right|^2 \neq \left|\left\langle \overline{B}{}_s^0(0)\middle|B_s^0(t)\right\rangle\right|^2. \tag{2.50}$$

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

As shown in Eq. 2.37, the previous inequality is equivalent to

$$\left|\frac{p}{q}\right|^2 |g_-(t)|^2 \neq \left|\frac{q}{p}\right|^2 |g_-(t)|^2 \quad \leftrightarrow \quad \left|\frac{q}{p}\right| \neq 1 \quad \leftrightarrow \quad |q| \neq |p| \,. \tag{2.51}$$

Assuming no direct CP violation, this leads to a deviation from unity of $\left|\lambda_f\right|$ in the Standard Model. CP is conserved in mixing of B mesons to a high degree, since $|q/p| \approx 1$ [24], but it has been observed to be violated in K^0 oscillations [2].

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

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

$$\Gamma\Big(B^0_{s\,(\rightsquigarrow \overline{B}^0_s)} \rightarrow f\Big)(t) \neq \Gamma\Big(\overline{B}^0_{s\,(\rightsquigarrow B^0_s)} \rightarrow f\Big)(t)\,. \tag{2.52}$$

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

$$\mathrm{Im}\big(\lambda_f\big) \neq 0\,. \tag{2.53}$$

Using Eq. 2.41 and 2.42, and assuming $|q/p| \approx 1$, the time dependent CP asymmetry can be expressed as:

$$a_{\mathrm{CP}}(t) = \frac{\Gamma_{\overline{B}^0_s \to f}(t) - \Gamma_{B^0_s \to f}(t)}{\Gamma_{\overline{B}^0_s \to f}(t) + \Gamma_{B^0_s \to f}(t)} = \frac{S_f \sin(\Delta m_s t) - C_f \cos(\Delta m_s t)}{\cosh\Big(\frac{\Delta\Gamma_s t}{2}\Big) - D_f \sinh\Big(\frac{\Delta\Gamma_s t}{2}\Big)}\,. \tag{2.54}$$

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

CP violation in the interference of mixing and decay has been observed in K^0 and B^0 mesons [25, 41].

Part II

EXPERIMENTAL SETUP

The Large Hadron Collider, located at CERN in Geneva, 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 40 million collisions per second. During the data-taking periods of interest for this work, approximately $4 \cdot 10^{13}$ b-hadrons have been produced by the LHC at each of the interaction points around which the experiments are located. The CMS detector is a large general-purpose detector, consisting of several sub-systems capable of measuring 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 and the steps to extract the high-level features used in later chapters.

3

THE CMS DETECTOR

The taggers presented in this dissertation, as well as the CPV analysis used as benchmark, are developed on a data sample of $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ candidates collected in proton-proton (pp) collisions provided by the Large Hadron Collider (LHC) and collected by the Compact Muon Solenoid (CMS) experiment.

LHC is, at the time of writing, the largest and most powerful particle accelerator ever built, and CMS is one of the detectors using its beams to measure different types of particles and their properties. The LHC is managed by the European Organization for Nuclear Research (CERN), near Geneva, whose main function is to provide the particle accelerator and other infrastructure necessary for particle physics research. In the field of particle physics, many important achievements have been made by CERN experiments, such as the discovery of weak neutral currents in the Gargamelle bubble chamber [42], the discovery of the W and Z bosons by the UA1 and UA2 experiments [43, 44], the discovery of direct CP violation in the NA48 experiment [37], the determination of the number of light neutrino families by the LEP experiments [45–48], and, most recently, the Higgs boson discovery by the ATLAS and CMS experiments [49, 50].

This section describes the experimental setup of the LHC and CMS experiments.

3.1 THE LARGE HADRON COLLIDER

The LHC is a dual-ring hadron accelerator, storage ring, and collider situated in the 26.7 km tunnel formerly built for the Large Electron-Positron Collider (LEP)[I], at a depth between 45 m and 170 m underground on the French-Swiss border near Geneva. The LHC is designed to accelerate protons to energies up to 7 TeV and heavy ions[II] at 2.76 TeV. The first particle collision was achieved in 2010, after commissioning in 2008. All the taggers and the ϕ_s measurement presented in this thesis were tuned for pp collisions, so the following description only considers protons.

3.1.1 *Acceleration chain*

Before injection into the LHC, protons are pre-accelerated by the accelerator chain shown in Fig. 3.1. A bottle of molecular hydrogen gas is used as the proton

---

[I] 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.

[II] Predominantly lead nuclei.

Table 3.1: Value of the LHC parameters for the three year of Run2 and by design [52] [53].

Parameter		2016	2017	2018	Design
Center-of-mass energy ($\sqrt{s}$)	[TeV]	13	13	13	14
Instant. luminosity (L)	[$10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$]	1.4	1.7	2.1	1
Norm. emittance (ϵ_n)	[μm]	2.2	2.2	1.9	3.75
β at the IP (β^*)	[cm]	40	40	30→25	55
Bunch time spacing	[ns]	25	25	25	25
N. of bunches (n_b)	-	2220	2556	2556	2808
Protons per bunch	[10^{11}]	1.1	1.2	1.1	1.15

source, 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 [51]. The protons are then 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 into the LHC at 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 was achieved in each direction.

In the LHC, the beams are accelerated 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 a rate of 16 MeV per turn. The acceleration process takes about 20 to 25 minutes and the time period for which the LHC has bunches circulating is called *"fill"*. The fill is typically maintained up to thirty hours. Around the beam pipes, 1232 dipole superconducting magnets are installed to bend the beams into a circular trajectory. Additionally, quadrupole and higher multipole[IV] magnets are used to focus the beam despite the electromagnetic repulsion of the protons. At the LHC interaction points, the beams can be confined to a diameter of the order of $\approx 10\ \mu$m. A cross section of an LHC dipole magnet is shown in Fig. 3.2.

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

$$\sqrt{s} = 2E_{\text{beam}}\,, \tag{3.1}$$

although the energy available for the hard parton scattering is smaller and depends on the amount of energy carried by the involved parton. The beams are

---

[III] This is slightly less than the final design energy of 7 TeV per beam mentioned before. In the recently started Run3, this was increased to 6.8 TeV.

[IV] Sextupoles and octupoles.

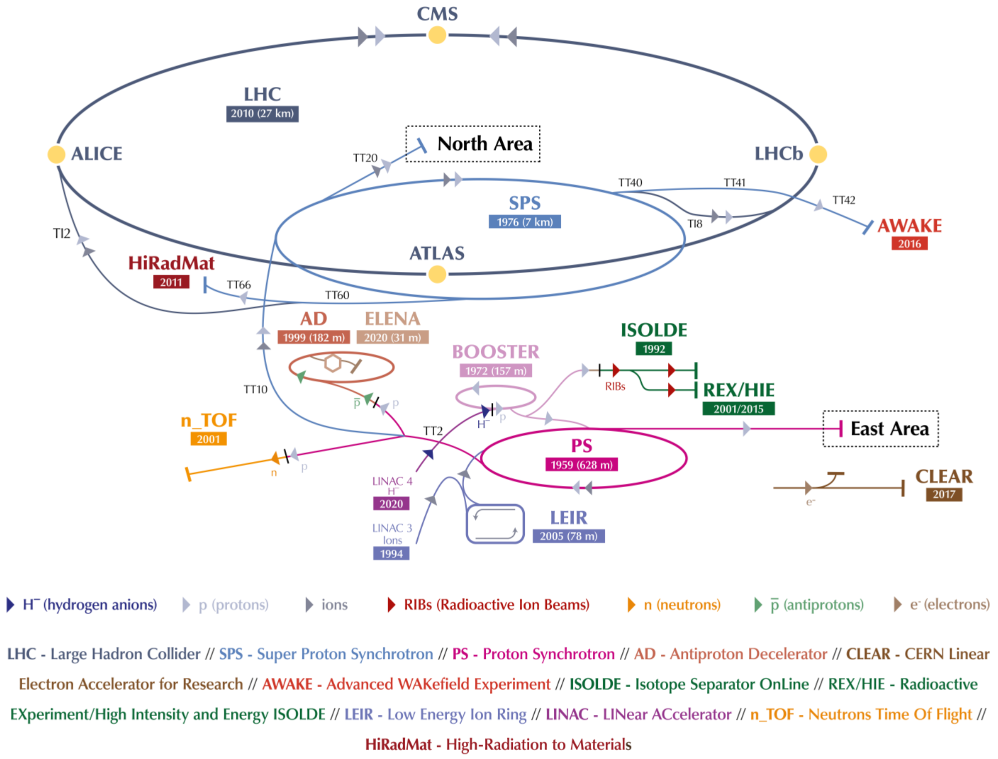


Figure 3.1: The CERN accelerator complex. Figure from Ref. [54].

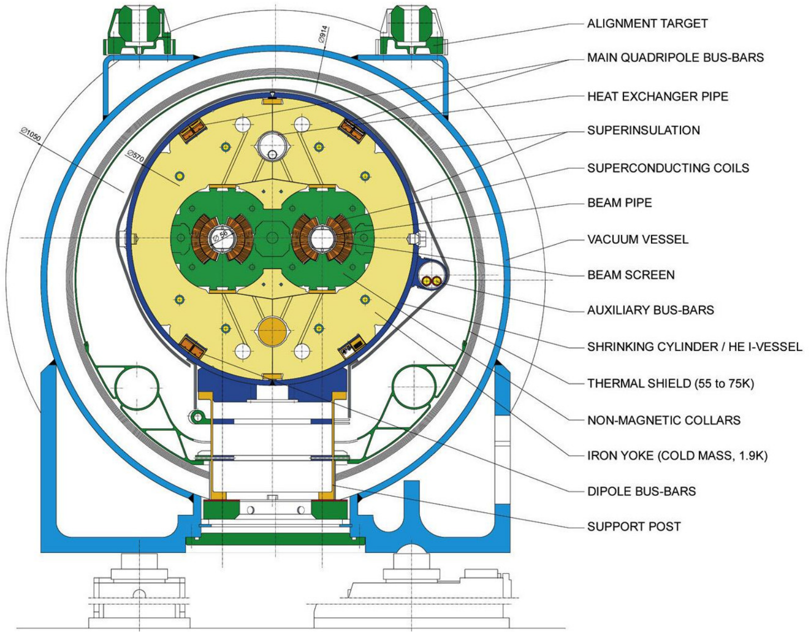


Figure 3.2: Cross section of an LHC dipole magnet. Figure adapted from Ref. [55].

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 [56],
- *ATLAS*, a general-purpose detector [57],
- *CMS*, a general-purpose detector (thoroughly described in Section 3.2),
- *LHCb*, a detector operating at low luminosity, dedicated to heavy flavor physics [58].

Both the ATLAS and CMS experiments have similar physics programs, but they achieve them with different strategies in the detector layout. The main difference is that in CMS both the tracking and calorimetry systems are housed inside the superconducting solenoid, whereas in ATLAS only the tracking system is. On average, the performance of the two detectors is similar, but the performance on individual analyses may differ. While the LHCb experiment would be a prime candidate for comparisons, since this dissertation concerns heavy flavor physics phenomenology, the ATLAS detector is more similar in terms of phase space and absence of a dedicated hadronic particle identification (PID) system.

3.1.2 *Particle production*

At every bunch crossing, only few of the protons actually collides. In most cases, the particles simply graze each other in what is called an *elastic* collision, in which the proton structure is unaltered. 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 interesting physics processes occur. The rate of collision events for a given process is given by:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = L\,\sigma\,, \tag{3.2}$$

where σ is the process cross section (dimensionally an area, measured in barn[V] or smaller fractions), and L is the instantaneous luminosity of the accelerator. Instantaneous luminosity measures the ability of a particle accelerator to produce interactions. The machine luminosity only depends on the beam parameters and can be calculated as [59]:

$$L = \frac{N_b^2 n_b f_{\mathrm{rev}} \gamma_r F}{4\pi \epsilon_n \beta^*}\,, \tag{3.3}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic Lorentz factor, ϵ_n is the normalized transverse beam emittance, and β^* is the beta function at the

---

[V] $1\mathrm{b} = 10^{-24}\mathrm{m}^2$.

interaction point[VI]. The factor F describes a reduction due to crossing angle at the interaction point and can be computed as

$$F = 2\left[1 + (\theta_c \sigma_z / 2\sigma^*)^2\right]^{-\frac{1}{2}}, \tag{3.4}$$

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. The design instantaneous luminosity of the LHC is $10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$, but has been exceeded in the 2017 and 2018 runs by a factor of ≈ 2 [60]. The nominal luminosity described by Eq. 3.3 is typically achieved only at the beginning of a fill and decays over time due to collisions and beam losses. The instantaneous luminosity decreases exponentially over time, with a time scale, called "beam lifetime", which in the 2017 and 2018 runs was equal to approximately 15 hours.

Following Eq. 3.2, the total number of events for a given process is given by:

$$N = \int \frac{\mathrm{d}N}{\mathrm{d}t}\,\mathrm{d}t = \sigma \int L\,\mathrm{d}t \equiv \sigma\,\mathcal{L}_{\mathrm{int}}, \tag{3.5}$$

where $\mathcal{L}_{\mathrm{int}}$ is the so-called "integrated luminosity" that is proportional 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[VII], which corresponds to an average of ≈ 30 collisions per bunch crossing [61]. An overview of the inelastic proton-proton cross-section as a function of $\sqrt{s}$ [62] can be found in Fig. 3.3. Processes of interest typically have a very small cross section, thus it is extremely rare than more than one occurs in the same bunch crossing. The remaining, *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. 3.4 and 3.5, while the mean number of interactions per bunch crossing is shown in Fig. 3.6.

BOTTOM QUARK PRODUCTION AT THE LHC At the LHC bottom quarks are predominantly produced in $pp \rightarrow b\bar{b}X$ 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. 3.7 and 3.8.

The predicted NNLO $b\bar{b}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV is $\sigma(pp \rightarrow b\bar{b}X) \approx 500\ \mu\mathrm{b}$ [63], and its estimated value as a function of the center

---

[VI] 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 distance s from the interaction point, is given by $\sigma(s) = \sqrt{\epsilon_n \beta(s)}$, where $\beta(s) = \beta^*[1 + (s/\beta^*)^2]$.

[VII] The value of $\sigma^{pp}_{\mathrm{inelastic}} \approx 70$ mb is the "CMS nominal" value for CMS analyses [61] and is determined by finding the best agreement with measured data.

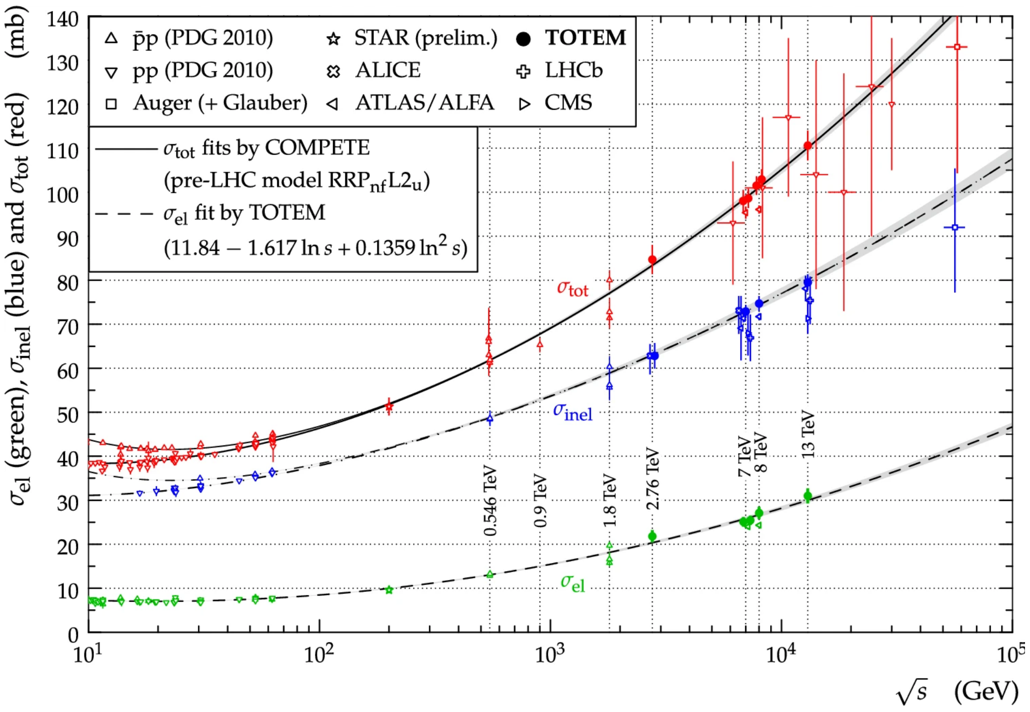
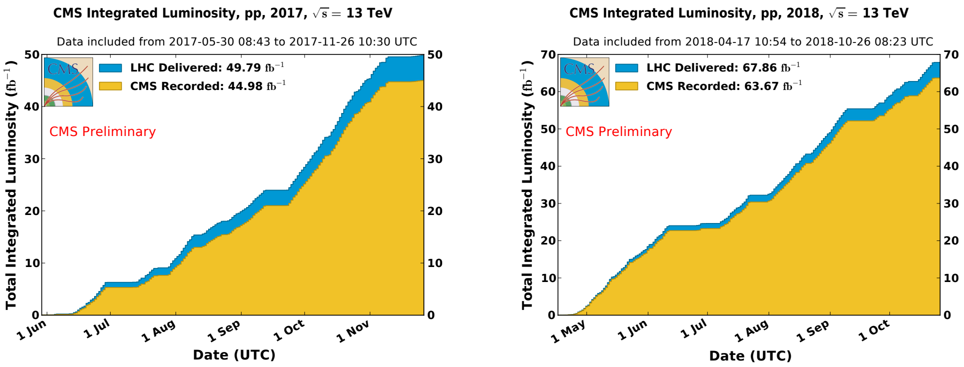


Figure 3.3: Overview of the elastic ($\sigma_{\rm el}$), inelastic ($\sigma_{\rm inel}$) and total ($\sigma_{\rm tot}$) cross-section for pp collisions as a function of $\sqrt{s}$. The continuous black lines (lower for pp, upper for $p\bar{p}$) represent the best fits of the total cross section data by the COMPETE collaboration. The dashed line results from a fit of the elastic cross section data. The dash-dotted lines refer to the inelastic cross section and are obtained as the difference between the continuous and dashed fits [62]

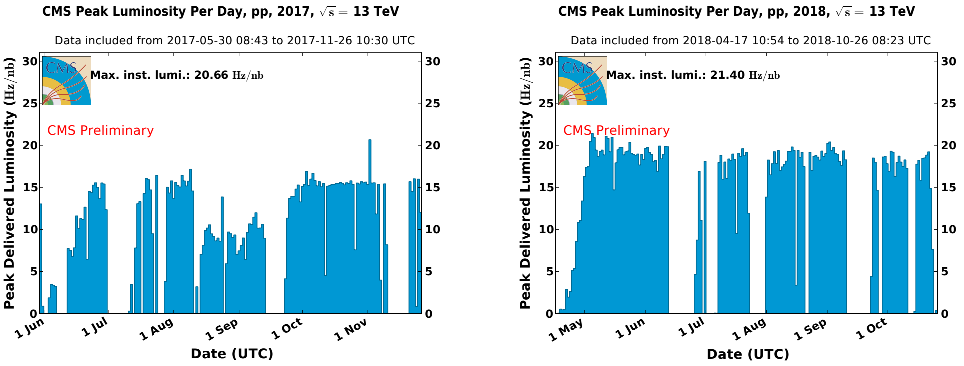
.



(a) 2017 data-taking period.

(b) 2018 data-taking period.

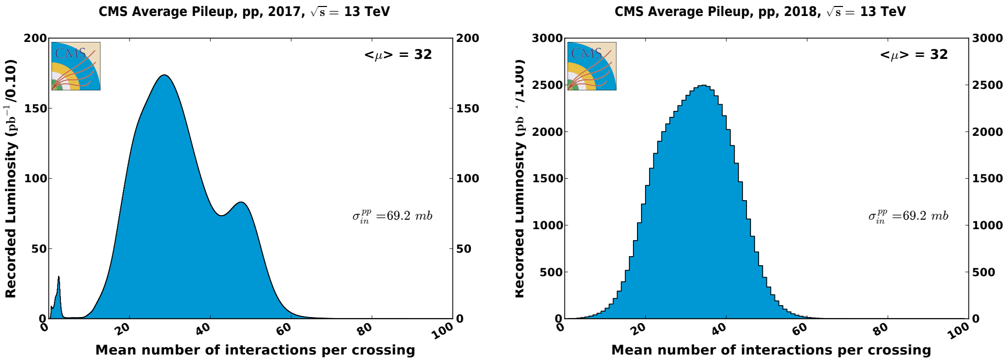
Figure 3.4: Cumulative measured luminosity versus day delivered by the LHC to CMS during the 2017 (left) and 2018 (right) pp run at 13 TeV. Figure from Ref. [61].



(a) 2017 data-taking period.

(b) 2018 data-taking period.

Figure 3.5: 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. [61].



(a) 2017 data-taking period.

(b) 2018 data-taking period.

Figure 3.6: 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. 3.5. Figure from Ref. [61].

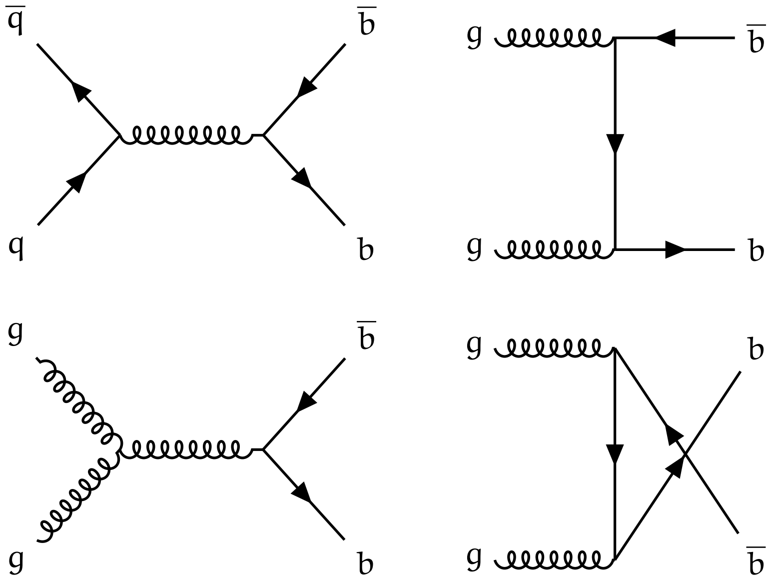


Figure 3.7: Leading order Feynman diagrams for b-quark pair production: $q\bar{q}$ annihilation (top left) and gluon-gluon fusion (others).

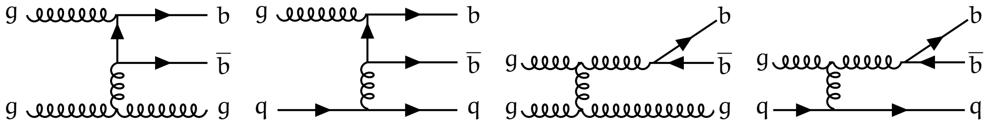


Figure 3.8: Next-to-leading order Feynman diagrams for $b\bar{b}$ pair production: flavor excitation (left and middle-left) and gluon splitting (middle-right and right).

of mass energy is shown in Fig. 3.9. The total number of $b\bar{b}$ pairs produced during the 2017 and 2018 data-taking periods can be calculated with Eq. 3.5, as:

$$N_{b\bar{b}} = \sigma(pp \to b\bar{b}X) \cdot \mathcal{L}_{\mathrm{int}} \approx 5.5 \cdot 10^{13}\,, \tag{3.6}$$

where the integrated luminosity values reported in Fig. 3.4 are used. Note that this is the total number of $b\bar{b}$ pairs produced in pp collisions and not the number of those produced in the geometric acceptance of CMS, which is of course lower.

The fact that bottom quarks are mostly produced in opposite flavor pairs has important consequences in the development of flavor tagging algorithms, as the second (non-signal) quark provides an additional handle for the determination of the flavor. This will be discussed in depth in Section 6.3.

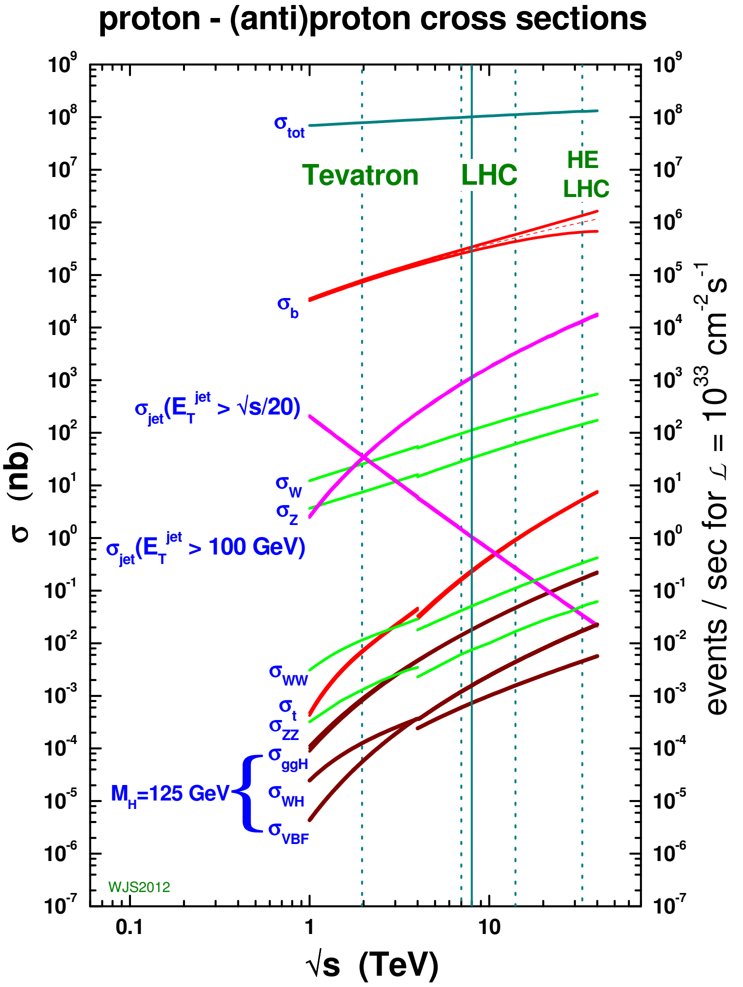


Figure 3.9: Production cross section at NLO in *pp* collisions of several process of interests as a function of the center-of-mass energy. Figure from Ref. [64].

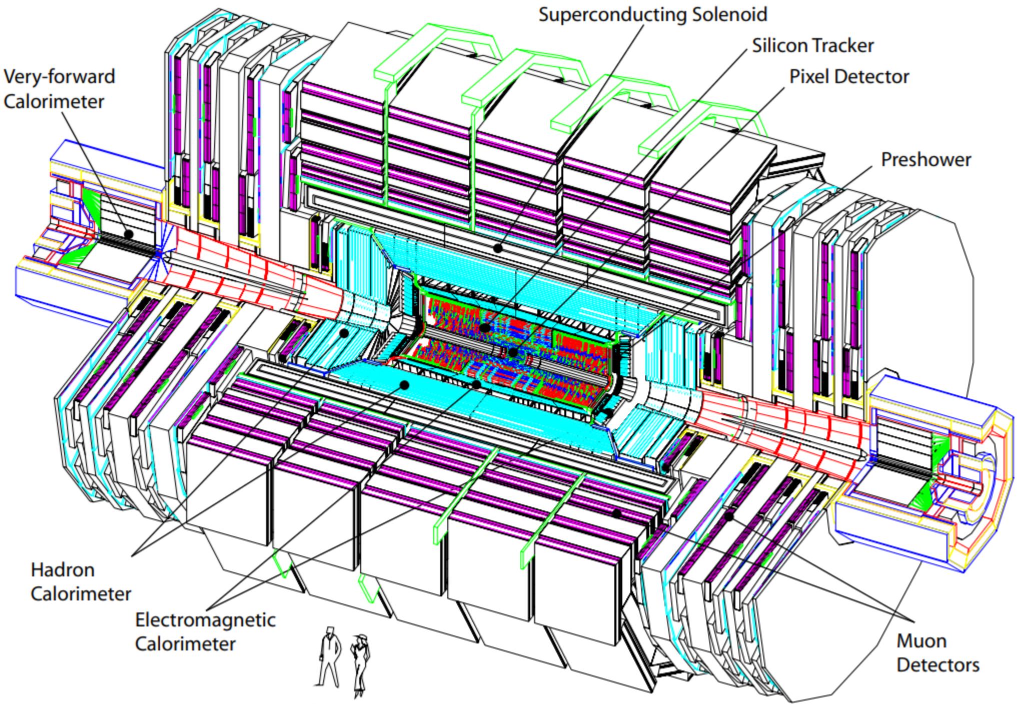


Figure 3.10: Overview of the CMS detector and its sub-systems. Figure from Ref. [65].

3.2 THE CMS DETECTOR

The Compact Muon Solenoid experiment is one of two large general-purpose particle physics detectors built at the LHC[VIII]. The goal of the CMS experiment is to investigate a wide range of physics, including the search for the Higgs boson, extra dimensions, and particles that could make up dark matter[65]. It is located on the north side of the LHC ring, near the city of Cessy, 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 and, with a total weight of about 14 000 t, is the heaviest LHC detector. An overview of the CMS detector and its subsystems is shown in Fig. 3.10. At the time of writing, the CMS Collaboration is composed of more than 6000 scientists, engineers, technicians, and students from around 250 institutions from more than 55 countries.

COORDINATE SYSTEM The coordinate system used by CMS to describe positions and directions is 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. In this dissertation, Cartesian, cylindrical, and spherical co-

---

[VIII] The other being ATLAS.

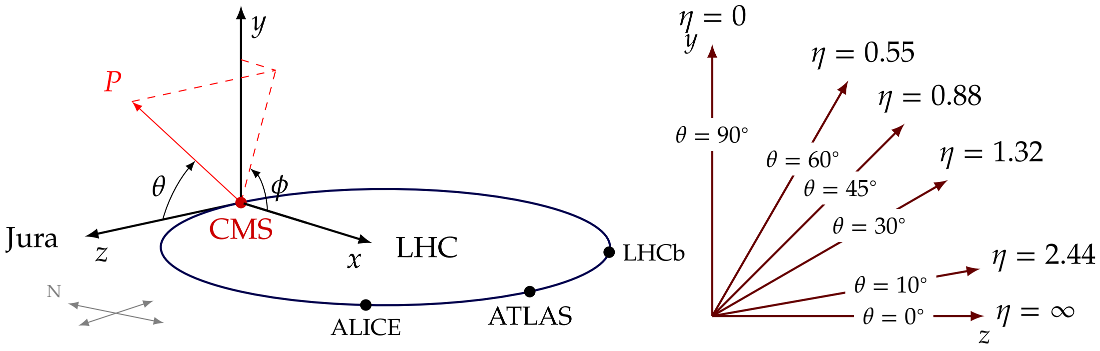


Figure 3.11: Illustration of the CMS coordinate system (left) and the relation between the pseudorapidity η and the polar angle θ (right). "Jura" refers to the mountain range near CERN. Figures from Ref. [66].

ordinates are used, depending on the cases. The azimuthal angle ϕ is measured in the x-y plane (often called "transverse plane") relative to the x axis. The coordinate system based on pseudorapidity and transverse momentum p_T (the projection on the transverse plane) is often preferred to the pure cylindrical or spherical coordinates. The pseudorapidity is defined as a function of the polar angle θ (measured from the z axis) as

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right). \tag{3.7}$$

Its main advantage is that the particle production rate in hadron colliders is approximately constant as a function of η. Additionally, pseudorapidity differences are approximately 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. 3.11.

3.2.1 *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[IX]. The detector is divided into a *barrel* component and two *endcaps*, which typically correspond to different types of detector technologies.

SUPERCONDUCTING SOLENOID The core characteristic of CMS is a large superconducting solenoid of 12.5 m length and 6.3 m internal diameter [67], which is used to generate a uniform magnetic field of 3.8 T inside the solenoid. The magnetic field is used to bend the trajectory of charged particles, which enables

---

[IX] 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.

the measurement of their momenta and the sign of their charge. The solenoid is made up of four layers of coiled superconducting NbTi, the same technology used in the LHC magnets, cooled to 4.6 K with liquid helium, and conducts 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. As complement to the magnet, a 12 500 t iron return yoke is used, with 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. In the return yoke the magnetic field saturates to a value of 1.8 T. A mapping of the magnetic field generated by the superconducting solenoid is shown in Fig. 3.12.

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 momenta, charge, and trajectories of the charged particles [69–72]. The tracker is surrounded by the calorimeter, which comprises a lead tungstate crystal electromagnetic calorimeter (ECAL) [73] and a brass-scintillator hadron calorimeter (HCAL) [74]. The ECAL detects particles that interact electromagnetically, such as electrons, positrons, and photons, whereas the HCAL is designed to detect charged and neutral hadrons, i. e., protons, neutrons, pions, and kaons. Finally, outside the solenoid, the gas detectors that make up the muon system are interleaved to the layers of the return yoke, which has the additional purpose of absorbing most of the particles different from muons [75]. A transverse slice of the CMS detector and its subsystem is shown in Fig. 3.13. The following sections present further details of each subdetector ordered by distance from the interaction point, while the trigger system and physics object reconstruction are described in Chapter 4.

3.2.2 *Tracking system*

The CMS tracking system ("tracker") is installed near the interaction point and is entirely based on silicon semiconductor technology [69, 70]. 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[X]. Since this analysis uses only data collected between 2017 and 2018, this section describes only the CMS *Phase-1* tracker, which includes the upgraded inner pixel detector installed in March 2017 [71, 72].

Being 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 bunch crossing (25 ns), which implies a rate density of 1 MHz/mm^2 at

---

[X] The vertex reconstruction procedure is described in Section 4.2.1.

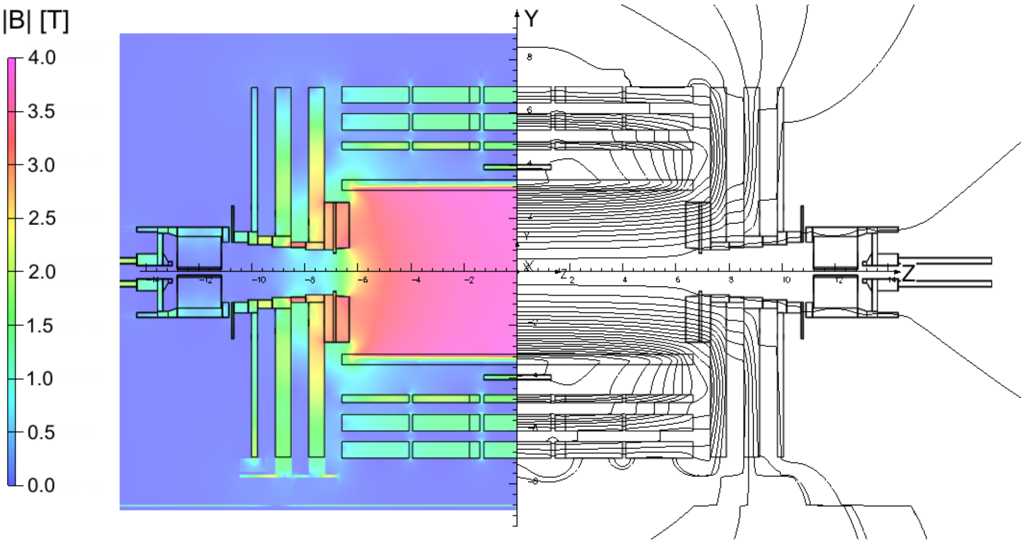


Figure 3.12: Value of $|B|$ (left) and field lines (right) predicted on a longitudinal section of the CMS detector at a central magnetic field of 3.8 T. Figure from Ref. [68].

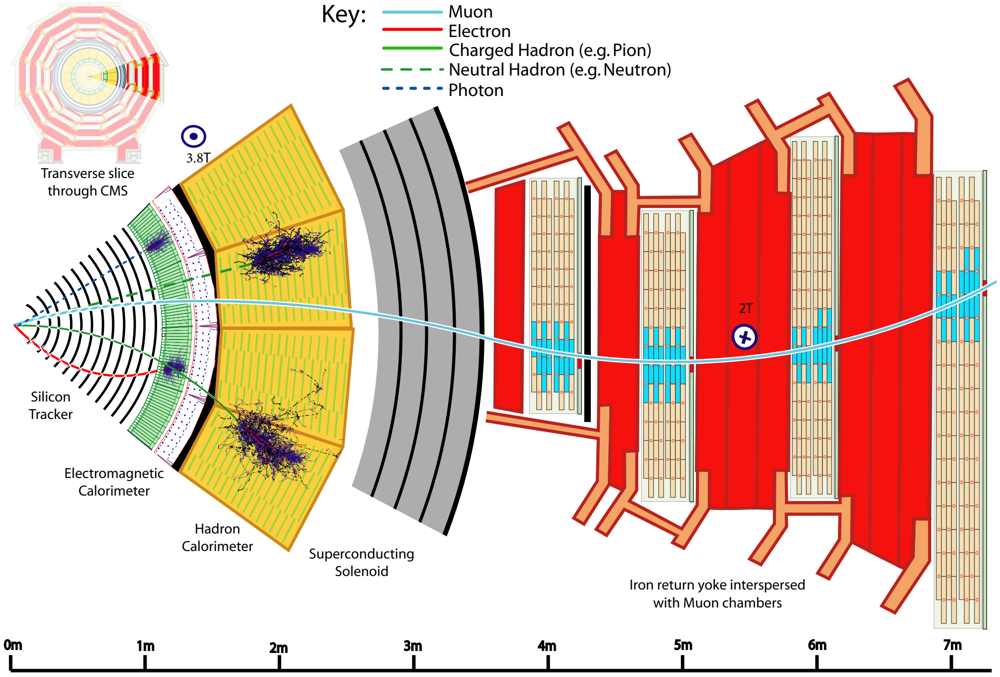


Figure 3.13: View of a transverse slice of the CMS detector with its sub-systems. A schematic view of the specific particle interactions, from the beam interaction region to the muon system, is also shown. The muon and the charged pion are positively charged, and the electron is negatively charged. Figure from Ref. [76].

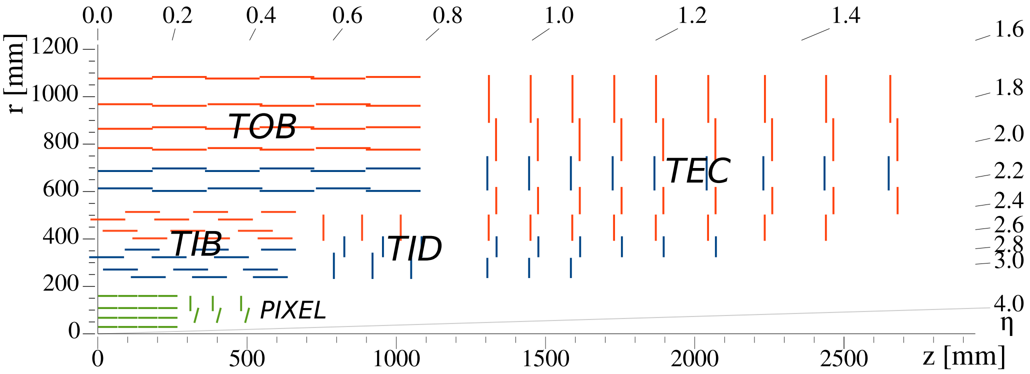


Figure 3.14: Sketch of one quarter of the CMS Phase-1 tracking system, in r-z view, showing the tracker pixel (PIXEL), inner barrel (TIB), inner disk (TID), outer barrel (TOB), and endcap (TEC) subsystems. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively. Figure from [78].

$r = 4$ cm, 60 kHz/mm^2 at $r = 22$ cm and 3 kHz/mm^2 at $r = 115$ cm. For a good efficiency in the track reconstruction, a low hit occupancy ($< 1\%$), defined as the fraction of sensors hit at each bunch crossing, is required. For this reasons, the innermost subdetector consists of high-granularity pixel detectors (up to 16 cm in radius), while microstrip detectors can be used in the outer region which has reduced flux. In total, the tracking system covers a pseudorapidity region of up to $|\eta| < 2.5$, has a length of 5.4 m, a diameter of ≈2.4 m, and an active silicon area of approximately 200 m^2, making it the largest silicon tracker ever constructed$^{\text{XI}}$. All tracking components are cooled to a temperature of $-10\,^\circ$C to improve radiation hardness. The overall used material, including support structures, cables, and the 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$^{\text{XII}}$ [77]. A schematic view of the Phase-1 tracker is shown in Fig. 3.14.

PIXEL TRACKER The layout of the inner pixel detector covers a pseudorapidity region of $|\eta| < 2.5$ in the innermost region of the tracker, 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, amounting to a total net active area of the pixel tracker of 1.9 m^2. The high granularity allows one to resolve large amounts of particles

---

[XI] At the time of writing

[XII] The radiation length X_0 is defined as the constant of the exponentially decaying energy distribution of an interacting high-energy particle $\mathrm{d}E/\mathrm{d}x = -E/X_0$.

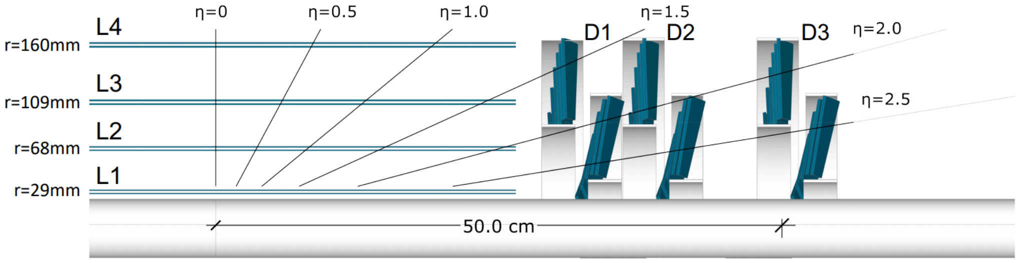


Figure 3.15: Layout of one quadrant of the CMS Phase-1 pixel detector. The four barrel layers are indicated with L1, L2, L3 and L4, whereas the three endcap disks with D1, D2, and D3. Figure from [72].

under high pileup conditions and to precisely reconstruct tracks and vertices in three dimensions. The spatial resolution of the inner tracker in the r-ϕ plane and z direction is 10 μm and 20 μm, respectively [72]. Within each layer, the modules overlap to minimize the holes in the acceptance. The inner pixel detector is shown in Fig. 3.15.

STRIP TRACKER The outer layers of the tracking system consists of silicon microstrips organized into four different subsystems and covers the radial region between 20 and 116 cm where 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 is composed by four layers of 320 μm thick microstrips sensors, covering radii up to 50 cm. Sensors in the TIB 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 (TIDs), consisting of three disks at each endcap, installed from 80 to 90 cm in the z direction. The 320 μm thick microstrip sensors in the TIDs are 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 $|\eta| < 2.5$.

Surrounding the inner tracking system[XIII] are the tracker outer barrel (TOB) in the central region and the tracker endcaps (TECs) 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 $|\eta| < 2.5$.

Tracking performance is discussed in Section 4.2.1, where track reconstruction and vertexing algorithms are also described.

[XIII] The pixel tracker, the TIB and the TID.

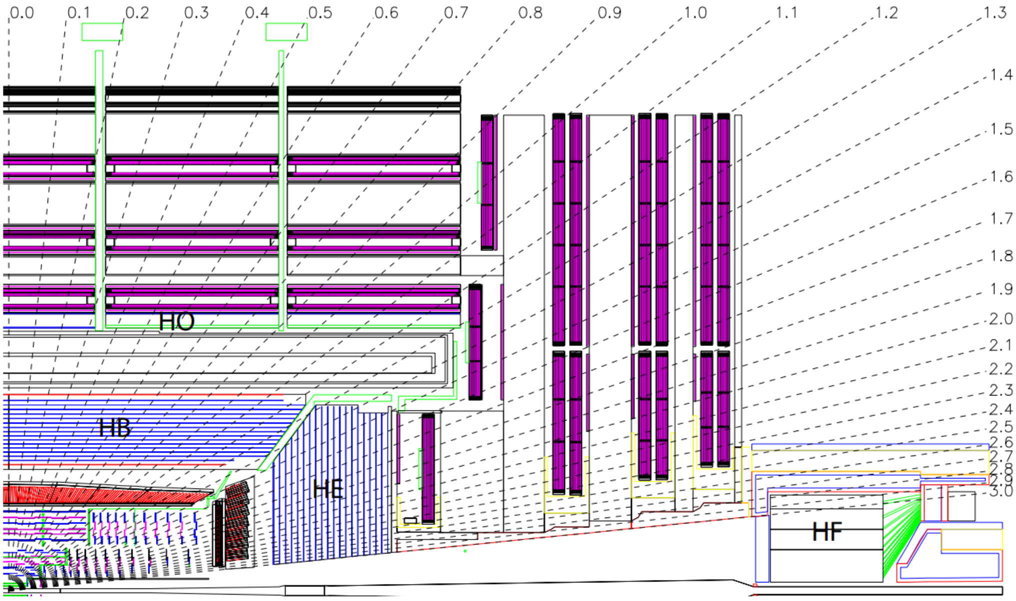


Figure 3.16: Cross section of one quadrant of the CMS calorimetry system showing the locations of the HCAL barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. The radial numbers indicate pseudorapidity values. Figure from Ref. [65].

3.2.3 *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 subdetectors with complementary objectives and 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. The calorimetry system is fully hermetic in the pseudorapidity region $|\eta| < 5$, to provide a reliable measurement of the missing transverse energy associated with neutrinos. An overview of its geometric layout is shown in Fig. 3.16.

3.2.3.1 *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) [73]. 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 distance of ± 314 cm from the interaction point. In addition, pre-shower detectors (ES) are installed in front of the endcaps in the region $1.653 < |\eta| < 2.6$ to deal with their reduced granularity.

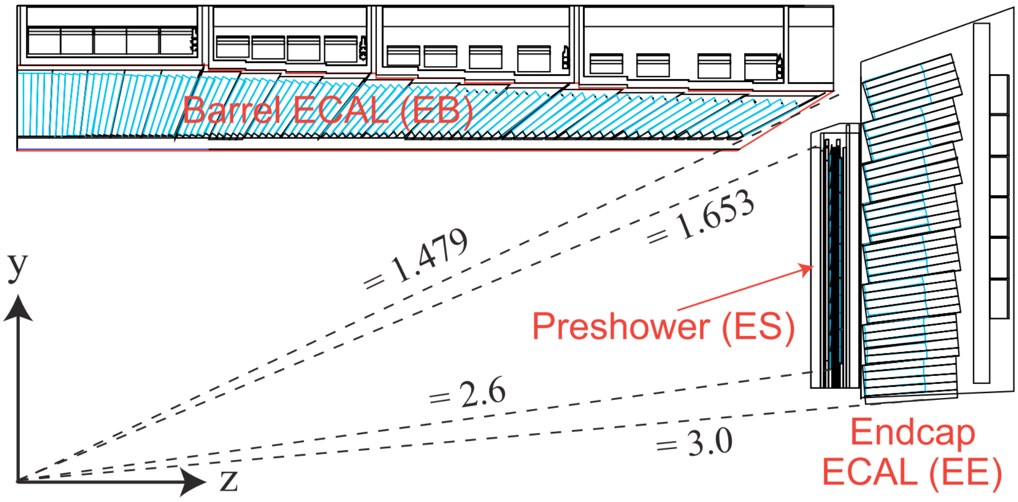


Figure 3.17: Transverse section through the ECAL, showing the geometrical configuration of the ECAL barrel (EB), endcap (EE), and preshower (ES) detectors in terms of the pseudorapidity. Figure from Ref. [79].

They consist of a 20 cm thick sampling calorimeter. In a sampling calorimeter an absorber material and an active material are interleaved. The absorber material induces a hadronic shower and the active material samples the shower along its length. In this case, the ES detectors are composed of alternating layers of lead absorber planes and silicon microstrip detectors. An overview of the geometrical configuration of the ECAL is shown in Fig. 3.17.

The energy measurement in the ECAL is performed by means of scintillation from the lead tungstate crystals. Lead tungstate is characterized by a high density ($\rho = 8.28\ \mathrm{g/cm^3}$), a short radiation length ($X_0 = 0.89$ cm) and a small Molière radius[XIV] ($R_M = 2.19$ cm). Moreover, the scintillation decay time is very short, such that 80% of the light is collected during a single 25 ns bunch spacing. The crystals in the EB have a front cross section of $2.2 \times 2.2\ \mathrm{cm^2}$ and a length of 23 cm, corresponding to ≈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 $2.86 \times 2.86\ \mathrm{cm^2}$ 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 (APDs) in the EB and vacuum phototriodes in the EE. On average, photoelectrons are produced at a rate of ≈4.5 per MeV. To stabilize the number of scintillation photons emitted by the crystals and the amplification of the APDs, which are both temperature-dependent, a cooling system is installed to keep the crystal temperature stable at $18.00 \pm 0.05\,^\circ\mathrm{C}$.

[XIV] 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 that contains on average 90% of the shower's energy deposition. Two Molière radii contain 95% of the shower's energy deposition.

The main purpose of the ECAL is, together with the tracking detector, to identify electrons and photons and to facilitate high-resolution energy measurements. Its energy resolution has been measured in an electron test beam and is found to be [65]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%\,, \tag{3.8}$$

where E is in GeV and $\oplus$ is the squared sum operator. The three terms in Eq. 3.8, each with different energy scaling, refer to the stochastic, noise, and irreducible components of resolution, respectively.

3.2.3.2 *Hadronic calorimeter*

The HCAL is a hermetic sampling calorimeter designed to measure the energy of charged and neutral hadrons, allowing a basic division of the particle types and the measurement of the missing energy associated with neutrinos [74]. The HCAL consists of four different subdetector systems: the HCAL barrel (HB), endcaps (HE), forward (HF), and outer (HO). The HCAL geometrical configuration has been already shown in Fig. 3.16.

The HB and HE consist of alternating layers of non-magnetic brass absorber[XV] and plastic scintillators. Traversing particles create showers in the brass layers, which induce detectable light in the subsequent scintillators. The scintillation light is then driven by wavelength-shifting fibers to a 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 into 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 structure similar to that of 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 hermeticity of the HCAL system in order to improve the measurement of the missing transverse energy. Due to the high radiation environment in this region (about 100 Mrad/year), the HF is made of robust quartz fibers as active material, steel absorber plates and is read 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 the absorbing material. The total thickness of the ECAL+HCAL calorimeter system is approximately 12 and 10 interaction lengths in the barrel and endcaps, respectively.

The energy resolution of the HCAL has been measured to be [80]:

$$\frac{\sigma_E}{E} = \frac{110\%}{\sqrt{E}} \oplus 9\%\,, \tag{3.9}$$

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

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

3.2.4 *Muon system*

In CMS, the muon system constitutes the outermost component of the apparatus and is designed to reconstruct the momenta and charge of muons over a wide angular range up to $|\eta| < 2.4$ [75]. In the ideal case, only muons[XVI] can survive the calorimetry system and reach the muon detectors, due to their long lifetime ($\tau_\mu \approx 2.2\ \mu$s) and scarce interaction with matter[XVII], but as described in Section 7.1.1 other particles can reach the detectors and be misidentified. The muon system is interleaved with the magnet return yokes and is composed of three different types of gaseous detectors depending on the region. An overview of its geometrical configuration is shown in Fig. 3.18.

The first type of muon detectors are the *drift tubes* (DT). 250 *drift tubes* are positioned in the barrel region of the detector ($|\eta| < 1.2$). They are organized into four layers, called "stations", and 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 $r-z$ coordinates while the other two r-ϕ. The exception is the outermost station, which does not contain the z-measuring planes. The single-point resolution of the DT system is measured as 80–120 μm [81].

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 not homogeneous, *cathode strip chambers* (CSC) are chosen for their radiation hardness and fast response. CSCs are multi-wire proportional chambers, which use gas under strong electric fields both as active and amplification medium. Four CSC stations are stationed at each endcap, for a total of 540 chambers. CSCs are positioned perpendicular to the beam axis and overlap in ϕ to avoid gaps in muon acceptance. Each CSC consists of six layers, all capable of measuring all three spatial coordinates (r-ϕ-z). The spatial resolution of a hit in the CSC system is 40-150 μm [81].

The last part of the muon system is composed by *resistive plate chambers* (RPC) that cover the pseudorapidity range $0.0 < |\eta| < 1.9$. RPCs consist of two parallel gas chambers operated in avalanche mode and can achieve a time resolution of a few nanoseconds, which allows the muon system readout to be assigned to a bunch crossing with rates up to 1 kHz/cm^2. At the same time, RPCs have a coarser spatial resolution of ≈1 cm with respect to DTs and CSCs.

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 [81]. The momentum resolution for muons up to 100 GeV is

---

[XVI] And neutrinos, which are however not detectable.

[XVII] 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.

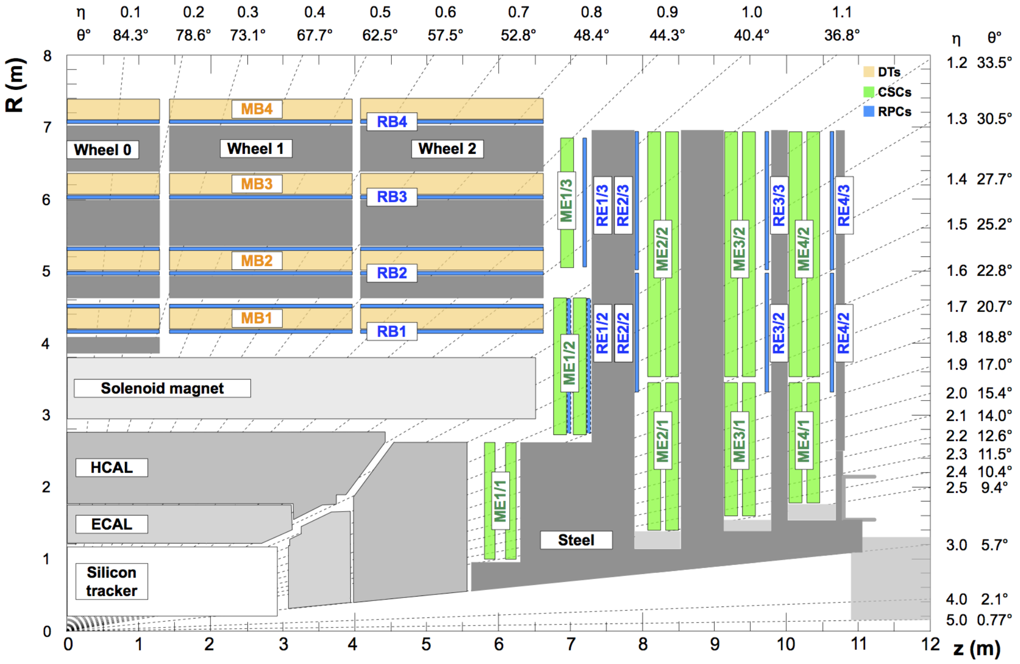


Figure 3.18: Longitudinal *r*-*z* cross section of one quadrant of the CMS detector with the muon system highlighted. The DT and CSC stations are labeled MB ("Muon Barrel") and ME ("Muon Endcap"), respectively. RPC stations are mounted in both the barrel and endcaps of CMS, and labeled RB and RE, respectively. Figure from Ref. [81].

≈1% in the barrel and ≈3% in the endcaps [81]. The reconstruction of muons is described in Section 4.2.2.

4

TRIGGER AND OBJECT RECONSTRUCTION

Protons in the LHC beam are grouped into small groups called bunches. At the interaction points, collisions between bunches occur at a frequency of 40 MHz, and each bunch crossing creates about 30 proton interactions Fig. 3.6. The rate of processes interesting for physics analysis is, however, several orders of magnitude lower. If CMS were to store all the information on all collisions (commonly called an "event"), it would far exceed the limits for processing and storing the data[I].

The CMS data acquisition (DAQ) and trigger systems are designed to select only a small fraction of events for storage and further analysis [83–87], reducing the acquisition rate by a factor of $\approx 10^6$.

A selected event is initially saved on permanent storage as a sum of all readout channels of the CMS detectors[II], the so called RAW format. For physics analyses, raw information need to be combined into "physics objects", in the procedure known as reconstruction.

This chapter will provide a brief description of the CMS trigger system and the procedures used in the reconstruction of physics objects. A short summary of the Monte Carlo (MC) techniques used to generate the simulated samples used in this work is also presented.

4.1 TRIGGER AND DATA ACQUISITION

The trigger system in CMS is composed of two stages: the "Level-1" (L1) trigger operates on fast-response data from the calorimeters and muon detectors, while the "high-level" trigger (HLT) uses a full, albeit rough, reconstruction of the whole event.

The L1 trigger is used to reduce the rate down to 100 kHz, leading to a maximum processing time of 4 μs, in which it must decide whether an event should be tentatively accepted (and passed to the further HLT stage) or rejected. Given the tight time constraint, only calorimeter and muon detectors are used in the L1 trigger. An overview of the logic of the CMS L1 trigger system is shown in Fig. 4.1. Data from the calorimeters and muon detectors are first processed regionally and then passed to the global trigger, which makes the final trigger decision. Accepted events are passed on to the HLT, via the Data Acquisition System (DAQ).

[I] Each event in CMS, taking into account all subdetectors has size of about 1 MB [82], leading to an output rate of ~40 TB/s. For comparison, the maximum sustainable event rate during Run2 was ~ 1 kHz.

[II] Since 2017, ≈135 millions.

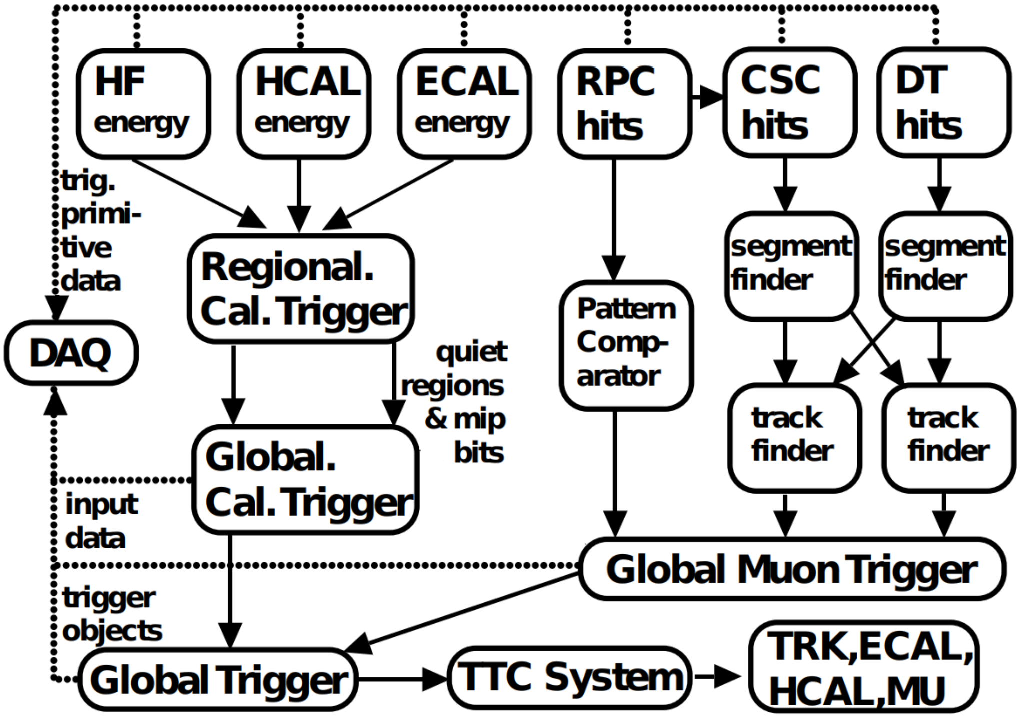


Figure 4.1: Overview of the CMS L1 trigger system, to be read from top to bottom. Data from the HCAL and ECAL calorimeters are processed first regionally and then globally. Hits from the RPC, CSC, and DT chambers are processed either via a pattern comparator or via a system of segment- and track-finders 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. [88].

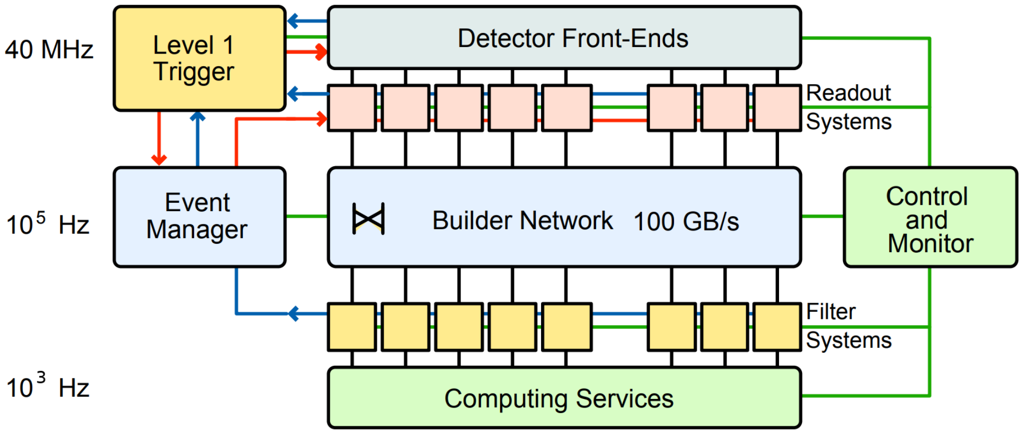


Figure 4.2: Schematic overview of the CMS DAQ and trigger systems. The trigger (left), data acquisition (center), and detector control systems (right) work in parallel to select and collect events. Figure from Ref. [65].

While the L1 Trigger consists of custom-designed electronics, the HLT is implemented in software and executed on a farm consisting of around 22 000 CPU cores. The L1 rate (100 kHz) constrains the maximum processing time for an event, considering the CPU power available, to about 220 ms.

The event selection at the HLT is performed in a similar way to that used in the offline processing. For each event, objects such as tracks, electrons, muons, and jets are reconstructed, and identification criteria are applied in order to select only those events which are of possible interest for data analysis. The processing of data in the HLT is based around the concept of "HLT paths", a sequence of algorithmic steps of increasing complexity and computational cost, which can either filter the event based on the available information or reconstruct additional physics objects for further processing. An event is accepted for storage and further analysis if it is accepted by at least one path. The processing benefits from the sharing of objects, as an object only needs to be reconstructed once for use in several paths. The HLT paths used as selection in the development of the tagger described in Part III, as well as the analysis described in Part IV, are discussed in Section 5.1.

The CMS DAQ system is closely integrated with the trigger system and is responsible for collecting and processing data from all subsystems [86, 87]. It is designed to readout a fraction of subdetectors at 40 MHz (for the L1 trigger) and the full CMS output at 100 kHz (for HLT and storage). A schematic representation of the CMS DAQ system is presented in Fig. 4.2.

4.2 RECONSTRUCTION OF PHYSICS OBJECTS

Reconstruction is the name of the process in which the raw information from the CMS subdetectors is combined into higher-level objects representing hypotheses of stable particles, assigning momentum, energy, direction, and estimating the particle type. Many algorithms are in use in CMS, depending on the type of object reconstructed, and all the information is finally correlated using the “Particle Flow” (PF) reconstruction algorithm [76]. The following sections briefly describe the reconstruction of ”low-level” objects, such as tracks, then the PF algorithm, and finally the reconstruction of higher-level objects such as decay vertices, electrons and jets.

4.2.1 *Track reconstruction*

Track objects represent the trajectory of charged particles within the tracking detectors. Tracks in CMS are reconstructed from hits using the so-called “Combinatorial Track Finder” [89], an adaptation of the combinatorial Kalman filter [90–92]. Track finding is executed in multiple passes called ”iterations” in a process called *iterative tracking*. Initial iterations search for tracks that are the easiest to find (e.g. with large p_T and produced near the interaction point). After each iteration, the hits associated with reconstructed tracks are removed from the set of all hits, simplifying the job of subsequent iterations. Later iterations reconstruct tracks with more complex trajectories (e.g. low p_T or produced far from the interaction point). Some iterations use feedback from the reconstruction of other physics objects, such as for example reexamining regions where the calorimeters detected large energy deposits (which may represent a jet, described in Section 4.2.5) or where a muon is expected to pass extrapolating its trajectory from muon detectors. Finally, the sets of tracks from all iterations are joined. During this stage, different tracks are compared to each other and merged if they are compatible with originating from the same particle.

Track reconstruction starts from ”seeds”, small groups of 2-4 hits in the tracking detector. Seeds for early iterations use hits from the inner detectors (e.g., the pixel tracker) since they provide a higher 3D resolution. Not all tracks can be reconstructed this way, as the particle may be created outside the inner region, so later iterations also (or exclusively) use the outer tracker detectors. The parameters of the seed trajectory are fitted (as an helix, because of the magnetic field) and then extrapolated to the nearest tracker layer. If a compatible hit is found in the layer, the parameters and covariances are updated according to the Kalman Filter rules; otherwise, the extrapolation continues marking the hit as missing. When the last tracker layer is reached, or when too many hits (1-2) are missing, the propagation is stopped. The track candidates are fitted using a Kalman Filter to obtain the final estimate of the track parameters and their uncertainties. Finally, the tracks are classified on the basis of quality requirements. Particularly relevant for this thesis is the *High Purity* flag, which is assigned to tracks passing specific quality requirements on χ^2, number of hits, missing hits, and, for iterations reconstructing non-displaced tracks, compatibility with production near the interaction point[III].

An overview of the different iterations, their input seeds and their physics target can be found in Tab. 4.1. Plots of the tracking efficiency in various iterations, as a function of p_T and displacement, can be found in Fig. 4.3.

Step name	Input seeds	Target tracks
Initial	pixel quadruplets	prompt, high p_T
LowPtQuad	pixel quadruplets	prompt, low p_T
HighPtTriplet	pixel triplets	prompt, high p_T recovery
LowPtTriplet	pixel triplets	prompt, low p_T
DetachedQuad	pixel quadruplets	from B hadron decay, $r \leq 5$ cm
DetachedTriplet	pixel triplets	from B hadron decay, $r \leq 5$ cm
MixedTriplet	pixel+strip triplets	displaced, $r \leq 7$ cm
PixelLess	inner strip pairs	displaced, $r \leq 25$ cm
TobTec	outer strip pairs	displaced, $r \leq 60$ cm
JetCore	pixel pairs in jets	high-p_T jets
Muon inside-out	muon-tagged tracks	muons
Muon outside-in	standalone muons	muons

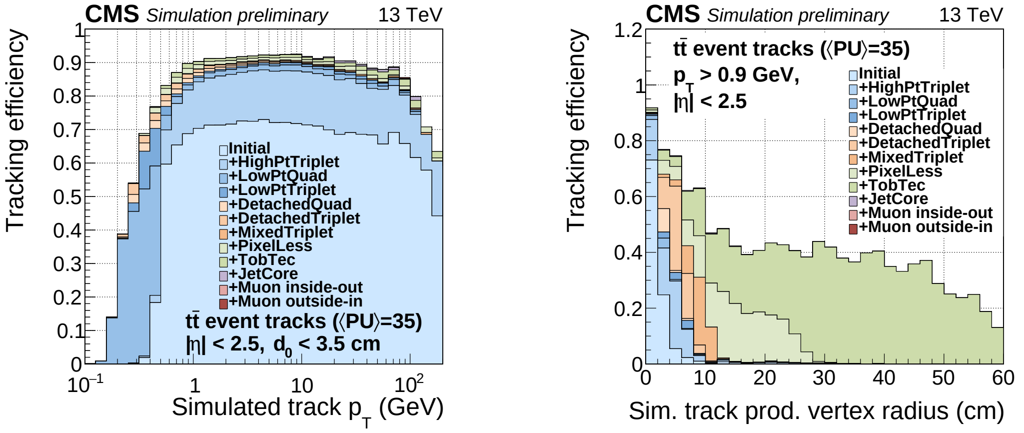
Table 4.1: List of the tracking iterations used in CMS since 2017, with the corresponding seeding configuration and the target tracks. From Ref. [94].

INTERACTION VERTEX RECONSTRUCTION Interaction vertices, also called primary vertices (PV), are the CMS representation of the pp interactions, including the pileup ones. PV reconstruction is performed entirely in the tracking system, and is composed of three main steps: track selection, clustering of tracks that appear to be produced at the same vertex, and fit of the vertex position using its associated tracks.

Tracks are selected based on the probability of having been produced near the interaction region, using cuts on the impact parameter of the track, the number of hits in the innermost detectors, and the track χ^2. To maximize efficiency, no requirement is imposed on the p_T of the tracks.

Clustering is based on the z-coordinate at the point of closest approach to the interaction region, and is implemented through Deterministic Annealing [95]. In Deterministic Annealing, the z-position of the tracks and the (unknown) position of the vertices are used to build a "free energy" function F (similar to the free energy of statistical mechanic), with a temperature coefficient T. The function is structured such as that when $T \to \infty$ all tracks are compatible with the existence of a single vertex, while when $T \to 0$ each track is hard assigned to a single vertex.

[III] A more precise description of the selections applied to High Purity tracks can be found in Ref. [93].



(a) Tracking efficiency as a function of p_T.

(b) Tracking efficiency as a function of displacement.

Figure 4.3: Tracking efficiency in CMS divided by tracking iteration. From Ref. [94].

In each iteration of the algorithm, the free energy is minimized as a function of the vertex positions and track assignment weights, then the temperature is reduced. In the initial state, only one vertex is considered, and new vertices are added whenever the decrease in temperature turns the minimum of F into a saddle point[IV]. The process is continued down to a temperature threshold, which is chosen to balance the resolving power and the possibility of incorrectly splitting true vertices.

Finally, once the tracks have been clustered into vertex candidates, all vertices with more than two tracks are fitted using the *adaptative vertex fitter* [96], to compute an estimate of the vertex parameters, including the 3D position and its uncertainty.

A plot showing the performance of vertex reconstruction in simulated data, with different numbers of simulated vertices, can be found in Fig. 4.4.

CMS also computes an estimation of the position of the interaction point, called the *beam spot* (BS). The beam spot represents the profile of the 3D region in which the beams collide. The position of the beam spot center can be determined in two ways. The first method is to use the position of the PVs defined above, whose position follow a probability distribution around the center. The second is to use a χ^2 fit based on the tracks transversal impact parameter. The two methods provide consistent results and are thus used in combination. Unlike PVs, which are computed in each event, the BS is averaged over a small period of time ($\mathcal{O}(10\text{ s})$), called a luminosity section.

[IV] This allows the algorithm to find the number of vertices automatically, and not just their position.

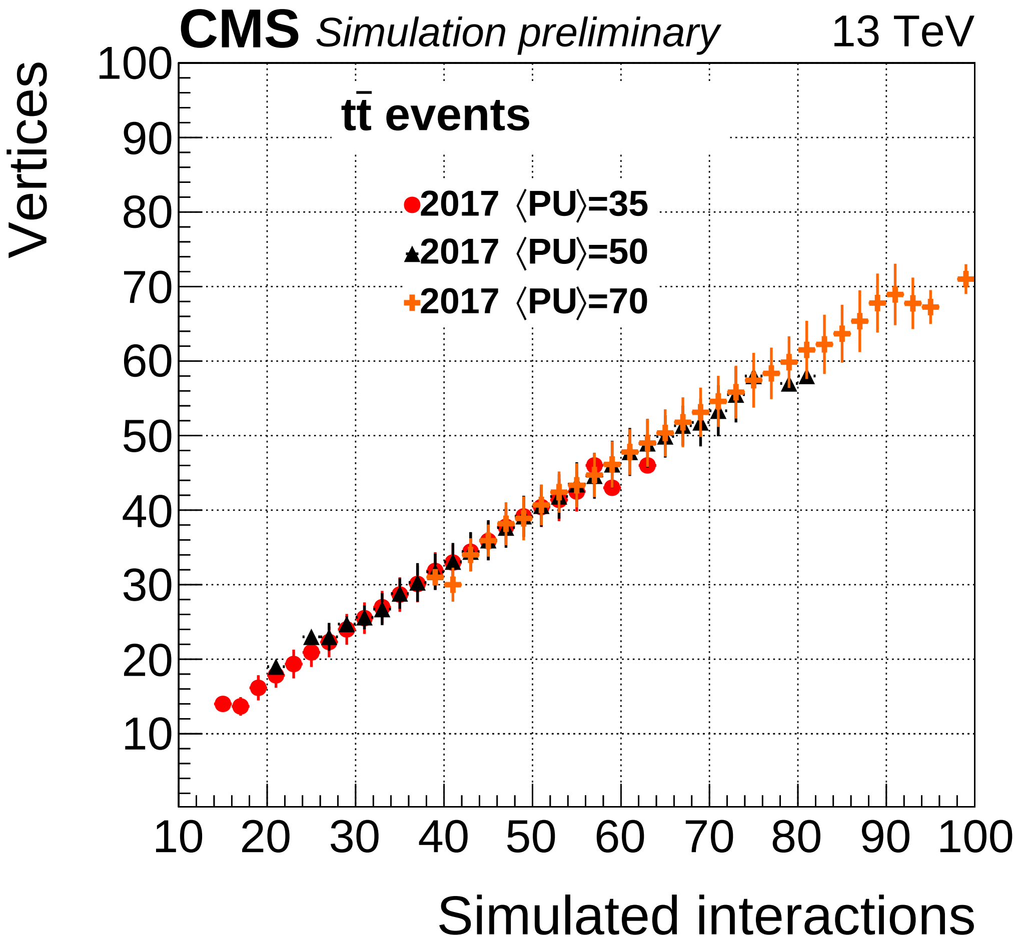


Figure 4.4: Performance of the CMS vertex reconstruction algorithm in simulated data, as a function of the amount of simulated interactions. The three colors refer to three samples with different averages for the number of simulated interactions.

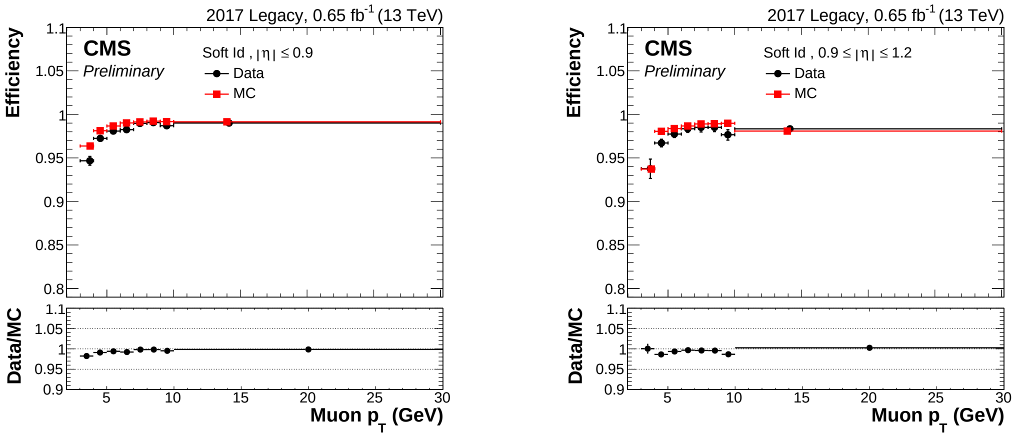
4.2.2 *Muon reconstruction*

Muons can be reconstructed using information from both the tracking system and the muon spectrometer. The following three different types of muon can be reconstructed depending on which subdetector information is used:

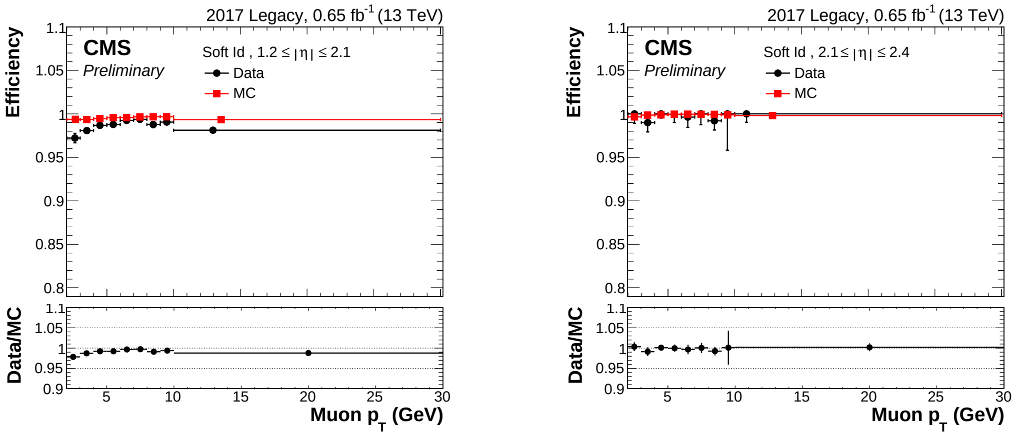
- *Standalone muons* are reconstructed by clustering hits in each DT or CSC to form track segments, which are then used as seeds for pattern recognition (based on a Kalman filter) using all DT, CSC, and RPC hits along the trajectory of candidate muons. This type of muon typically has high purity, but also low precision in their momentum measurement. Since the tracking system is very efficient, standalone muons originating from collisions are very rare ($\approx$1%) and most of them are cosmic ray muons traveling through the muon system.
- *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. This algorithm is especially useful for the identification of low p_T muons, which may not leave a sufficient number of hits inside 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, whereas the muon segment provides only information on the particle type.
- *Global muons* are build by extrapolating standalone-muon tracks inwards to the tracking system[V]. A link to an inner track is established if it yields high compatibility, as determined by a Kalman filter approach. The muon track is then refitted by combining information from both subsystems, thus improving the measurement of the muon momentum. This type of muon has high purity, while its reconstruction efficiency is highly dependent 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. To reduce the background from hadronic fakes, an MVA-based selection is provided, which will be described in Section 7.1.1. Overall, about 99% of all muons produced within the geometric acceptance of the muon system are reconstructed. A more detailed view of the efficiency of muons at the energies considered in this thesis can be found in Fig. 4.5.

[V] This approach is called *outside-in* and is opposite to the one used to reconstruct tracker muons (*inside-out*).



(a) $|\eta| < 0.9$ (b) $0.9 < |\eta| < 1.2$.



(c) $1.2 < |\eta| < 2.1$ (d) $2.1 < |\eta| < 2.4$.

Figure 4.5: Muon efficiency in CMS in slices of $|\eta|$. These effiency were estimated using the Tag and Probe method on J/ψ resonances. From Ref. [97].

4.2.3 *The Particle Flow algorithm*

The purpose of the PF algorithm is to combine measurements from all CMS subdetectors in a global reconstruction of all physics objects in the event, as well as a tentative identification of the type of particle that produced the signal. The PF algorithm can be structured into three consecutive steps. First, the information from the subsystems is collected to form the so-called "PF objects". These include tracks of charged particles in the tracking system and in muon detectors and clusters of energy deposited in the calorimeters. Then, the PF objects are spatially correlated with a *link algorithm*. Finally, particle hypotheses are inferred, and complex objects, such as vertices, jets, and missing energy, are computed.

LINK ALGORITHM The link algorithm is the core of the PF reconstruction and connects tracks (both in the tracker and muon detectors) and clusters in the calorimeters, 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 the computation time.

The tracks reconstructed on 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 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.

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

1. *Muons* are selected among those reconstructed in Section 4.2.2, with additional cuts based on isolation (surrounding tracks and surrounding energy deposits in the calorimeter) and tight cuts on the muon MVA.
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. Geometric 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[VI].

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 already been shown in Fig. 3.13 in Chapter 3.

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

4.2.4 *Electron reconstruction*

Electrons in CMS are reconstructed using information from both the tracker and ECAL, using the PF algorithm to join the two sources [99].

Most of the energy is deposited in the calorimeter, but while passing through the tracker the electron may interact with the material emitting bremsstrahlung photons, which can themselves be converted into an electron-positron pair. Due to this, a dedicated algorithm needs to be used to combine the clusters from the individual particles into a single object and recover the energy of the primary electron or photon. Additionally, the trajectory of an electron main track has a variable curvature due to the loss of momentum while emitting bremsstrahlung photons. For this reason, electron tracks are not reconstructed using the normal Kalman Filter procedure from Section 4.2.1, and instead use a custom algorithm based on the Gaussian Sum Filter (GSF) for the estimation of track parameters.

Electron and photon reconstruction is done simultaneously, as the two type of objects leave an extremely similar signal (due to bremsstrahlung and photon conversion), aside from the lack of originating GSF track. The first step for the reconstruction is the identification of clusters of crystals with sufficient energy in the ECAL, which are then merged (geometrically) into superclusters (SC). Then, the GSF algorithm is applied, using as seeds compatible hits in the pixel detector and compatible generic tracks (from Section 4.2.1). Tracks from photon conversion into e^+- e^- pairs are also searched among the generic tracks. Clusters, GSF tracks, and conversion tracks are imported in the PF algorithm, which links them together into photon and electron candidates. Finally, electrons are defined as the candidates from PF that contain a GSF track.

To limit the background from fake electrons (misidentified jets and photon conversions), an MVA selector (hereafter named "`eleIDNIV2Val`") is provided, based on the energy radiated by the GSF track, the distance between the expected

---

[VI] Some species of charged hadrons are identifiable by exploiting the energy loss $\mathrm{d}E/\mathrm{d}x$ of particles traversing the inner tracker [98]. 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.

(from the track trajectory) and measured cluster position, the ratio of energy deposited in HCAL and ECAL and the quality of the GSF track fit. By default, no selection is applied to maximize efficiency, and during analysis the cut must be optimized based on the analysis requirements[VII].

4.2.5 *Jet reconstruction*

Jets are narrow cones of collimated hadrons produced in the hadronization of quarks or gluons.

In this thesis, all jets considered were reconstructed starting from PF objects, using the sequential recombination algorithm anti-k_T [VIII]. Sequential recombination algorithms are the main type of algorithm used in LHC experiments. They reconstruct jets using a bottom-up approach, starting to combine the closest particles based on a distance metric defined as

$$d_{ij} = \min\left(p_{T,i}^{2p}, p_{T,j}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2} \quad \text{and} \quad d_{iB} = p_{T,i}^{2p}\,, \tag{4.1}$$

where d_{ij} is the distance between particles i and j, d_{iB} is the distance between particle i and the beam, ΔR_{ij} is the euclidean distance in the η-ϕ plane between particles i and j and R is a configurable parameter called the jet radius[IX]. p is another configurable parameter which controls the weight of the different particles in the clustering: when p is positive (+1 in the k_T algorithm), jet clustering starts around soft components, while when p is negative (-1 in the anti-k_T algorithm) it starts around high p_T components. After all distances are computed, the next step is to compare the smallest d_{ij} and the smallest d_{iB}. Then:

- if $d_{ij} < d_{iB}$, merge particles i and j into a single particle, remove i and j from the list of available particles, and proceed to the next iteration;
- if $d_{ij} > d_{iB}$ call particle i a jet, remove it from the list of particles, then proceed to the next iteration.

The iterations stop when all particles have been clustered.

The jets in this dissertation are reconstructed using the anti-k_T algorithm [100]. An advantage when using anti-k_T is the behaviour of low and high-p_T particles in the clustering: soft components have little effect on the boundaries of the jet, while multiple hard components will either be merged, if they are close enough compared to the jet radius, or produce clipped cones, based on the difference in momentum, if they are not. Thanks to this property, anti-k_T jets are resistant

---

[VII] For this thesis electron objects are only relevant in the context of the electron-based flavor tagger, and the cuts applied are discussed in Chapter 8.

[VIII] Jets in CMS can also be reconstructed using only the calorimeters informations (CALO jets) or calorimeters plus tracks (Jet-Plus-Track, JPT, jets).

[IX] Chosen as 0.4 for jets used in this thesis.

both against emission of soft particles in the jet, which would make the jet area dependent on the jet momentum, and against the influence of overlayed soft events (e.g. pileup), which would smear the measurement of the jet p_T.

4.2.6 *Secondary vertex reconstruction*

Secondary vertices (SV) in CMS represent the decay points of particles with relatively long lifetimes (compared to the CMS position resolution), such as b-hadrons.

While the reconstruction of a decay vertex without information on the particles produced is possible, and used in CMS for tasks such as the classification of jets, all SVs considered in this dissertation were reconstructed will full knowledge of the decay product. This knowledge enables the use of the more precise Kinematic Fit algorithm. In kinematic fits, the fitter can add constraints based on the conservation of energy and momentum. Additionally, when the fit contains resonances, the kinematic fitter can also use their nominal mass (and decay width) as ground truths, which improves the mass resolution.

The search for secondary vertices starts from PF objects, to make use of the particle kind as estimated from CMS, but mostly uses information from the tracks associated to the objects (precise position, parameters uncertainty, behaviour in the CMS magnetic field). In case of decays with internal resonances (e.g. $B_s^0 \to J/\psi\,\phi(1020)$), vertex reconstruction is performed bottom-up. Initially, only non-composite decays are reconstructed (e.g. $J/\psi \to \mu^+\mu^-$ or $\phi(1020) \to \mathrm{KK}$). Those decays are then combined with other decays or PF objects to reconstruct the more complex vertices. At each stage, vertices are skimmed based on the mass (to be compared with the resonance mass hypothesis), vertex fit quality and momentum. This allows a considerable reduction in the amount of combinatorial background, as well as processing time.

The main decays used in this thesis are $B_s^0 \to J/\psi\,\phi(1020)$, $B^+ \to J/\psi\,K^+$ and $B^0 \to J/\psi\,K^*(892)^0$. Precise selections applied to the vertices are described in Section 5.2.1.

4.3 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 train the tagging algorithms presented in Part III and to parameterize the acceptance effects and background in the analysis presented in Part IV. Specific simulated samples used in this thesis are reported in Section 5.4.

In CMS, simulation of pp collision events is performed in multiple steps. The first step, called *generation*, is responsible for the simulation of the physical pro-

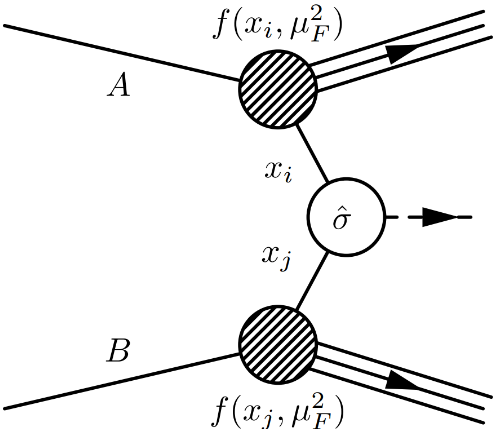


Figure 4.6: Schematic representation of the factorization in a hard-scattering process between two hadrons. Figure from Ref. [102].

cesses included in the event. The interaction of the two protons is seen 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 [101]. 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 $\sqrt{s}$, is computed as

$$\sigma(pp \to X \,|\, \sqrt{s}) = \sum_{i,j=q,\bar{q},g} \int_0^1 \hat{\sigma}\Big(ij \to X \,|\, \sqrt{\hat{s}}, \mu_R^2, \mu_F^2\Big) f(x_1, \mu_F^2) f(x_2, \mu_F^2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \ , \tag{4.2}$$

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[X], and μ_R is the renormalization scale that defines the scale for the strong interaction coupling constant α_s[XI]. A schematic representation of this factorization is shown in Fig. 4.6.

Parton distribution functions describe the probability for 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 simulation by Monte Carlo event generators that model each aspect of an event up to a given perturbation order [103]. Particle decays and other probabilistic processes are generated with fixed physics parameters, which typically correspond to the current world-average values, and with appropriately distributed random variables.

[X] 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 $\hat{\sigma}$.

[XI] Typically $\mu_F = \mu_R$ is chosen.

Above the QCD renormalization scale, i. e., when $q^2 > \mu_R^2$, the strong coupling constant is sufficiently small to allow perturbative QCD calculations. In these regimes, the parton showering approach is used to describe the evolution of QCD processes to a cutoff value by recursively adding $g \to q\bar{q}$ and $q \to qg$ splitting. At the renormalization scale, perturbation approaches cannot be used any longer, and phenomenological methods are implemented to proceed 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[XII], and pileup interaction. In CMS, the effect of pileup is accounted for by overlaying the hard-scattering process to simulated minimum bias events[XIII], where the pileup multiplicity per event is based on the distribution observed in real data.

A schematic representation of a proton-proton collision from the event-generator point of view is shown in Fig. 4.7.

The second step, called *simulation*, models the response of the detector to a given set of simulated final-state particles. First, the final-state particles are propagated in a full simulation of the CMS detector using the GEANT4 package [105], which models the propagation of particles through matter taking into account all types of interaction with matter, the production of secondary particles, and the influence of electromagnetic fields.

Next, the readout of all subsystems 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. Pileup interactions are overlaid at this point. Finally, the simulated samples are reconstructed using the same software as the collision data and can be analyzed with common analysis tools. Typically, the simulated samples contain also information on the generated 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 8.240 Monte Carlo event generator [106] with the underlying event tune CP5 [107] and the parton distribution function set NNPDF3.1 [108]. Final-state photon radiation is accounted for with PHOTOS 215.5 [109, 110]. The b-hadron decays are modeled with the EVTGEN 1.6.0 package [111].

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

[XIII] This process is executed overlaying the response of the detectors instead of the generated physical processes

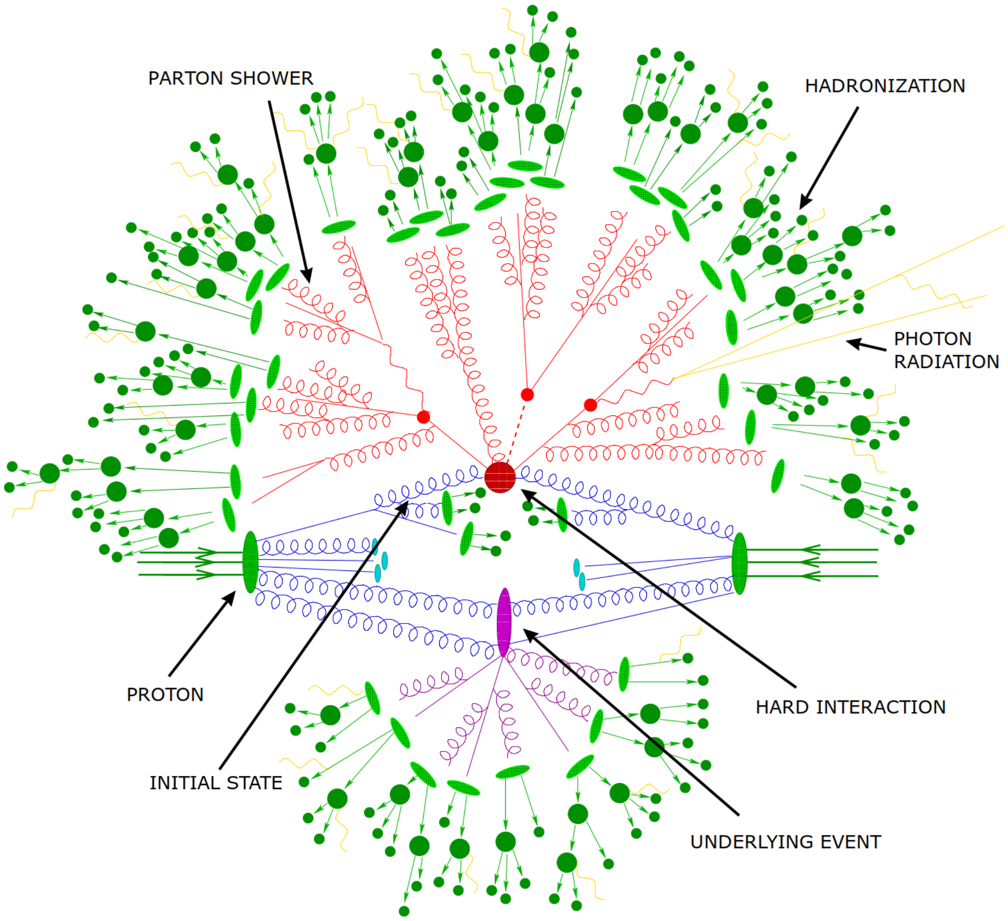


Figure 4.7: Sketch of a proton-proton collision from an event-generator point of view. Figure adapted from Ref. [104].

5

SAMPLES AND EVENT RECONSTRUCTION

This chapter describes the samples used both in the development of the tagging algorithms in Part III and in the analysis in Part IV. The tagging algorithms were designed with CP violation measurements in b-meson systems in mind, so in both cases the target is the identification and reconstruction of high quality b-mesons. In particular, this thesis makes use of $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$, $B^+ \to J/\psi\,K^+ \to \mu^+\mu^- K^+$ and $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ decays.

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:

$$S = \frac{N_{\text{sig}}}{\sqrt{N_{\text{sig}} + N_{\text{bkg}}}}\,, \tag{5.1}$$

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

The basic preprocessing step in any sample is the application of the techniques from Section 4.2.6 to extract candidate b-meson decays. To save processing time, a single pass is done on each sample, reconstructing the vertices for all considered decays. The pre-processed data samples are split according to the requested decay into the samples used in the analysis. Note that since the decays are reconstructed independently, a single event can be used simultaneously in any of the decay subsamples. This is, however, rare. From now on, the three decay-split analysis subsamples will be called *dataset_Bs*, *dataset_Bp* and *dataset_B0*. Of the three, only *dataset_Bs* is extensively used, while the other two samples are only used for specific tasks such as the calibration of the tagging algorithms (procedure described in Section 6.6) and the estimation of the efficiency (described in Section 13.2.6).

This chapter describes the procedure to go from the objects described in Chapter 4 to the samples used for tagger development and CP violation analysis, including a description of the selection optimization.

5.1 TRIGGER SELECTION

As explained in Section 4.1, only a fraction of pp collisions are selected by the CMS trigger system and recorded for further analysis. In this work, three high-level triggers are used, all optimized for the detection of b -hadrons that decay to J/ψ mesons. The aim of a trigger path is to reduce the activation rate as much as possible (since the trigger rate is a very limited resource that must be shared with all trigger paths) while maintaining high efficiency of triggering *good* events. The large difference between the triggers used here is how the rate is reduced.

The first two triggers considered are very similar and are called

$$\texttt{HLT\_DoubleMu4\_JpsiTrkTrk\_Displaced} \tag{5.2}$$

$$\texttt{HLT\_DoubleMu4\_JpsiTrk\_Displaced}, \tag{5.3}$$

hereafter `HLT_JpsiTrkTrk` and `HLT_JpsiTrk` respectively. Both triggers are meant to activate in case of b-hadrons decaying into a J/ψ plus additional tracks.

Most of the selections applied in these triggers, summarized in Tab. 5.1 and Tab. 5.2, are fairly standard. Note, however, that the HLT algorithms do not use L1 information, and as such, there is no direct relationship between the L1 muon objects and the HLT ones. Momenta cuts are a very common selection criterion in high energy particle physics, as they reduce background from soft processes and restrict the kinematic acceptance to the region of interest. At HLT level, vertices are available in the selection, so both triggers apply requirements on the $\mu\mu$ vertex to select decays into a J/ψ, with a mass range of 2.9–3.3 GeV, and a probability to originate from a common vertex larger than 10%[I]. To reduce the triggering rate, both paths require a large displacement, in particular a significance $L_{xy}/\sigma(L_{xy}) > 3$. This is common for B-Physics triggers, as b-hadron tend to have a relatively long decay time and this cut helps reducing the background from particles originating close to the interaction point (prompt). This is, however, a compromise, because while b-hadrons are less affected by the cuts compared to the backgrounds, this significantly impacts the trigger efficiency of the signal, producing a visible turn-on curve in the distribution of the proper decay time, and precludes the possibility of studies in the prompt region[II]. The cuts on the tracks is where `HLT_JpsiTrkTrk` and `HLT_JpsiTrk` differ the most. `HLT_JpsiTrkTrk` is intended to trigger *only* in the presence of $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-$ decays[III], so it can add cuts to the Trk-Trk vertex, specifically a cut to the invariant mass (0.95–1.3 GeV) and on the vertex probability ($> 80\%$). `HLT_JpsiTrk` instead is more generic in what it accepts, but that comes at the cost of a larger displacement requirement: the track is required to have a transverse impact parameter $d_0 > 2$ cm. Finally, in both triggers, the fit of the total J/ψ plus tracks vertex is required to have $\chi^2/\mathrm{NDoF} < 10$.

[I] 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.

[II] Resolution studies in Section 13.2.5 were, for example, performed without selections in these triggers, since they are executed using prompt decays.

[III] This trigger was developed specifically for the analysis described in Part IV.

Variable	Either			
Min. number of muons	2	2	2	2
$p_T(\mu)$	> 0 GeV	> 0 GeV	> 4.5 GeV	> 4 GeV
$\eta(\mu)$	< 1.5	< 1.4	–	–
$\Delta R(\mu_1 \mu_2)$	< 1.4	< 1.4	< 1.2	< 1.2

Table 5.1: L1 trigger selection requirements for `HLT_JpsiTrkTrk` and `HLT_JpsiTrk` triggers. The 4 L1 seeds are in OR from the point of view of the HLT path. While the triggers may seem to overlap, each can be activated at different times during the data taking, typically in response to the decreasing instantaneous luminosity after a fill.

Table 5.2: HLT selection requirements for `HLT_JpsiTrkTrk` and `HLT_JpsiTrk`. The J/ψ candidate is formed by a pair of oppositely charged muons μ^+ and μ^-. The charged tracks are indicated with $h^\pm$.

Observable	`HLT_JpsiTrkTrk`	`HLT_JpsiTrk`
$\lvert\eta(\mu^\pm)\rvert$	< 2.5	< 2.5
$p_T(\mu^\pm)$	> 4 GeV	> 4 GeV
$p_T(\mu^+\mu^-)$	> 6.9 GeV	> 6.9 GeV
$m(\mu^+\mu^-)$	$\in [2.9,\ 3.3]$ GeV	$\in [2.9,\ 3.3]$ GeV
J/ψ vertex prob.	$> 10\%$	$> 10\%$
$L_{xy}^{J/\psi} / \sigma_{L_{xy}^{J/\psi}}$	> 3	> 3
$p_T(h^\pm)$	> 0.8 GeV	> 1.2 GeV
$d_0(h^\pm)$	–	2 cm
$m(h^+h^-)\rvert_{m(h^\pm)=m(K^\pm)}$	$\in [0.95,\ 1.30]$ GeV	–

Table 5.3: L1 trigger selection requirements for the `HLT_JpsiMu` trigger. 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

The third trigger, called

$$\texttt{HLT\_Dimuon0\_Jpsi3p5\_Muon2,} \tag{5.4}$$

hereafter `HLT_JpsiMu`, uses a different strategy for the rate reduction: instead of maximizing the background rejection at the cost of signal efficiency, it places minimal cuts on the event, two muons forming a J/ψ vertex, but requires the presence in the event of a third muon, unrelated to the J/ψ. The third muon can have several uses (in CMS this path is, for example, used also for decays containing two J/ψ), but in the work presented in this thesis it has a very specific role, that of a tagging muon to be used for the OS-mu based flavor tagging (described in Chapter 7). Thanks to this cut, the rate of both signal and background is significantly reduced, but the signal events passing the selection have a very high probability (more than 50%) to be well tagged, and thus useful in a CP violation measurement. Due to the already small rate, no displacement cuts are thus necessary.

`HLT_JpsiMu` selections are summarized in Tab. 5.3 and Tab. 5.4. `HLT_JpsiMu` 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 $m(\mu_1\mu_2) < 9$ GeV. There is no p_T requirement on the additional muon at L1. At HLT level, the three muons are required to be within the geometric acceptance of the CMS muon system $|\eta| < 2.5$ 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%[IV]. The third muon is required to have $p_T > 2$ GeV.

Events that pass the trigger selection are stored for further analyses. As explained in Section 4.1, 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

---

[IV] All combinations of the three muons are tried.

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. To avoid biases due to the trigger selection, all three trigger objects need to be matched with the respective muon candidates used in the analysis. In practice, 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 $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.1$.

Trigger paths can be used to split each sample into two subsamples[V]. To preserve the good properties in the prompt region of the `HLT_JpsiMu` trigger, events tagged with that path are alone in one of the subsamples. To avoid double counting and overlaps, `HLT_JpsiMu` events are vetoed from the second subsample.

For simplicity, the names *dataset_JpsiMuon* and *dataset_JpsiTrkTrk* will refer from now on to the two subsamples of *dataset_Bs*.

5.2 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) [112], ROOT [113], TMVA [114], and RooFit [115]. CMSSW and ROOT form the general analysis framework, while TMVA and RooFit are used for multivariate analysis (MVA) and fitting, respectively.

5.2.1 *Offline selection 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 algorithm starts by assigning initial cut values set by the user, then cuts are changed randomly within a predefined n-dimensional matrix of allowed ranges, also set by the user. In this way, the *population* evolves to maximize the *fitness function*, which in our case is the classical definition of signal significance from Eq. 5.1, although more exotic choices can be adopted. The requirements on the following selection observables have been optimized:

- b-meson p_T, ct, and vertex probability;
- muons p_T and η;
- kaons p_T and η.

[V] `HLT_JpsiTrkTrk` and `HLT_JpsiTrk` are never used together in the same sample, as they are meant to trigger in different decays.

The allowed ranges have been chosen to be at least as tight as the respective HLT selection described in Section 5.1.

Optimization is performed only for $B_s^0 \to J/\psi\,\phi(1020)$ decays. The selection for $B^0 \to J/\psi\,K^*(892)^0$ decays follow the same rules as that of the B_s^0, except the $\phi(1020)$ selection is replaced with a $K^*(892)^0$ and the mass window is increased to 0.09 GeV to accomodate the wider resonance. In $B^+ \to J/\psi\,K^+$ decays instead the cuts on the $\phi(1020)$ are replaced with cuts of $p_T > 1.6$ GeV, #hits ≥ 4 on the kaon track.

The best configuration of selection cuts, along with the other requirements, is reported in Tab. 5.5, Tab. 5.6 and Tab. 5.7, where the cuts common between *dataset_JpsiMuon* and *dataset_JpsiTrkTrk* are presented in a separate table to ease reading. The requirement on $p_T(\mu^+\mu^-)$ for *dataset_JpsiTrkTrk* has been added to reproduce the HLT selection.

It is interesting to note that most of the optimal kinematic requirements on the final state particles are equal to the HLT selection[VI], indicating that the HLT cuts probably limit the statistical power of the analysis.

5.2.2 *Candidate selection*

The decay vertices of b-mesons used in this thesis are fitted using the method described in Section 4.2.6. Only one candidate per event is kept, and the selection is made based on the vertex fit probability, selecting the one having the greatest probability among the vertices passing the selections from Section 5.2.1

PRIMARY VERTEX SELECTION To compute the proper decay length, each selected event is also required to have a reconstructed Primary Vertex (PV). The PV closest in 3D to the line that passes through the SV and parallel to the B_s^0 momentum is selected (“*distance*” method). Several other procedures have been studied to identify the PV but have been found to lead to a higher uncertainty of the proper decay duration or to produce bias in the $ct \sim 0$ region. Two of them are worth commenting on:

- The PV chosen with the popular *pointing angle* method leads to bias in the low ct region as shown in Fig. 5.2. This is likely due to the fact that this selection method mathematically prefers positive ct values over negative ones (Fig. 5.2). While this is not a problem for the b-meson candidates used in this thesis, which are selected with a ct cut of $ct > 60\,\mu$m, it is problematic for studying the ct resolution using the prompt background. For this reason, the distance method was preferred.

- The center position of the beam spot was also considered as the point of origin to measure ct, but it was found to be nearly equivalent to using the

---

[VI] Lower boundary for the allowed range of the optimization.

Table 5.4: HLT selection requirements for `HLT_JpsiMu`. The J/ψ candidate is formed by a pair of oppositely charged muons μ^+ and μ^-. The muon not used to form the J/ψ candidate is indicated with μ_3.

Observable	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.	$> 0.5\%$

Table 5.5: Offline selection requirements common between *dataset_JpsiMuon* and *dataset_JpsiTrkTrk*.

Candidate	Variable	Requirement	Optimised
Event	PV found	Yes	
	Good data-taking cond.	Yes	
	Trigger obj. match	Yes	
$B_s^0 \to J/\psi\,\phi(1020)$	$m(\mu^+\mu^-K^+K^-)$	$\in [5.24,\ 5.49]$ GeV	
	$p_T(B_s^0)$	> 9.5 GeV	Yes
	B_s^0 vertex prob.	$> 2\%$	Yes
	$\sigma_{ct}(B_s^0)$	$< 50\ \mu$m	
$J/\psi \to \mu^+\mu^-$	Muon reco.	*Global*	
	Muon id.	*Loose*	
	$\lvert\eta(\mu)\rvert$	< 2.4	Yes
	$\left\lvert m(\mu^+\mu^-) - m_{J/\psi}^{\text{w.a.}}\right\rvert$	< 150 MeV	
$\phi(1020) \to K^+K^-$	Track quality	*High purity*	
	$\lvert\eta(K)\rvert$	< 2.5	Yes
	$\left\lvert m(K^+K^-) - m_{\phi(1020)}^{\text{w.a.}}\right\rvert$	< 10 MeV	
	N. of tracker hits	≥ 4	

Table 5.6: Offline selection requirements specific for *dataset_JpsiMuon*.

Variable	Requirement	Optimised
HLT	`HLT_JpsiMu`	
$p_T(\mu)$	> 3.5 GeV	Yes
$p_T(K)$	> 1.15 GeV	Yes
$ct(B_s^0)$	$> 60\ \mu$m	Yes

PV chosen with the distance method. The beam spot position is, however, not well simulated in the Monte Carlo samples used, and to avoid potential biases (since all development happens in MC samples) the distance method was again preferred.

A scheme showing the various methods for the selection of the PV are shown in Fig. 5.1.

PROPER DECAY LENGTH MEASUREMENT Once the PV has been chosen, the proper decay length is measured in the transverse plane as:

$$ct = c\, m^{\text{w.a.}}_{B_q} \cdot \frac{L_{xy} \cdot p_T}{|p_T|^2}\,, \tag{5.5}$$

where $m^{\text{w.a.}}_{B_q}$ is the world average value of the B_q mass [24] (with $q = [\mathrm{u}, \mathrm{d}, \mathrm{s}]$), p_T is here the transversal projection *vector* of the momentum, and L_{xy} is the reconstructed transverse decay length:

$$L_{xy} \equiv \bar{r}_{xy}(\text{SV}) - \bar{r}_{xy}(\text{PV})\,, \tag{5.6}$$

with $\bar{r}_{xy}(\text{SV})$ and $\bar{r}_{xy}(\text{PV})$ being the secondary and primary vertex positions in the transverse plane, respectively.

The proper decay length uncertainty (σ_{ct}) is obtained by fully propagating the uncertainties in the decay length and the momentum of the *b*-meson candidate[VII]. The L_{xy} contribution to σ^2_{ct} contains both the uncertainties in the positions of the PV and SV, and can be written as

$$\sigma^2_{ct\,|\,L_{xy}} = \left(\frac{c\, m^{\text{w.a.}}_{B_q}}{|p_T|^2}\right)^2 p_T \cdot \left(\sigma_{\text{pos}}(PV) + \sigma_{\text{pos}}(SV)\right) \cdot p_T\,, \tag{5.7}$$

where σ_{pos} indicates the 2D position covariance matrix on the vertices.

The p_T of the vertex is defined as the sum of the p_T^{trk} of the tracks, whose uncertainty can be propagated to σ^2_{ct} using the helix parameters:

$$\sigma^2_{ct\,|\,p_T} = (c\, m^{\text{w.a.}}_{B_q})^2 \sum_{\text{tracks}} \nabla_{\text{helix}} \left(\frac{L_{xy} \cdot p_T^{\text{trk}}}{|p_T|^2}\right) \cdot \sigma^{\text{trk}}_{\text{helix}} \cdot \nabla_{\text{helix}} \left(\frac{L_{xy} \cdot p_T^{\text{trk}}}{|p_T|^2}\right), \tag{5.8}$$

where the sum runs over the decay components used in the vertex fit, $\sigma^{\text{trk}}_{\text{helix}}$ is the covariance matrix of the track parameters and ∇_{helix} is the gradient over the parameters of the track helix.

The two components are then summed in quadrature. The uncertainty on L_{xy} is found to dominate for most of the ct spectrum, and the uncertainty in the *b*-meson momentum takes over only for very high values of ct ($ct \gtrsim 3$ mm).

[VII] The uncertainty in the PDG mass is negligible compared to the other sources.

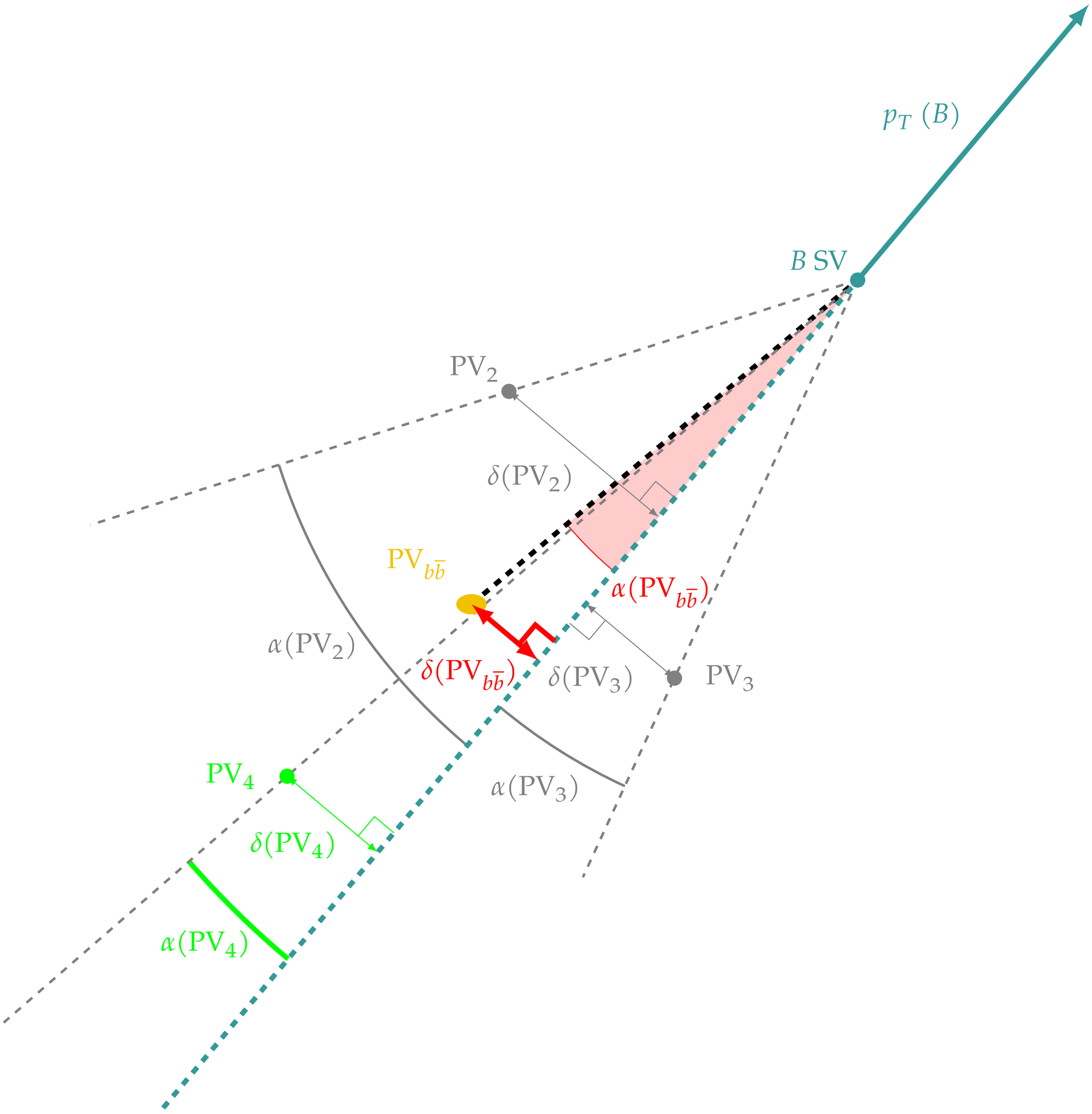


Figure 5.1: Comparison of the methods used for finding the PV. The *distance* method compares the δ of the various PV and would select the yellow PV ($\mathrm{PV}_{b\bar{b}}$), while the *pointing angle* method compares the α and would select the bright green PV (PV_4). The *beam spot* method bypasses this using a single point averaging all PVs.

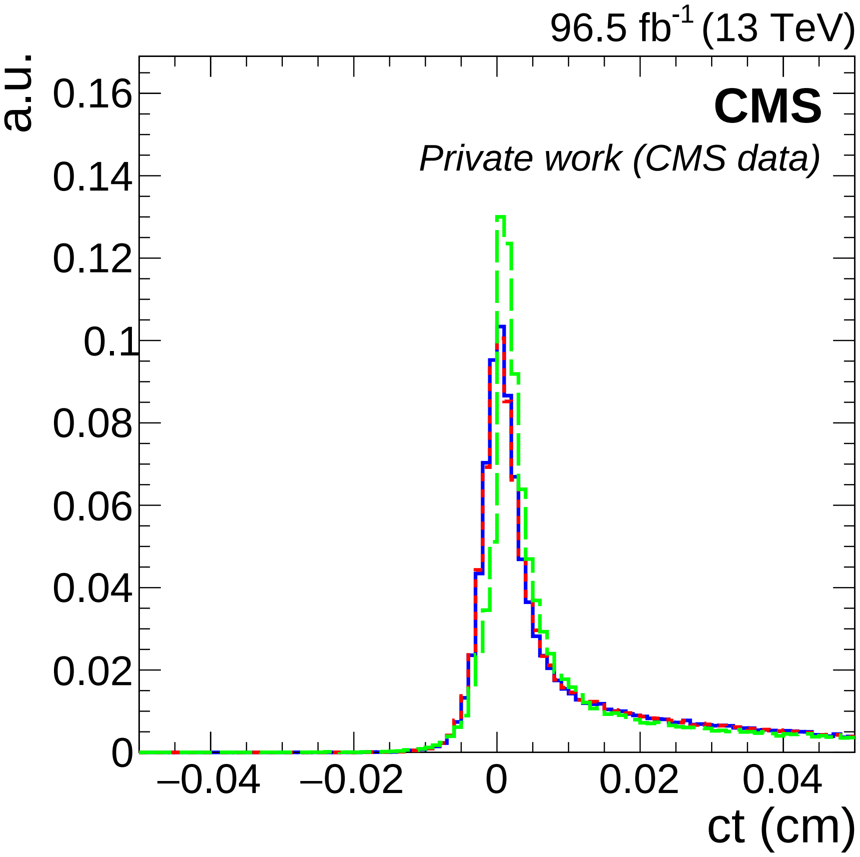


Figure 5.2: Proper decay length distribution for prompt candidates in *dataset_JpsiMuon* (with the *ct* selection cut removed) for different choice of PV. The distance method is shown as the plain blue line, the pointing angle is shown in long dashed green, and the BeamSpot is shown in short dashed red (overlapping with the distance method).

DECAY LENGTH SELECTION REQUIREMENTS To reduce the contribution from *prompt* background, a cut is applied on the proper decay length. 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 randomly 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 the considered b-meson 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-meson have a relatively long average lifetime ($\approx$ 1.5 ps), which makes them travel $\mathcal{O}$(mm) before decaying. After optimization, the cuts were placed at $ct > 60\,\mu$m for events selected with the `HLT_JpsiMu` trigger, and $ct > 100\,\mu$m for `HLT_JpsiTrkTrk`. The difference is expected since events at 100 μm are already affected by the trigger displacement selection.

5.3 SELECTED B_s^0 DATA SAMPLE

A total of 623 121 $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ candidates are selected, divided into 67 908 and 555 213 for *dataset_JpsiMuon* and *dataset_JpsiTrkTrk*, respectively. Tab. 5.8 reports a more detailed breakdown of the number of selected candidates, together with a preliminary estimation of the number of signal candidates obtained with a fit to the invariant mass distribution using a Johnson's S_U distribution[VIII]. The integrated luminosity and the selected number of events for each data-taking period are reported in Tab. 5.9. The 2017 data show an overall lower event rate, with the latter part (2017-F) displaying the lowest value. This is a consequence of the 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.

The B_s^0, J/ψ, and $\phi(1020)$ mesons invariant mass and proper decay length distributions for the selected candidates are shown in Fig. 5.3 and Fig. 5.4, respectively. It should be noted that the invariant mass distribution of J/ψ candidates is practically background-free, which is not the case for B_s^0 and $\phi(1020)$ candidates. This observation suggests that the majority of J/ψ candidates are genuine and most of the $B_s^0 \to J/\psi\,\phi(1020)$ background comes from either combination of prompt J/ψ and two random charged tracks or a $\phi(1020)$ ("prompt background") 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 ("B^0 background").

5.4 SIMULATED SAMPLES

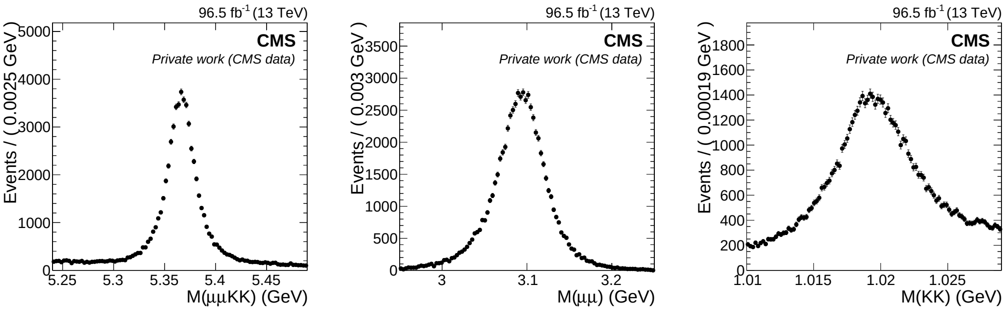
Several simulated samples are used in the tasks described in this dissertation. Simulated events are used to train the various tagging algorithms described in Part III, as well as the parametrization of the efficiencies and background templates used for the fit in Part IV. The main simulated samples used in this thesis are:

- *Regular* $B_s^0 \to J/\psi\,\phi(1020)$ decays, which are used in all the analysis steps, from optimizing the tagging tools to developing the fit model for the CP violation analysis. These samples have been generated with values of physics parameters close to the world-averages ones. Hereafter, these samples will only be indicated as "simulated $B_s^0 \to J/\psi\,\phi(1020)$ samples", without further specification.
- $B_s^0 \to J/\psi\,\phi(1020)$ decays, where $\Delta\Gamma_s$ is forced to be equal to 0 ps^{-1}, which are used to measure the angular efficiency in the CP violation analysis.

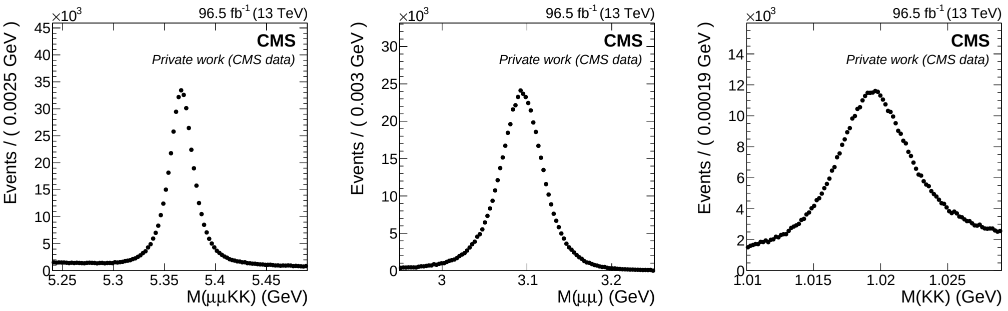
[VIII] The mass fit is described more in depth in Chapter 13.

Table 5.7: Offline selection requirements specific for *dataset_JpsiTrkTrk*.

Variable	Requirement	Optimised
HLT	`HLT_JpsiTrkTrk`	
HLT veto:	`HLT_JpsiMu`	
$p_T(\mu)$	> 4 GeV	Yes
$p_T(K)$	> 0.9 GeV	Yes
$p_T(\mu^+\mu^-)$	> 6.9 GeV	
$ct(B_s^0)$	$> 100\ \mu$m	Yes
$ct(B_s^0)/\sigma_{ct}$	> 3	



(a) B_s^0 invariant mass distribution for *dataset_JpsiMuon*.
(b) J/ψ invariant mass distribution for *dataset_JpsiMuon*.
(c) $\phi(1020)$ invariant mass distribution for *dataset_JpsiMuon*.



(d) B_s^0 invariant mass distribution for *dataset_JpsiTrkTrk*.
(e) J/ψ invariant mass distribution for *dataset_JpsiTrkTrk*.
(f) $\phi(1020)$ invariant mass distribution for *dataset_JpsiTrkTrk*.

Figure 5.3: Invariant mass distribution after all selection requirements for the B_s^0, J/ψ, and $\phi(1020)$ candidates.

Table 5.8: Number of selected candidates and number of $B_s^0 \to J/\psi\,\phi(1020)$ signal candidates estimated with a fit to the invariant mass distribution.

Data set	N. of candidates	N. of signal candidates
dataset_JpsiMuon (2017)	26 785	$20\,850 \pm 210$
dataset_JpsiMuon (2018)	41 123	$32\,250 \pm 260$
dataset_JpsiTrkTrk (2017)	141 001	$115\,200 \pm 440$
dataset_JpsiTrkTrk (2018)	414 212	$322\,970 \pm 770$
dataset_JpsiMuon (tot.)	67 908	$53\,100 \pm 340$
dataset_JpsiTrkTrk (tot.)	555 213	$438\,170 \pm 890$
Total	623 121	$491\,270 \pm 950$

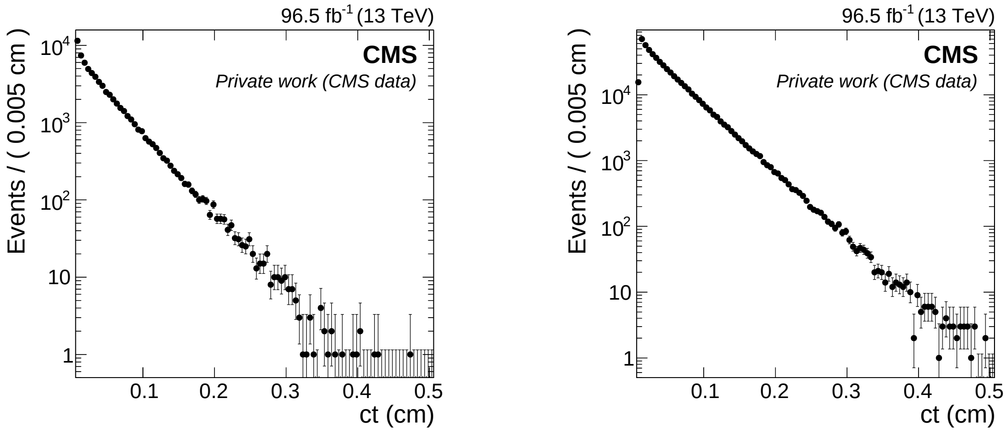
Table 5.9: Number of selected $B_s^0 \to J/\psi\,\phi(1020)$ signal candidates for each data-taking period.

Data-taking period	$\mathcal{L}_{\text{int}}$ [fb^{-1}]	N(*dataset_JpsiMuon*)	N(*dataset_JpsiTrkTrk*)
2017-C	9.57	5950 ± 110	36510 ± 250
2017-D	4.25	2729 ± 79	16900 ± 170
2017-E	9.31	5580 ± 110	31170 ± 230
2017-F	13.54	6590 ± 120	30790 ± 230
2018-A	14.03	7900 ± 140	70900 ± 350
2018-B	7.06	3949 ± 95	38100 ± 270
2018-C	6.89	3654 ± 87	37450 ± 260
2018-D	31.83	16750 ± 190	176480 ± 570

Setting $\Delta\Gamma_s = 0$ is equivalent to requiring 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[IX]. Since the tagging algorithms are not sensitive to the specifics of the B_s^0 decay, these samples can also be used in the training of the tagging algorithms, effectively doubling the number of events available. These samples are referred to as "$B_s^0 \to J/\psi\,\phi(1020)$ ($\Delta\Gamma_s = 0$)" hereafter.

- $B^+ \to J/\psi\,K^+$ decays, used in the calibration of the Same Side tagger (Chapter 10), and to evaluate the tagging systematic uncertainties for other opposite side taggers (Chapter 14).
- $B^0 \to J/\psi\,K^*(892)^0$ decays, used for the estimation of the shape of the B^0 background bkg component distributions.

[IX] 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 it impossible to measure the time and angular efficiency separately.



(a) B_s^0 proper decay length distribution for *dataset_JpsiMuon*

(b) B_s^0 proper decay length distribution for *dataset_JpsiTrkTrk*

Figure 5.4: Proper decay length distribution after all selection requirements for the $B_s^0 \rightarrow J/\psi\,\phi(1020)$ candidates.

Table 5.10: Number of generated events by physics process 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.

Simulated process	$N_{\text{evt}}^{2017} \cdot 10^6$	$N_{\text{evt}}^{2018} \cdot 10^6$
$B_s^0 \rightarrow J/\psi\,\phi(1020)$	95	97
$B_s^0 \rightarrow J/\psi\,\phi(1020)$ $(\Delta\Gamma_s = 0)$	93	93
$B^+ \rightarrow J/\psi\,K^+$	25	25
$B^0 \rightarrow J/\psi\,K^*(892)^0$	74	74
$B_s^0 \rightarrow J/\psi\,\phi(1020)$ (no filters)	99	

- $B_s^0 \rightarrow J/\psi\,\phi(1020)$ decays where no selection filters are applied at generation level and no reconstruction is performed. This sample is used in the angular efficiency measurements as the denominator of the ratio. The absence of generation filters, such as asking that the final-state particles are contained in the CMS geometrical acceptance, is essential to avoid biases. All other samples are produced requiring the final-state particles inside the CMS geometrical acceptance. This sample is generated with the same physics parameters as the $B_s^0 \rightarrow J/\psi\,\phi(1020)$ $(\Delta\Gamma_s = 0)$ sample.

The simulated processes and the number of generated events are summarized in Tab. 5.10. All MC samples are produced with the procedure described in Section 4.3.

Part III

FLAVOR TAGGING

An essential component of measuring mixing-related CP violation in neutral-meson systems is the determination of the initial flavor of the meson, called tagging. Flavor tagging is a complex process, for which many strategies exist, with different performances and trade-offs. The following chapters provide an introduction to the basic concepts of flavor tagging, including the two classes of algorithms that can be employed and methods to estimate their performances in a way that relates to the measurement precision. Four tagging algorithms are discussed in detail in this part, with focus on the jet-based and same side algorithms that were developed during the work in this thesis. Finally, a formalism for the combination of different, independent taggers is provided, and the performance is evaluated in the context of a possible application of the framework developed.

6

INTRODUCTION AND COMMON TOOLS

In a CPV analysis, flavor tagging is the process in which the signal particle is assigned a label estimating its quark content. The final flavor of a meson can often be estimated using the products of its decay, but the initial flavor, if changing in-flight, can only be estimated using specific algorithms ("taggers").

In oscillation analysis, flavor tagging at production time is essential because the measurement is sensitive to the difference between the initial and final flavors of the meson. In CP-neutral analyses such as the ϕ_s measurement that will be shown in Part IV the core parameters are sensitive to the difference in the decay rate between the initial B_s^0 and initial $\overline{B}_s^0$ mesons.

This chapter presents an overview of the principles behind flavor tagging and a description of the two main classes of flavor tagging algorithms. The following chapters contains the details of the four main taggers used in CPV analyses at CMS. A special focus is posed on the opposite side jet and same-side taggers, which were fully developed in the context of this thesis. Finally, a formalism for the combination of the results of multiple taggers is presented, and the final performance is evaluated in the datasets used in the ϕ_s measurement. To provide an absolute scale of the performance of the taggers presented here, the final performances will be compared with what was reported in similar ϕ_s measurements from other LHC experiments (ATLAS and LHCb), as well as previous CMS results.

6.1 FLAVOR TAGGING FUNDAMENTALS AND DEFINITIONS

Flavor tagging algorithms work on a probabilistic basis. The tagging decision is assigned at each event, and it is defined as:

$$\xi_{\text{tag}} = \begin{cases} +1 & \text{for } B_q \ (\overline{b}q \text{ content}) \, , \\ -1 & \text{for } \overline{B}_q \ (b\overline{q} \text{ content}) \, , \\ \ \ 0 & \text{if no tagging decision is made.} \end{cases} \tag{6.1}$$

Events are first classified as "tagged" and "untagged" (UT), depending on the outcome of the tag decision. Tagged events are further classified based on whether the tagging decision is correct (i.e., matches the true flavor of the b-meson) or not. Events where the inferred flavor is the same as the true flavor are called "rightly tagged" (RT), and other events are "wrongly tagged" (WT).

These categories are used to construct the following figures of merit:

- The "*tagging efficiency*" is defined as the fraction of events tagged, regardless of the correctness of the inference, as

$$\varepsilon_{\text{tag}} = \frac{N_{\text{tag}}}{N_{\text{tot}}}\,, \tag{6.2}$$

  where N_{tag} and N_{tot} are the number of tagged and total events, respectively, defined as:

$$N_{\text{tag}} = N_{\text{RT}} + N_{\text{WT}}\,, \tag{6.3}$$

$$N_{\text{tot}} = N_{\text{RT}} + N_{\text{WT}} + N_{\text{UT}}\,. \tag{6.4}$$

- The "*mistag rate*" is the fraction of mistagged events, defined as

$$\omega_{\text{tag}} = \frac{N_{\text{WT}}}{N_{\text{tag}}} = \frac{N_{\text{WT}}}{N_{\text{RT}} + N_{\text{WT}}}\,. \tag{6.5}$$

- The "tagging power" is a rescaled efficiency that takes into account the fraction of mistagged events, as

$$P_{\text{tag}} = \varepsilon_{\text{tag}}\big(1 - 2\omega_{\text{tag}}\big)^2 = \varepsilon_{\text{tag}}\mathcal{D}^2_{\text{tag}}\,, \tag{6.6}$$

  where $\mathcal{D}_{\text{tag}}$ is the *tagging dilution*, a measurement of the performance degradation due to mistagged events.

Two immediate consequences of these definitions are as follows:

$$\omega_{\text{tag}} = 0.5 \leftrightarrow P_{\text{tag}} = 0\,, \tag{6.7}$$

$$\omega_{\text{tag}} = 0 \leftrightarrow P_{\text{tag}} = \varepsilon_{\text{tag}}\,. \tag{6.8}$$

The tagging power is the most widely used figure of merit for flavor-sensitive analyses, as it directly correlates with the statistical power. We can prove this by starting from the true and estimated tagging asymmetries:

$$\widetilde{A}_{\text{flav}} = \frac{\widetilde{N}_B - \widetilde{N}_{\overline{B}}}{\widetilde{N}_B + \widetilde{N}_{\overline{B}}} \quad \text{and} \quad A_{\text{flav}} = \frac{N_B - N_{\overline{B}}}{N_B + N_{\overline{B}}}\,, \tag{6.9}$$

using the "tilde" ($\sim$) to indicate true quantities, in contrast to measured ones. The observed number of B and $\overline{B}$ events can be written as a function of the true ones as:

$$\begin{aligned} N_B &= \varepsilon_B\widetilde{N}_B(1 - \omega_B) + \varepsilon_{\overline{B}}\widetilde{N}_{\overline{B}}\omega_{\overline{B}}\,, \\ N_{\overline{B}} &= \varepsilon_{\overline{B}}\widetilde{N}_{\overline{B}}(1 - \omega_{\overline{B}}) + \varepsilon_B\widetilde{N}_B\omega_B\,, \end{aligned} \tag{6.10}$$

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

$$A_{\text{flav}} = \frac{\varepsilon_{\text{tag}}(\overbrace{1-2\omega}^{\equiv \mathcal{D}_{\text{tag}}})(\widetilde{N}_B - \widetilde{N}_{\overline{B}})}{\varepsilon_{\text{tag}}(\widetilde{N}_B + \widetilde{N}_{\overline{B}})} = \mathcal{D}_{\text{tag}}\widetilde{A}_{\text{flav}} \,. \tag{6.11}$$

The term dilution comes from the fact that mistagged events effectively "dilute" the tagging asymmetry by a factor $\mathcal{D}_{\text{tag}} \equiv 1 - 2\omega_{\text{tag}}$. Neglecting the uncertainty on the mistag rate, A_{flav} is measured with a statistical uncertainty of[I]:

$$\sigma_{A_{\text{flav}}} \approx \sqrt{\frac{1 - A^2_{\text{flav}}}{\varepsilon_{\text{tag}} N_{\text{tot}}}} \,, \tag{6.12}$$

where N_{tot} is the total number of events, as defined in Eq. 6.4. Using Eq. 6.12, the uncertainty on the true asymmetry can be evaluated as:

$$\sigma_{\widetilde{A}_{\text{flav}}} = \frac{\sigma_{A_{\text{flav}}}}{\mathcal{D}_{\text{tag}}} = \frac{\sqrt{1 - A^2_{\text{flav}}}}{\sqrt{\varepsilon_{\text{tag}} N_{\text{tot}}}\mathcal{D}_{\text{tag}}} \approx \frac{1}{\sqrt{\underbrace{\varepsilon_{\text{tag}}\mathcal{D}^2_{\text{tag}}}_{\equiv P_{\text{tag}}} N_{\text{tot}}}} = \frac{1}{\sqrt{P_{\text{tag}} N_{\text{tot}}}} \,, \tag{6.13}$$

where $\sqrt{1 - A^2_{\text{flav}}} \approx 1$ is assumed.

From these equations, it is clear that the tagging power behaves as an *effective efficiency*, constructed in such a way that a diluted data sample with tagging power P_{tag} has the same statistical power as a perfectly tagged sample (i. e., $\omega_{\text{tag}} = 0$) with $\varepsilon_{\text{tag}} = P_{\text{tag}}$.

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

6.1.1 *Flavor tagging with per-event mistag probability*

The mistag rate and tagging power defined in Eq. 6.5 and Eq. 6.6 are average quantities evaluated in a given data set, but can also be evaluated on a per-event basis to improve performance by assigning a "*per-event mistag probability*" (ω_{evt}).

[I] The uncertainty on A_{flav} is obtained by propagating the uncertainty on N_B and $N_{\overline{B}}$ in the A_{flav} definition reported in Eq. 6.9 (right). The number of events is assumed to follow a Poisson distribution.

The effective tagging power $\hat{P}_{\text{tag}}$ can be constructed as the average of the tagging power of each event, as

$$\hat{P}_{\text{tag}} = \frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tot}}} P_{\text{evt}} = \frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tot}}} \underbrace{\varepsilon_{\text{evt}}}_{0\text{ or }1} \mathcal{D}^2_{\text{evt}} = \frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tag}}} \mathcal{D}^2_{\text{evt}}\,, \tag{6.14}$$

where the "evt" subscript indicates per-event quantities and the efficiency of a single event ε_{evt} is either 0 or 1 for untagged and tagged event, respectively. The sum in Eq. 6.14 can be extended to N_{tot} by defining $\mathcal{D}_{\text{evt}}$ as equal to zero for untagged events. Accordingly, an effective dilution and an effective mistag rate are defined as:

$$\widehat{\mathcal{D}}_{\text{tag}} \equiv \sqrt{\frac{\hat{P}_{\text{tag}}}{\varepsilon_{\text{tag}}}}\,, \tag{6.15}$$

$$\widehat{\omega}_{\text{tag}} = \frac{1}{2}\left(1 - \sqrt{\frac{\hat{P}_{\text{tag}}}{\varepsilon_{\text{tag}}}}\right), \tag{6.16}$$

where the total efficiency $\varepsilon_{\text{tag}} = \sum_{\text{evt}}^{N_{\text{tag}}} \varepsilon_{\text{evt}} = N_{\text{tag}}/N_{\text{tot}}$ is equivalent to the averaged one defined in Eq. 6.2. The average tagging dilution $\mathcal{D}_{\text{tag}}$, already introduced in Eq. 6.6, can also be defined in terms of $\mathcal{D}_{\text{evt}}$ as:

$$\mathcal{D}_{\text{tag}} = \frac{1}{N_{\text{tag}}} \sum_{\text{evt}}^{N_{\text{tag}}} \mathcal{D}_{\text{evt}}\,. \tag{6.17}$$

To describe the spread of the $\mathcal{D}_{\text{evt}}$ distribution in a given sample, we can introduce the tagging dilution variance, defined as:

$$\begin{aligned}
\sigma^2_{\mathcal{D}_{\text{tag}}} &= \frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tag}}} (\mathcal{D}_{\text{evt}} - \mathcal{D}_{\text{tag}})^2 \\
&= \frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tag}}} \mathcal{D}^2_{\text{evt}} + \underbrace{\frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tag}}} \mathcal{D}^2_{\text{tag}}}_{\substack{=\frac{N_{\text{tag}}}{N_{\text{tot}}}\mathcal{D}^2_{\text{tag}} \\ =\varepsilon_{\text{tag}}\mathcal{D}^2_{\text{tag}}}} - \underbrace{\frac{2}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tag}}} \mathcal{D}_{\text{evt}}\mathcal{D}_{\text{tag}}}_{\substack{=\frac{2\mathcal{D}_{\text{tag}}}{N_{\text{tot}}}\frac{N_{\text{tag}}}{N_{\text{tag}}}\sum_{\text{evt}}^{N_{\text{tag}}}\mathcal{D}_{\text{evt}} \\ =2\varepsilon_{\text{tag}}\mathcal{D}^2_{\text{tag}}}} \\
&= \frac{1}{N_{\text{tot}}} \sum_{\text{evt}}^{N_{\text{tag}}} \mathcal{D}^2_{\text{evt}} - \varepsilon_{\text{tag}}\mathcal{D}^2_{\text{tag}} \\
&= \hat{P}_{\text{tag}} - \varepsilon_{\text{tag}}\mathcal{D}^2_{\text{tag}}
\end{aligned} \tag{6.18}$$

Since variance is by definition a positive value, Equation 6.18 shows that, by construction, the effective tagging power $\hat{P}_{\text{tag}}$ is greater or at least equal to the average one defined in Eq. 6.6. Performance improvement, compared to the averaged approach, is given by $\sigma^2_{\mathcal{D}_{\text{tag}}}$. The greater the variance of the dilution distribution, the greater the performance enhancement. This translates to the necessity of not only evaluating ω_{evt}, but also maximizing the variance of its distribution, that is, separating as much as possible right- and wrong-tagged events.

In this work, the probability of mistagging for each tagger is estimated with Deep Neural Networks (DNN).

6.2 DEEP NEURAL NETWORKS

Neural network refers to a class of algorithms which attempts to approximate a function through the use of linear combination of the inputs and non-linear transformations. The basic building block of each neural network is the neuron. Each neuron is connected to several inputs and produces a single output that can be sent to multiple other neurons. The output of the neuron is given by a weighted sum of its inputs, passing through a nonlinear function called the activation function. Neurons are typically organized into layers, and in the simplest case of a feedforward network the layers are ordered such that each layer can only take as input neurons from the previous one.

In Deep Neural Networks (DNN), between the input data and the output neuron there are a certain number of layers called hidden layers.

6.2.1 *Neural network training*

Training, or learning, is the process in which the weights of a neural network are optimized to improve its performance. Practically, this is done by defining a *loss* function which is repeatedly evaluated and minimized (numerically) during training.

A common choice for the minimization is to use the so-called stochastic gradient descent (SGD), a family of algorithms that searches the minimum of the function by moving iteratively in the parameter space, following the value of the function gradient at each step. The "stochastic" part of the name is used because each step evaluates the gradient using a small subset of the input data (called a "batch"), leading to an estimated gradient with a probability distribution around the true gradient value. The size of the batch is an important hyperparameter to tune during training. At each iteration, SGD updates the vector of weights $\vec{w}$, according to the gradient of the loss function Q, as:

$$\vec{w} \to \vec{w} - \alpha \nabla Q(\vec{w}) \,, \tag{6.19}$$

where α is the learning rate.

Common improvements over pure SGD are the addition of a ”momentum” term, which introduces in the weight update a term proportional to the previous update, as [116]:

$$\vec{w} \rightarrow \vec{w} - \alpha \nabla Q(\vec{w}) + \beta \Delta \vec{w} \,, \tag{6.20}$$

or with a learning rate depending on the first and second moments of the gradient, as in Adam [117].

In the hidden layers, the behavior of the network is not directly specified by the loss function: instead, the learning algorithm optimizes the weights using the hidden layers to decide how to best minimize the loss. This is one of the main advantages of DNNs compared to the use of high-level features, as the hidden layers can be thought of as performing automatic feature engineering starting from the input low-level ones.

OVERFITTING AND REGULARIZATION Due to their complexity, deep neural networks are prone to *overfitting*, a situation when the model adapts too well to the training data set and performs poorly outside of it. Overfitting can be detected by evaluating at each iteration the loss function in a completely separate data set, called the validation 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 under- and over-fitted models is shown in Fig. 6.1.

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 addition of `Dropout` layers and the *early stopping*.

`Dropout` layers [118, 119] are layers of neurons that behave differently during training and evaluation. During training, each layer has a random probability p (called the ”dropout rate”) of setting its output to 0 and a probability $1 - p$ of passing it forward unchanged. This act randomly simplifies the model in each optimizer step, forcing the network to not favor any specific feature, but to learn general properties. During the evaluation, instead, the input is scaled by a factor of p, giving an output that on average is the same as that of the training. The dropout rate is the key hyperparameter to optimize in order to reduce overfitting without losing performance.

Early stopping refers instead to the monitoring of the value of a specific function (often the loss itself) in the validation dataset during training. The training is allowed to continue only as long as the trend of the monitor function is not negative. After a certain number of steps (another tunable hyperparameter) in which the monitor function does not improve, the training is forcefully stopped.

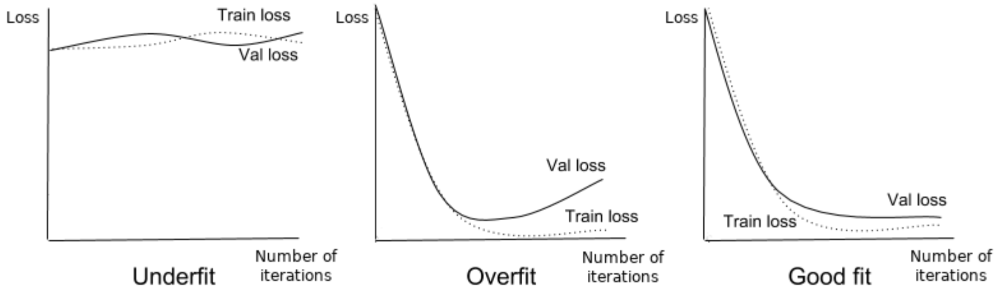


Figure 6.1: Learning curves for underfitted (left), overfitted (middle) and well fitted (right) models.

6.2.2 *DNN as probability estimators*

All the networks presented in this thesis are used for binary classification and are optimized through "supervised learning". This means that for each input event we have a corresponding true label, which the neural network is meant to estimate. A common approach is to use the neural network to estimate the "class probability", which means training it so that

$$s_{\mathrm{DNN}}(\vec{x}\,|\,\vec{w}) \sim P(y\,|\,\vec{x})\,, \tag{6.21}$$

where s_{DNN} is the output of the neural network, $\vec{x}$ is the vector of input features, $\vec{w}$ is the vector of neural network weights, y is the true label, and $P(y|\vec{x})$ is the probability of having label y when the input is $\vec{x}$ (what we call "class probability").

For the output layer to be interpreted as a single probability, a single output neuron with a *sigmoid* activation can be used[II]

$$\mathtt{sigmoid}(z) = \frac{1}{\mathrm{e}^{-z}+1} \tag{6.22}$$

The output of this layer satisfies the properties of a probability ($\mathtt{sigmoid}(z) \in [0,\,1]$), an additional step is needed before the approximation from Eq. 6.21 is correct, the use of the cross-entropy as loss function.

The cross-entropy is a function of the output, which can be seen as the likelihood of the conditional Bernoulli distribution of the labels. The label y, given the inputs $\vec{x}$, follows a Bernoulli distribution with probability

$$P_{bernoulli} = \begin{cases} P(y|\vec{x}) & y = 1 \\ 1 - P(y|\vec{x}) & y = 0 \end{cases}, \tag{6.23}$$

[II] When the events can be classified in more than $N > 2$ categories, the similar *softmax* function is used on a size N output layer. When $N = 2$ the two approaches can be proven to be mathematically equivalent.

and negative log-likelihood (that is, the cross-entropy)

$$\begin{aligned}\mathcal{L}_{\mathrm{CE}}^{\mathrm{dataset}} &= -\sum_{y=1}\log\big(P(y|\vec{x})\big) - \sum_{y=0}\log\big(1-P(y|\vec{x})\big)\\ &= -\sum y\cdot\log\big(P(y|\vec{x})\big) + (1-y)\cdot\log\big(1-P(y|\vec{x})\big)\\ &\sim -\sum y\cdot\log\big(s_{\mathrm{DNN}}(\vec{x}|\vec{w})\big) + (1-y)\cdot\log\big(1-s_{\mathrm{DNN}}(\vec{x}|\vec{w})\big)\,,\end{aligned} \tag{6.24}$$

where in the last passage we applied Eq. 6.21, replacing the real probability with the neural network score, which, in the desired outcome of training, will approximate it.

By minimizing the logarithmic likelihood as a function of the weights $\vec{w}$, the score of the neural network converges to the true probability of the label and can be used in the estimation.

6.2.3 *DNN calibration*

Despite the minimization of cross-entropy, the neural network score is not guaranteed to be a good approximation of the label probability: despite the regularization techniques described in paragraph 6.2.1, a high complexity neural network can overfit the likelihood [120]. Furthermore, the training is executed in simulated samples to avoid biases from background events or wrong ”true” labels, and. due to unavoidable simulation mismodeling, the predicted mistag probability is not the same as that measured in data.

For a correct estimation, the neural network score goes through an additional step of calibration, where it is transformed using a calibration function. For this thesis, the procedure chosen is Platt scaling [121]. In Platt scaling, the calibration function is a linear function in the *logit* space[III]

$$\texttt{logit}(p) = \log\frac{p}{1-p} \tag{6.25}$$

This has been proven to be an excellent strategy for calibrating DNNs, which presents several good properties such as constraining the output in the $[0,\ 1]$ range. This allows the calibrated mistag probability to still be treated as a probability.

6.3 TAGGING STRATEGIES

Flavor tagging techniques can generally be classified into two broad categories depending on the *side* of the event studied: *same-side* (SS) and *opposite-side* (OS) taggers. A schematic overview of the various SS and OS algorithms developed in this work is shown in Fig. 6.2.

[III] Since the logit function is the inverse of the *sigmoid*, this is equivalent to a linear calibration before the last activation function

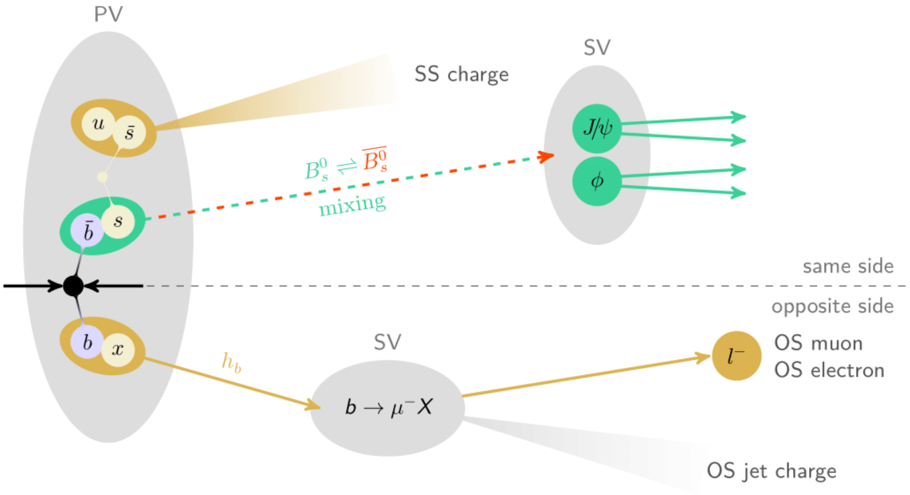


Figure 6.2: Schematic overview of the various flavor tagging techniques studied in this thesis.

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 time, 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:

$$\begin{array}{l} K^+ \xrightarrow{\text{tag}} \text{signal } B_s^0 \\ K^- \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \end{array} . \tag{6.26}$$

In the case of a signal B^0 meson, analogous techniques can be used to exploit charged pions and protons. Since CMS does not have particle identification (PID) capabilities to distinguish kaons from pions, a different approach is used in this work, as described in Chapter 10.

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. 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 analyzed. Once the flavor of the OS b-hadron is known, the tag decision is simply given by charge conjugation as:

$$\begin{array}{l} \text{OS } b \xrightarrow{\text{tag}} \text{signal } B_s^0 \\ \text{OS } \bar{b} \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \end{array} . \tag{6.27}$$

Opposite-side tagging is affected by the following irreducible sources of dilution:

- *multiple $b\bar{b}$ pair production*, where more than one pair of heavy quarks is produced ($b\bar{b}\,b\bar{b}$ or $b\bar{b}\,c\bar{c}$) and a heavy quark not correlated with B_s^0 is studied [IV];
- *mixing of the OS b-hadron*, which reverses the correlation between the reconstructed OS b-hadron initial flavor and the B_s^0.

These two combined effects 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, the average mistag rates for OS lepton tagging techniques are about $\omega_{\text{tag}} = 30 \sim 40\%$ ($\mathcal{D}_{\text{tag}}^2 = 0.04 \sim 0.15$).

6.4 ALGORITHMS OVERVIEW

The following techniques have been developed in CMS for the CPV analysis presented in Part IV: OS-muon, OS-electron, OS-jet, and SS tagging algorithms.

LEPTON TAGGERS The OS-muon and OS-electron algorithms exploit semileptonic $b \to \ell^- X$ decays of the OS b-hadron. Tagging inference is performed through three steps:

1. *OS-lepton selection*.
2. *Tagging decision*, which exploits the charge of the OS lepton following the logic
$$\begin{aligned} &\text{OS } \ell^- \to \text{OS } b \xrightarrow{\text{tag}} \text{signal } B_s^0 \\ &\text{OS } \ell^+ \to \text{OS } \bar{b} \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \end{aligned} . \tag{6.28}$$
3. *Per-event mistag probability estimation*, which is obtained using a DNN trained to distinguish mistagged events calibrated in data using self-tagging $B^+ \to J/\psi K^+$ decays. Since the neural network in these cases is trained to return the probability of being correctly tagged, the mistag is simply
$$\omega_{\text{evt}}^{\text{DNN}} = 1 - s_{\text{DNN}} \tag{6.29}$$

LEPTON-*LESS* TAGGERS Inclusive taggers such as the OS-jet and SS tagger do not have a single strong tagging variable like the lepton charge. As a consequence, a DNN is used to both make the tag decision and estimate the per-event mistag probability. Inference is performed with the following steps:

1. *Selection of the DNN input objects* (jets, tracks, etc.).

[IV] The second $b\bar{b}$ pair can be produced for example from PU collisions or from additional processes in the main collision (e.g. double gluon splitting).

2. *Tagging decision and per-event mistag probability estimation*, which is based on the output score ("s_{DNN}") of the DNN trained to distinguish B_s^0 from $\overline{B}_s^0$

$$s_{\text{DNN}} > 0.5 + \epsilon \xrightarrow{\text{tag}} \text{signal } B_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = 1 - s_{\text{DNN}} \tag{6.30}$$

$$s_{\text{DNN}} < 0.5 - \epsilon \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = s_{\text{DNN}} \tag{6.31}$$

   where ϵ is a small positive number that is used to remove the region of the DNN output where the mistag probability is the highest. Again, the DNN is calibrated in data, using a control channel.

In all cases, the signal b-meson and PV need first to be reconstructed.

6.5 DATA SETS

The tagging algorithms are developed in simulated MC events, using the samples described in Section 4.3. They are also reported in Tab. 6.1 and Tab. 6.2 for reference.

Table 6.1: Simulated MC samples corresponding to the 2017 data-taking conditions. The number of events refers to the generated ones before offline selection.

Process	Description/usage	#evt
$B_s^0 \to J/\psi\,\phi(1020)$	Signal	$95 \cdot 10^6$
$B_s^0 \to J/\psi\,\phi(1020)$	Signal ($\Delta\Gamma_s = 0$)	$93 \cdot 10^6$
$B^+ \to J/\psi\,K^+$	Calibration channel	$24 \cdot 10^6$

Table 6.2: Simulated MC samples corresponding to the 2018 data-taking conditions. The number of events refers to the generated ones before offline selection.

Process	Description	#evt
$B_s^0 \to J/\psi\,\phi(1020)$	Signal	$97 \cdot 10^6$
$B_s^0 \to J/\psi\,\phi(1020)$	Signal ($\Delta\Gamma_s = 0$)	$93 \cdot 10^6$
$B^+ \to J/\psi\,K^+$	Calibration channel	$25 \cdot 10^6$

TRIGGERS Of the four taggers, only the muon based one is trained on events that pass the `HLT_JpsiMu` trigger. The muon tagger performance is greatly enhanced when selecting this trigger, since the third required muon is used as a candidate to infer the initial flavor of B_s^0.

For events passing the `HLT_JpsiTrkTrk` trigger, all four taggers are used. However, since, as mentioned in Section 5.1, `HLT_JpsiMu` events are vetoed in *dataset_JpsiTrkTrk*, the performance of the muon tagger is severely reduced.

6.6 CALIBRATION PROCEDURE FOR OPPOSITE-SIDE METHODS

For correct modeling of probabilities in the data sample, the calibration function defined in Section 6.2.3 must be fitted in data. ω_{evt} is calibrated in data using self-tagging $B^+ \to J/\psi K^+$ decays, where the meson cannot oscillate and the charge of the final state kaon unambiguously determines the flavor of the B^+ meson. 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 the simulated decays of $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ [V].

The calibration procedure is the same for all opposite-side tagging algorithms, separately for the 2017 and 2018 data sets and for the two different triggers, and can be schematized in the following steps:

1. The total number of events (N_{tot}) is estimated by fitting the $\mu^+\mu^-K^+$ invariant mass distribution. The signal is modeled with a Johnson's S_U distribution[VI], whereas an inverse error function plus a first-order polynomial is used for the background.

2. The data sample is binned according to the mistag probability as estimated by the DNN ($\omega_{\text{evt}}^{\text{DNN}}$). For the OS-jet tagger, bins of width equal to 0.04 of $\omega_{\text{evt}}^{\text{DNN}}$ are used, while bins of with 0.02 are used for the other algorithms.

3. The charge of the kaon (which gives the flavor of the B^+ meson) and the tag inference are used to further classify events as correctly tagged and mistagged, based on their agreement or lack thereof.

4. A simultaneous fit to each bin is performed to extract all the $N_{\text{RT}}^{\text{bin}}$ and $N_{\text{WT}}^{\text{bin}}$ yields. The shape of the mass distribution is shared between all bins, while the yields $N_{\text{RT}}^{\text{bin}}$ and $N_{\text{WT}}^{\text{bin}}$ are determined for each bin. The fit obtained in Step 1 is used as a starting point for the shape. This strategy ensures the stability of the yields even in very low statistics bins. Examples of such fits are reported in Fig. 6.3.

5. The efficiency and mistag fraction of each bin are computed as

$$\varepsilon_{\text{tag}}^{\text{bin}} = \frac{N_{\text{RT}}^{\text{bin}} + N_{\text{WT}}^{\text{bin}}}{N_{\text{tot}}^{\text{bin}}} \tag{6.32}$$

$$\omega_{\text{tag}}^{\text{bin}} = \frac{N_{\text{WT}}^{\text{bin}}}{N_{\text{RT}}^{\text{bin}} + N_{\text{WT}}^{\text{bin}}} \tag{6.33}$$

[V] In a CPV analysis, these eventual differences would be treated as systematic uncertainties

[VI] Described more in depth in Chapter 13.

6. The $(\omega_{\text{evt}}^{\text{DNN}}, \omega_{\text{tag}}^{\text{bin}})$ pairs are fitted to obtain the calibration function:

$$\omega_{\text{tag}}^{\text{bin}} = f(\omega_{\text{evt}}^{\text{DNN}})\,. \tag{6.34}$$

where the calibration function f is given by the Platt scaling

$$f(\omega_{\text{evt}}) = \texttt{sigmoid}(a + b \cdot \texttt{logit}(\omega_{\text{evt}}))\,, \tag{6.35}$$

7. Unless stated differently[VII], the per-event mistag probability to be used in the analysis is, then,

$$\omega_{\text{evt}} \equiv f(\omega_{\text{evt}}^{\text{DNN}})\,. \tag{6.36}$$

8. Using bins, the total tagging power from Eq. 6.14 can also be approximated as the weighted sums of the $P_{\text{tag}}^{\text{bin}}$, the average tagging power inside each bin

$$\hat{P}_{\text{tag}} \approx \frac{1}{N_{\text{tot}}} \sum_{\text{bin}} N_{\text{tot}}^{\text{bin}}\, P_{\text{tag}}^{\text{bin}} = \frac{1}{N_{\text{tot}}} \sum_{\text{bin}} N_{\text{tot}}^{\text{bin}}\, \varepsilon_{\text{tag}}^{\text{bin}} (1 - 2\,\omega_{\text{tag}}^{\text{bin}})^2\,, \tag{6.37}$$

which is equivalent to assigning mistag $\omega_{\text{tag}}^{\text{bin}}$ to all events in each bin. This formulation has the advantage of depending only on the number of correct- and wrong-tagged and untagged events inside the bin, which can be easily estimated even in presence of background, with a simple fit to the meson mass spectrum. Similarly, the efficiency ε_{tag} can be computed[VIII] as

$$\varepsilon_{\text{tag}} = \frac{1}{N_{\text{tot}}} \sum_{\text{bin}} N_{\text{tot}}^{\text{bin}} \varepsilon_{\text{bin}}\,. \tag{6.38}$$

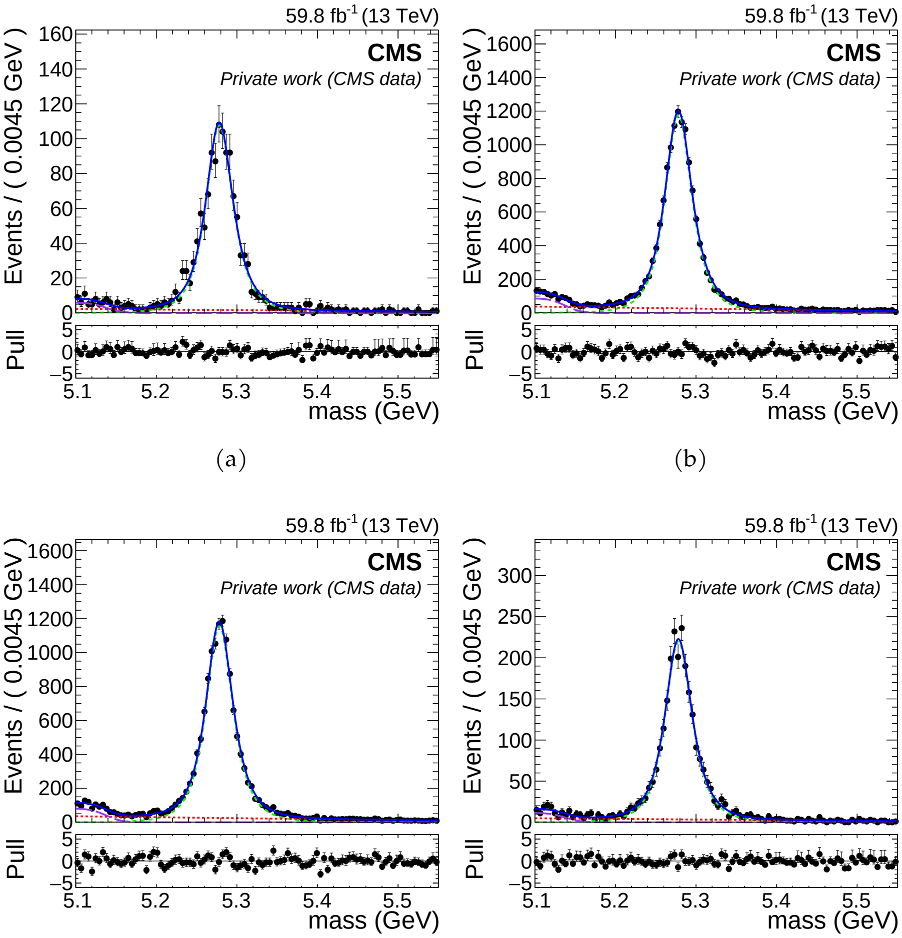
In simulated samples, this is found to be in excellent agreement with $\hat{P}_{\text{tag}}$ defined in Eq. 6.14. This method will be referred from here on as $\sum P_{\text{bin}}$.

The calibration procedure can also be applied to MC samples (both $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\,\phi(1020)$) with some key differences:

- the signal b-hadron is matched to the corresponding GEN object and the correct flavor is given by the generation information;
- the various event yields are given by a simple counting instead of a mass fit since no background is present.

[VII] This is the case for the SS tagger

[VIII] This is not an approximation, since ε_{tag} is linear in the number of events.



(a) (b)

(c) (d)

Figure 6.3: Examples of fit to the $B^+ \to J/\psi K^+$ invariant mass distribution.

7

OPPOSITE-SIDE MUON TAGGING

The principles of OS-lepton tagging were described in Section 6.4. The three core steps are OS-lepton selection, tagging decision[I], and mistag probability estimation. This section will describe two neural networks: the first one (*DALIHA_UL*) is used in the muon selection to reduce contamination from misidentified hadrons, while the second (*DeepMuonTagger*) is used for flavor tagging.

7.1 OS-MUON SELECTION

OS-muon candidates are searched in events with a reconstructed b-meson decay and a selected PV, as described in Chapter 5, with tracks associated with the reconstructed b-meson excluded from the pool of candidates.

Candidates must leave a signal both in the tracker and in the muon detector (”Global” muons), have $p_T \geq 2$ GeV, and $|\eta| \leq 2.4$. To reduce the background from PU interactions, the absolute value of the candidate muon's impact parameter with respect to the selected PV is required to be smaller than 1 cm in the longitudinal direction. OS-muon candidates are required to pass the so-called “Loose” working point in the *DALIHA_UL* DNN discriminator described in Section 7.1.1. For candidates reconstructed as Global Muons (described in Section 4.2.2), this corresponds to an efficiency for genuine muon of 98%, and 36% for ”fake” muons (mostly misidentified kaons and pions)[II]. The classifier efficiencies and the working point used are shown in Fig. 7.1.

The selection requirements are summarized in Tab. 7.1. In the case of multiple OS-muon candidates, the one with the highest p_T is chosen. $\Delta R_{\eta,\phi}$ in the table refers to the euclidean distance in the $\eta - \phi$ plane between the two objects.

About $\approx 50\%$ and $\approx 1.5\%$ of the events in *dataset_JpsiMuon* and *dataset_JpsiTrkTrk*, respectively, contain an OS-muon candidate. As a reminder, events passing `HLT_JpsiMu` are vetoed from appearing in *dataset_JpsiTrkTrk*, which greatly reduces the efficiency of the OS-muon tagger.

7.1.1 *Muon discriminator against hadronic fakes* (*DALIHA_UL*)

Long-lived particles, such as pions, kaons, protons, and electrons, can live enough to leave a signal in the muon system or decay in flight into a muon and therefore be reconstructed as (fake) muons by CMS. These fakes carry no charge-flavor

---

[I] Which exploits the charge of the OS lepton, e.g.: OS $\ell^- \rightarrow$ OS $b \xrightarrow{\text{tag}}$ signal B_s^0.

[II] The ”fake” muon efficiency is defined as the fraction of misidentified hadrons which also pass the discriminator selection.

correlation, therefore, degrading the tagging performance. This dilution source is very prominent in very low p_T regions such as the one where the OS-muon candidates are selected. To reduce this background, in the previous iteration of the ϕ_s measurement in CMS, a deep neural network (DNN) discriminator was developed to discriminate real muons from fake muons. For this work, the DNN has been retrained in the same samples used for tagger training, reported in Tab. 6.1 and Tab. 6.2. The model has been developed using the KERAS [122] machine learning library within the TMVA [114] toolkit. The discriminator is called *DALIHA_UL* (Discriminator Against LIght HAdron Ultra Legacy).

TRAINING DATA SET Muons reconstructed as Global Muon that satisfy the requirements reported in Tab. 7.2 are used to train *DALIHA_UL*. Tracks and muons associated with the filtered process characteristic of the MC sample (e.g. the two muons and kaons in $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ decays) are removed from the candidate pool. All training candidates are required to match a GEN particle with $\Delta R_{\eta,\phi} < 0.12$ and $\left|p_T^{\text{gen}} - p_T^{\mu}\right|/p_T^{\text{gen}} < 0.3$. Reconstructed muons matched to a genuine GEN muon are labeled as *Signal*, while those matched to a charged kaon, charged pion, electron, or proton are called *Background* (or *Fakes*). Since this discriminator will be used to select OS-muons, *Signal* muons are also required to be associated with a b-hadron ancestor at the GEN level.

A total of 7.8 M *Signal* events and 2.6 M *Background* events are selected. The dataset has been divided into subsets *Training*, *Validation* and *Test* subsets with respective fractions of 50%, 25% and 25%. The training dataset is used to train the model, the validation dataset is used to evaluate the performance in each training epoch, and the test dataset is used to evaluate the performance and overtraining.

INPUT FEATURE The discriminator accepts as input features:

- p_T: muon transverse momentum;
- η: muon pseudorapidity;

Table 7.1: Selection requirements for OS-muon candidates.

Variable	Requirement
Reconstruction	*Global*
Reconstructed b-meson tracks	Excluded
p_T	≥ 2 GeV
$\lvert\eta\rvert$	≤ 2.4
$\lvert d_z\rvert$ w.r.t. PV	≤ 1 cm
$\Delta R_{\eta,\phi}$(b-meson)	≥ 0.4
DALIHA_UL score	≥ 0.2 (Loose WP)

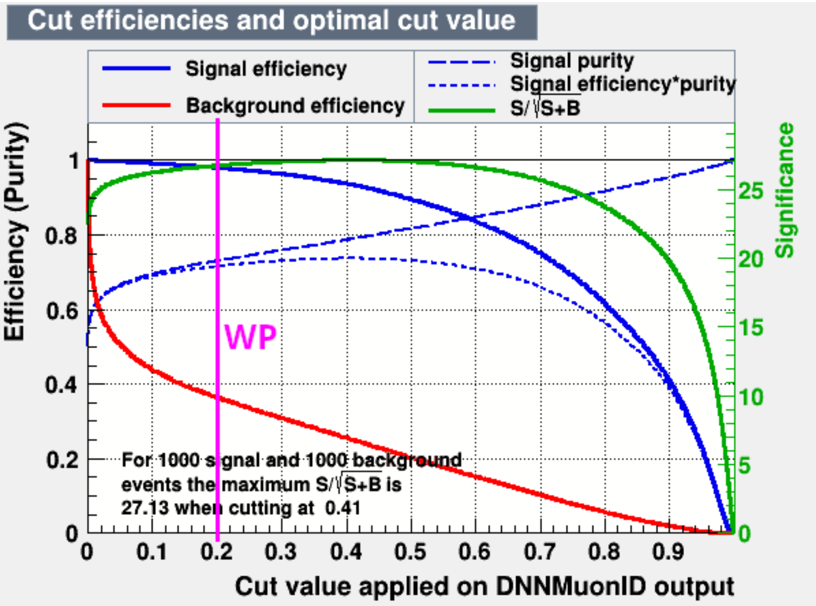


Figure 7.1: Classifier efficiencies for *DALIHA_UL* as a function of the cut on the classifier score. The solid magenta line represents the used working point to select the OS-muon candidate.

Table 7.2: Selection requirements for the training candidates for *DALIHA_UL*.

Variable	Requirement
Reconstruction	*Global*
p_T	≥ 2 GeV
$\lvert\eta\rvert$	≤ 2.4
Muon track quality	*High Purity*
Match to GEN	$\Delta R_{\eta,\phi} < 0.12$ $\lvert p_T^{\text{gen}} - p_T^{\mu}\rvert / p_T^{\text{gen}} < 0.3$
Signal definition	Match to a GEN muon Has a b-hadron ancestor
Background definition	Match to a GEN $\pi^{\pm}$, $K^{\pm}$, $e^{\pm}$, $p^{\pm}$

- *Iso*$_\mu$: muon ParticleFlow isolation with beta correction, defined as

$$\text{Iso}_\mu = \frac{\sum_{h^\pm} p_T + \max\Big(0,\ \sum_{h^0} E_T + \sum_\gamma E_T - 0.5 \sum_{h^\pm_{\text{PU}}} p_T\Big)}{p_T^\mu}; \tag{7.1}$$

  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;
  - $\sum_{h^\pm_{\text{PU}}} p_T$ 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 $-0.5 \sum_{h^\pm_{\text{PU}}} p_T$is the so-called "$\Delta\beta$" correction for contamination from PU neutral particles[III].

- *muoSegmComp*: a measure of the track's compatibility with the muon hypothesis;
- *muoChi2LM*: comparison between the momenta of the tracker and muon detector tracks, extrapolated at the muon detector innermost surface;
- *muoChi2LP*: comparison between the extrapolated positions of the tracker and muon detector tracks at the muon detector innermost surface;
- *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 standalone 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;

[III] 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 removed more easily, as they are not associated with the PV.

- *muoLWH*: amount of tracker layers with hits;
- *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;
- $muoVMuonHitComb = vDT/2 + vCSC_{\text{max}} + vRPC$, where vDT is the number of valid hits in the drift tubes, $vCSC_{\text{max}}$ is the number of valid hits in the Cathode Strip Chambers limited to a maximum of 6 hits per station and $vRPC$ is the number of valid hits in the Resistive Plate Chambers;
- *muoTrkKink*: comparison of the tracker hits position with the interpolated trajectory of the inner track (kink);
- $muoGlbKinkFinderLOG = \log(2 + glbKinkFinder)$ where $glbKinkFinder$ is 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 distributions of the input variables are shown in Fig. 7.2 and Fig. 7.3.

MODEL TRAINING AND PERFORMANCE The DNN model is constructed as three fully connected layers, each with 200 neurons and `ReLU` as activation function, each followed by a `Dropout` layer with a dropout probability equal to 20%. The `ReLU` (*REctified Linear Unit*, [123]) activation function is defined as

$$\texttt{ReLU}(x) = \max(0, x)\,. \tag{7.2}$$

Training is carried out in TMVA with the options reported in Tab. 7.3. The output distribution of the trained model is shown in Fig. 7.4. No noticeable overtraining is observed.

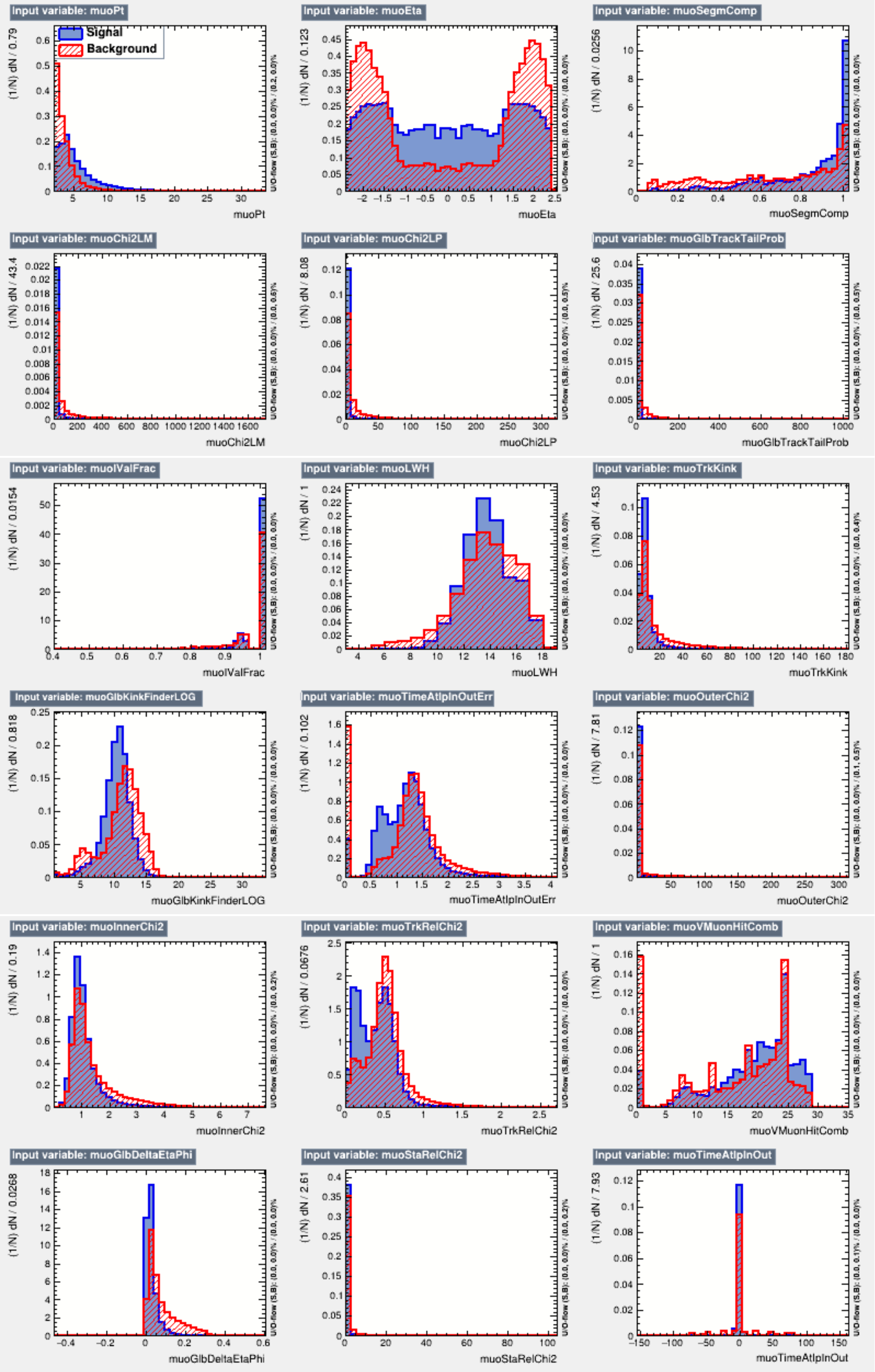


Figure 7.2: *DALIHA_UL* input features distribution, for Signal (blue) and Background (red). Part 1.

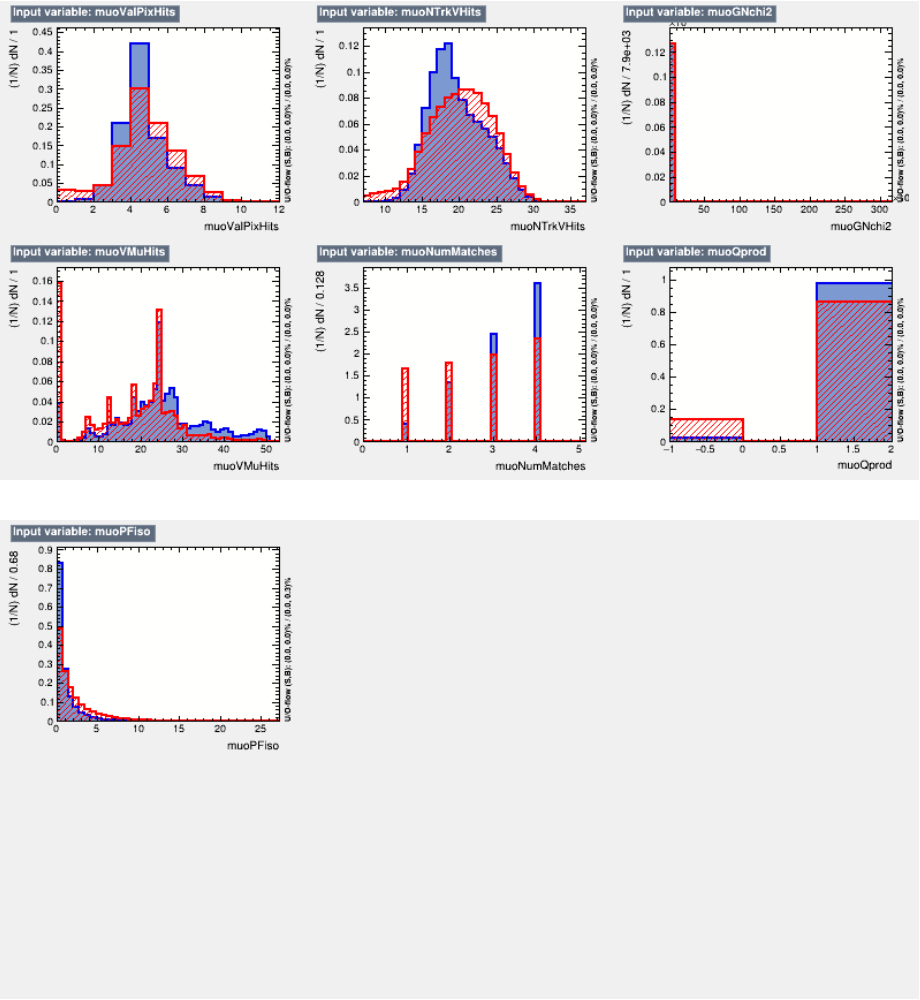


Figure 7.3: *DALIHA_UL* input features distribution, for Signal (blue) and Background (red). Part 2.

Table 7.3: Options used for training the *DALIHA_UL* DNN model. The technical description of these parameters can be found in the TMVA User Guide[124].

Option/hyperparameter	Value
Optimiser	Adam with $lr = 0.001$
Loss	Categorical cross-entropy
Batch size	1024
Signal weight	0.5
Background weight	1.0
N. of training epochs	100
N. of epochs before early stopping	10
Feature preprocessing	Norm. → Gauss. → Norm.

7.2 DNN OS-MUON TAGGER (*DEEPMUONTAGGER*)

The evaluation of the mistag probability for each event greatly improves the tagging performance, as described in Section 6.1.1. In this work, the mistag probability of the OS-muon tagger is predicted with two DNNs, one for *dataset_JpsiMuon* and one for *dataset_JpsiTrkTrk*, hereafter called "*DeepMuonTagger_JpsiMuon*" and "'*DeepMuonTagger_JpsiTrkTrk*'. The models have been developed using the Keras machine learning library within the TMVA toolkit and use 14 input features related to muon kinematics and surrounding activity.

7.2.1 *Training data set*

The two NNs are trained in $B^0_s \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-$ events simulated under the 2017 and 2018 data-taking conditions, mixed randomly. In all events, the B^0_s meson is identified using the MC truth and selected with requirements on the generated momenta. 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^0_s meson is required to have $p_T > 10$ GeV and $|\eta| < 2.5$. This selection, summarized in Tab. 7.4, ensures that reconstructed b-mesons are in the same kinematic region as the B^0_s candidates used in the CPV analysis. Training candidates are selected with the same requirements listed in Tab. 7.1. A total of 640k (140k) OS-muon candidates are selected for training for *DeepMuonTagger_JpsiMuon* (*DeepMuonTagger_JpsiTrkTrk*), out of which 190k (50k) are muons that carry the wrong flavor information. The data set is then further split into *training*, *validation* and *test* subsets with respective fractions of 60%, 20% and 20%. The training data set is used to train the DNN models, the validation data set is used to evaluate performance at each training iteration (also called "epoch") to control overfitting and trigger early stopping, while the test

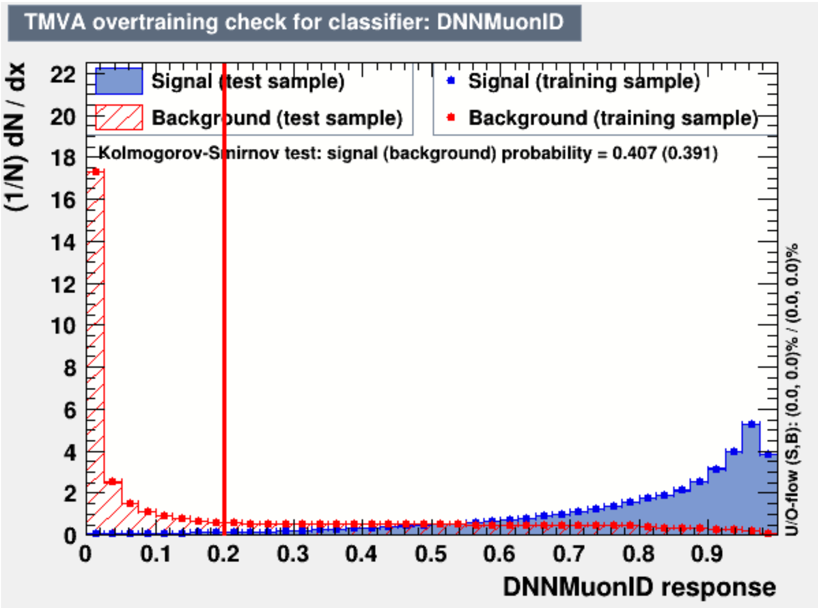


Figure 7.4: `DALIHA_UL` output distribution for Signal (blue) and Background (red). The solid red line represents the used working point to select the OS-muon candidate. No significant difference is observed between the solid lines (training sample) and the dots (test sample), showing the absence of overtraining.

data set is used to evaluate performance and overtraining after the training is finished.

7.2.2 *Input features*

The two NNs combine several input features in order to discriminate right-tag muon from wrong-tag ones and infer the mistag probability. A total of 14 variables are chosen due to their discrimination power against various dilution sources.

- Kinematics:

Table 7.4: Additional B_s^0 selection requirements for the *DeepMuonTagger* training data set. All requirements are applied to the generated quantities.

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

  - p_T: muon transverse momentum;
  - η: muon pseudorapidity.

  Information on the muon candidate kinematics helps to discriminate kaons and pions misidentified as muons. 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. Since these variables do not carry flavor-related information, they are more helpful in identifying muons from dilution sources with $\omega_{\text{evt}} \approx 50\%$.

- Hadron discriminator:
  - s_{muonID}: output score of the *DALIHA_UL* 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 ($\omega_{\text{evt}} \approx 50\%$).

- Impact parameter with respect to the PV:
  - d_{xy}: signed muon impact parameter in the transverse plane;
  - d_{z}: muon longitudinal impact parameter;
  - $\sigma(d_{\text{xy}})$: uncertainty on d_{xy};
  - $\sigma(d_{\text{z}})$: 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. Furthermore, muons associated with small impact parameters are more likely to have the correct flavor tag if they originate from OS B^0 mesons, as this species is subject to time-dependent flavor mixing. Although 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. In combination with other variables, d_z can also help identify 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 for inclusion in a multivariate method. The uncertainty on the impact parameter is also given to improve its modelling.

- Separation from the reconstructed b-meson:

- $\Delta R_{\eta,\phi}(\mu, B)$: distance in the η-ϕ plane between the muon and the reconstructed b-meson momenta.

The angular separation between the tagging muon candidate and the reconstructed b-meson is used in the selection as a veto for the same-side activity and can be used by *DeepMuonTagger* to construct more complex variables that relate the muon features to the B_s^0 flavor.

- Isolation:
  - Iso_μ: muon isolation, defined as in Eq. 7.1.

  Muon 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 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 muon surrounding activity, a jet-like object (hereafter “muon cone”) is constructed from all 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[IV]. Unless stated otherwise, the tagging muon is assumed to be part of the cone, if reconstructed by the PF algorithm (not all global muons are). The following variables are defined by combining the muon and muon cone properties:
  - $p_T^{\text{cone}} = \mathtt{Proj}_{xy}(\sum_{i\in\text{cone}} \vec{p}_i)$: transverse component of the cone axis[V],
  - $p_T^{\text{rel}} = \vec{p}_T^{\,\mu} \cdot (\vec{p}^{\,\text{cone}} - \vec{p}^{\,\mu}) / |\vec{p}^{\,\text{cone}} - \vec{p}^{\,\mu}|$: muon $\vec{p}_T$ projected on the cone axis, after subtracting the muon momentum,
  - $\Delta R_{\eta,\phi}(\mu, \text{cone})$: distance in the η-ϕ plane between the muon momentum direction and the cone axis,
  - E_μ / E_{cone}: ratio of the muon energy to the cone energy[VI].

  All of these variables have been shown to help separate direct semileptonic decays ($b \to \mu^- X$) from cascade decays ($b \to cX \to \mu^+ X'$) [125, 126].

- Cone charge:

$$Q_{\text{cone}} = Q_\mu \frac{\sum_i^{\text{cone}} p_{T,i} \cdot Q_i}{\sum_i^{\text{cone}} p_{T,i}}, \tag{7.3}$$

  where Q_μ is the charge of the muon, used as a multiplicative factor to remove the dependence of Q_{cone} on the tagging decision[VII]. The cone charge is

---

[IV] A *home-made* cone is used instead of jet objects, since these are not reconstructed in the kinematic region of interest.

[V] The cone axis is defined as the direction of the total momentum of the cone.

[VI] Since the tagging muon is not removed from the cone, E_μ/E_{cone} values of one are possible.

[VII] Reminder: the muon tagger DNN task is to discriminate right- and wrong-tag, not identifying the flavor of the OS b-hadron.

constructed following the same general strategy as the jet-tagging techniques and it helps identify the OS b-hadron flavor.

The distributions of the input features are shown in Fig. 7.5 and Fig. 7.6 for *DeepMuonTagger_JpsiMuon* and *DeepMuonTagger_JpsiTrkTrk*, respectively.

FEATURE PREPROCESSING The input features are preprocessed to increase the performance of the DNN methods. The variables are first normalized in the range $[-1, 1]$, then their distribution is transformed into a Gaussian ("Gaussianisation")[VIII] 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 regards all features to a similar extent without favoritism. Gaussianization is found to speed up learning and improve convergence properties when combined with a Gaussian random initialization of the model weights. The preprocessing procedures are handled with the TMVA toolkit.

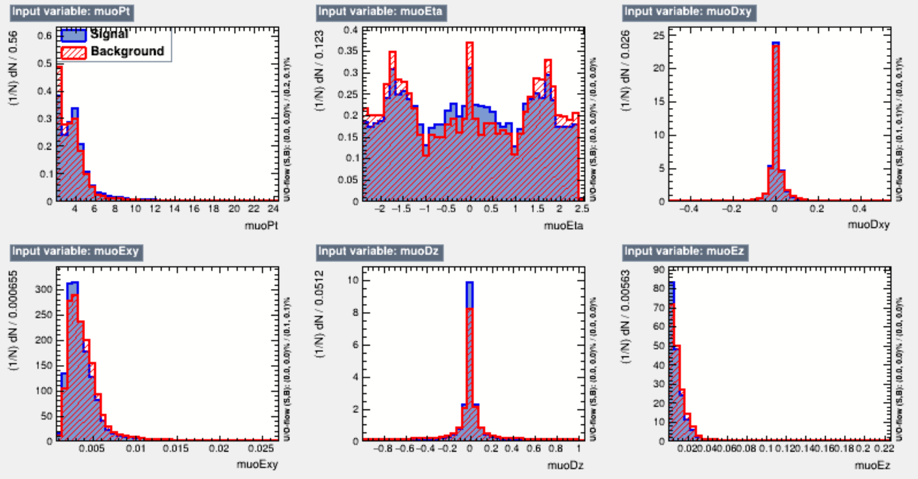
7.2.3 *Model structure and training*

Both *DeepMuonTagger* models are constructed with fully connected layers interlaced with `Dropout` layers to reduce overtraining. To construct the network with the best right- versus wrong-tag separation, several hyperparameters related to the DNN structure are optimized with a grid scan: number of fully connected layers, neuron for each layer (layer width), dropout probability, and batch size. The parameter configuration that maximizes the final tagging power and minimizes the normalized χ^2 of the calibration fit is chosen. After the optimization through grid scan, the only hyperparameter difference between the two models is the training batch size. The best parameter configuration chosen for the DNN is reported in Tab. 7.5, while other training options are reported in Tab. 7.6.

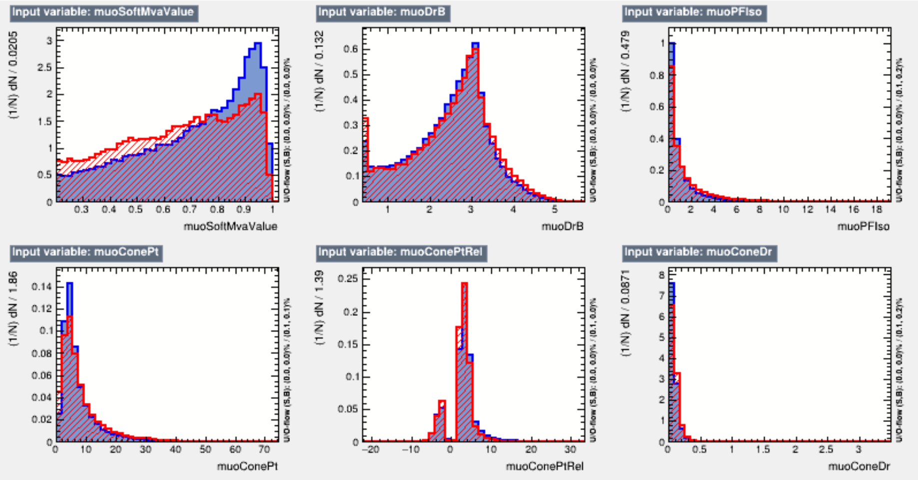
Table 7.5: Hyperparameters that define the structure of the *DeepMuonTagger* models.

Hyperparameter	`HLT_JpsiMu`	`HLT_JpsiTrkTrk`
N. of full conn. layers	2	2
Width of each layer	300	300
Dropout prob.	0.5	0.5
Batch size	512	256

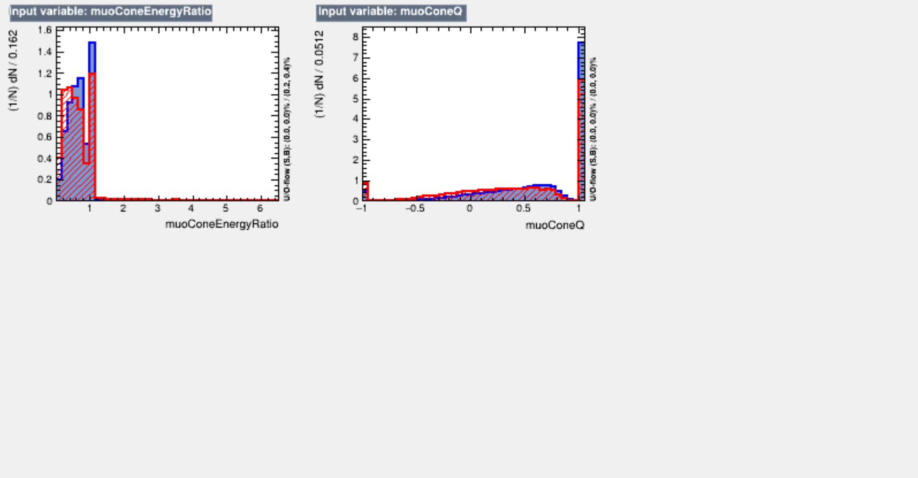
[VIII] Gaussianisation proceeds in two steps: first the input distribution is transformed into a uniform distribution using its cumulative distribution function. Next, the inverse error function is used to transform the uniform distribution into a Gaussian with zero mean and unity width. For more details, see the TMVA Users Guide[114].



(a) From left to right, top to bottom: p_T, η, d_{xy}, $\sigma(d_{xy})$, d_z and $\sigma(d_z)$.

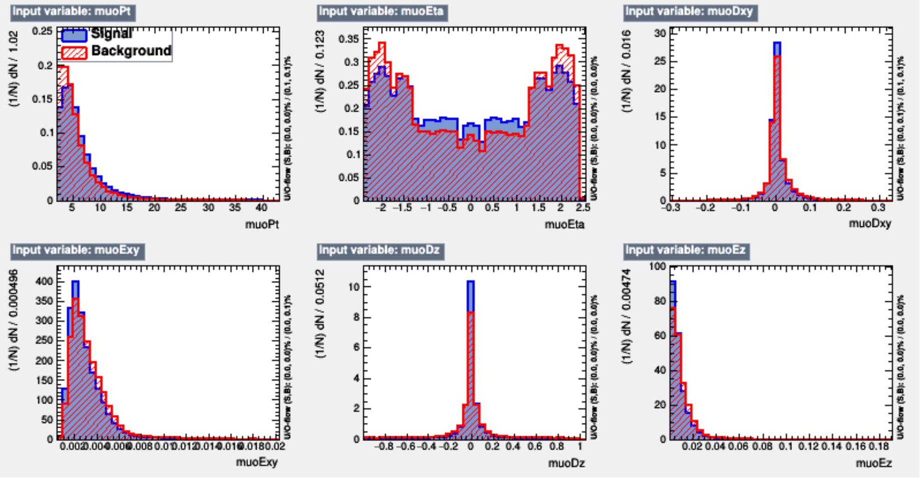


(b) From left to right, top to bottom: s_{muonID}, $\Delta R_{\eta,\phi}(\mu, B)$, Iso_{μ}, p_T^{cone}, p_T^{rel} and $\Delta R_{\eta,\phi}(\mu, \text{cone})$.

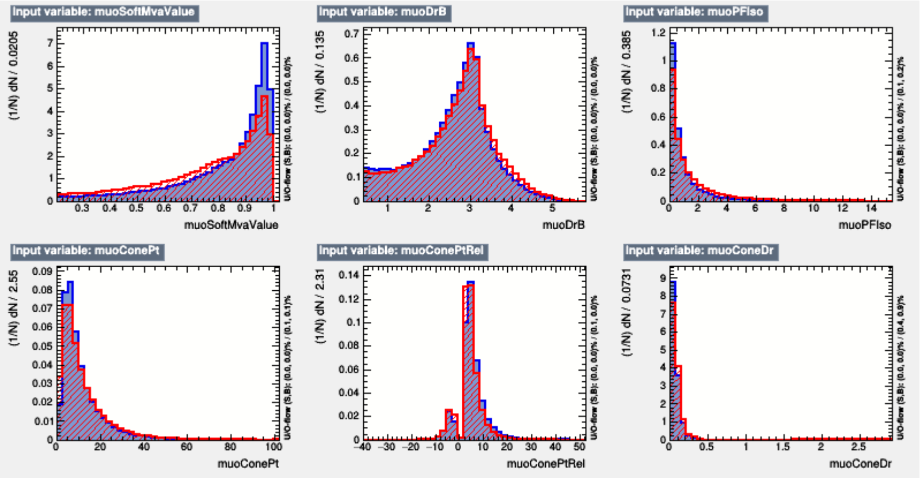


(c) From left to right: E_{μ}/E_{cone} and Q_{cone}.

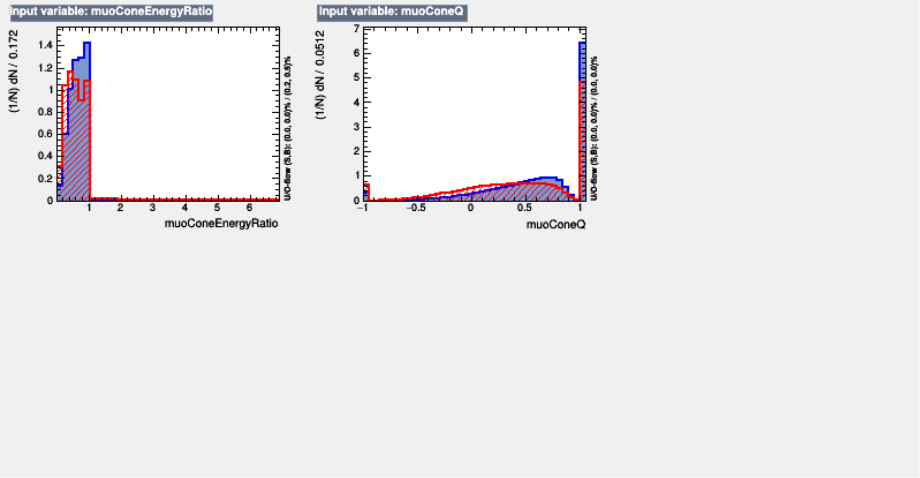
Figure 7.5: *DeepMuonTagger_JpsiMuon* input features distributions for right- (blue) and wrong-tag (red) OS-muon candidates. The distributions are normalised. The TMVA plotting tool automatically calls class-1 (in this case right-tag) "Signal" and class-2 (in this case wrong-tag) "Background".



(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 $\Delta R_{\eta,\phi}(\mu, \text{cone})$.



(c) From left to right: E_{μ}/E_{cone} and Q_{cone}.

Figure 7.6: *DeepMuonTagger_JpsiTrkTrk* input features distributions for right- (blue) and wrong-tag (red) tagging muon candidates. The distributions are normalised. The TMVA plotting tool automatically calls class-1 (in this case right-tag) "Signal" and class-2 (in this case wrong-tag) "Background".

Table 7.6: Training options for the *DeepMuonTagger* models.

Option/hyperparameter	Value
Optimiser	Adam, with $lr = 0.001$
Loss	Categorical cross-entropy
Metric	Accuracy
N. of training epochs	50
N. of epochs before early stopping	10
Feature preprocessing	Norm. → Gauss. → Norm.

7.2.4 *Model performance*

The performance of trained models is evaluated in the test data subset, which is statistically independent of the training and validation samples. The output score distributions are shown in Fig. 7.7. No signs of severe overfitting are observed when comparing the output distributions obtained from the training and test 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. 7.7 can still be qualitatively commented upon, remembering that the output score s_{DNN} is equal to $1 - \omega_{\text{evt}}$ (Eq. 6.28). The network is able to partially identify tagging muons that do not contain flavor information, such as PU and fake muons, by clustering them around $\omega_{\text{evt}} \approx 50\%$. This is a very helpful feature, since the flavor inference on these events can now be ignored in the final fit of the CPV analysis. 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, since 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. The small fraction of events with $\omega_{\text{evt}} < 15\%$ are likely well-reconstructed muons with a small transverse impact parameter, which is correlated to lower mixing probabilities, or muons from decays of charged b-hadrons, which do not oscillate.

In summary, *DeepMuonTagger* methods are able to partially separate the various dilution sources by exploiting their different properties. It is also capable to directly predict the right-tag probability s_{DNN} and, consequently, the mistag probability as

$$\omega_{\text{evt}} = 1 - s_{\text{DNN}} \,. \tag{7.4}$$

The predicted mistag probability can finally be used to reweight the flavor inference and better exploit the data.

7.2.5 *Model calibration*

The predicted mistag probability ω_{evt} is calibrated in data using self-tagging $B^+ \to J/\psi K^+$ decays, following the procedure described in Section 6.6. Calibration is performed separately for the 2017 and 2018 *dataset_JpsiMuon* and *dataset_JpsiTrkTrk* data sets. The results are shown in Fig. 7.8 and Tab. 7.7.

The calibration of the DNN methods is also verified with the same procedure using independent samples of simulated $B^0_s \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ decays, where the reconstructed B^0_s and B^+ mesons are matched to the generated ones to find their true flavor at production time. The results are shown in Fig. 7.9, 7.10 and Tab. 7.8.

In general, the measured mistag probability is accurately predicted by *DeepMuonTagger* over the entire range for all the simulated and data samples examined. In all cases, more than 90% of the tagged events fall in the $\omega_{\text{evt}} = [0.1,\ 0.5]$ interval. Residual differences are well approximated by the calibration function.

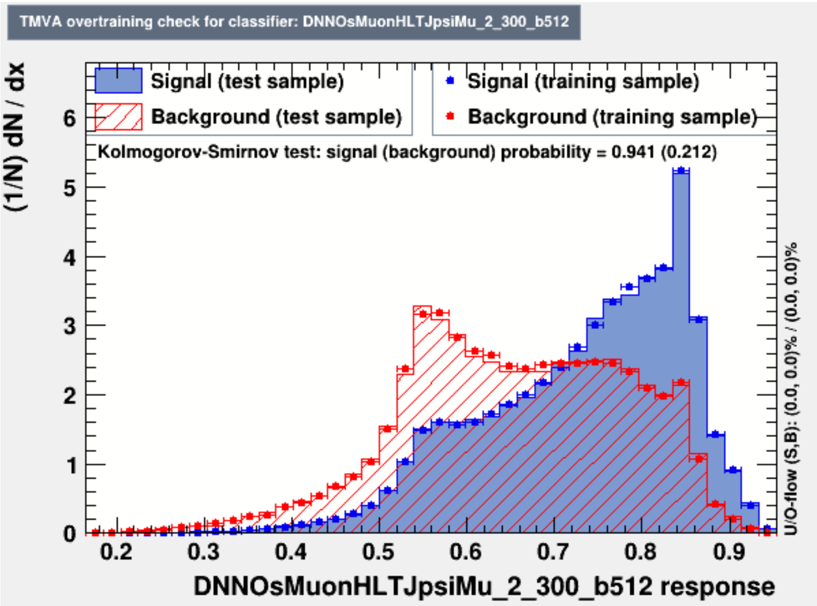
Differences in mistag probability calibration between the $B^0_s \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ samples and the statistical uncertainties in calibration fits are considered systematic uncertainties in the CPV analysis and are discussed in Chapter 14.

Table 7.7: Result of the mistag probability calibration fit for *DeepMuonTagger* on $B^+ \to J/\psi K^+$ decays from data. The fitted model is $f(\omega_{\text{evt}}) = \texttt{sigmoid}(a + b \cdot \texttt{logit}(\omega_{\text{evt}}))$ as defined in Eq. 6.35.

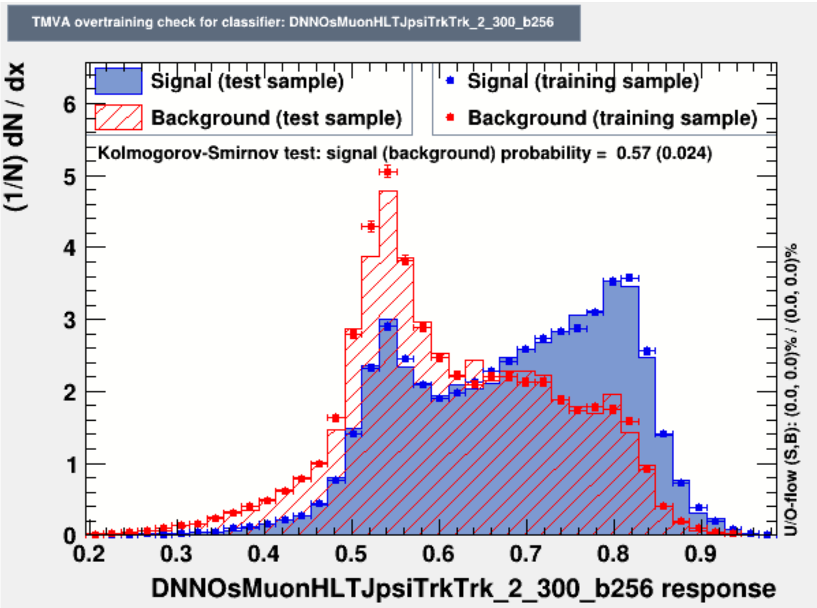
Trigger	Year (Data)	a	b	χ^2/d.o.f.
`HLT_JpsiMu`	2017	0.072 ± 0.012	1.012 ± 0.012	1.8
	2018	0.0986 ± 0.0094	1.0347 ± 0.0093	2.0
`HLT_JpsiTrkTrk`	2017	0.047 ± 0.026	0.956 ± 0.031	1.3
	2018	0.038 ± 0.016	0.951 ± 0.019	1.6

7.3 TAGGING PERFORMANCE

The tagging performance of the OS-muon tagger is evaluated on the $B^+ \to J/\psi K^+$ data samples used also to calibrate the probability of mistag using Eq. 6.37 and Eq. 6.38. The final tagging performance measured in the $B^+ \to J/\psi K^+$ data sample is reported in Tab. 7.9, where the relative statistics of the various data sets are taken into account.

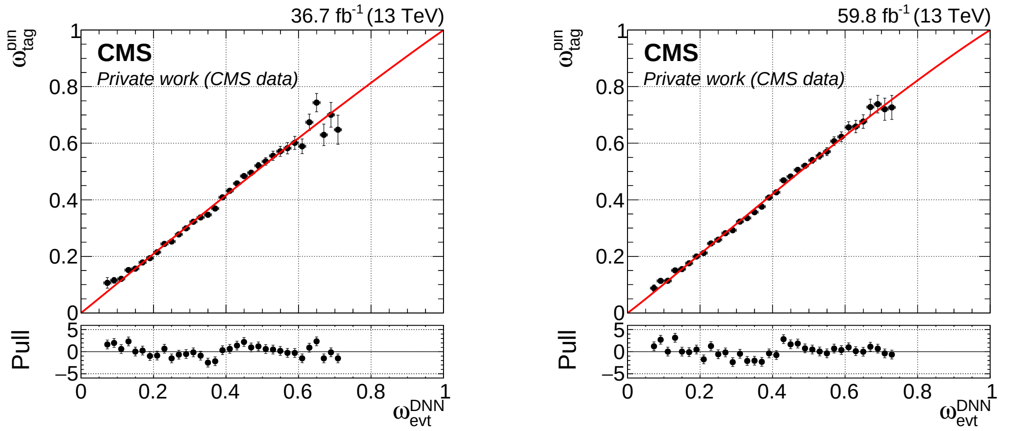


(a) *DeepMuonTagger_JpsiMuon* normalised output score.



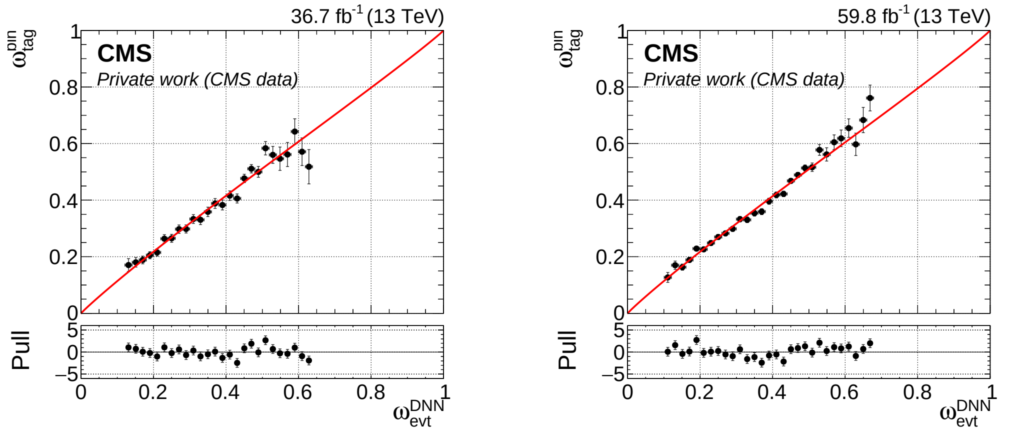
(b) *DeepMuonTagger_JpsiTrkTrk* normalised output score.

Figure 7.7: *DeepMuonTagger* normalised DNN output score for right- (blue) and wrong-tag (red) OS-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".



(a) *DeepMuonTagger_JpsiMuon* calibration plot for the 2017 data sample.

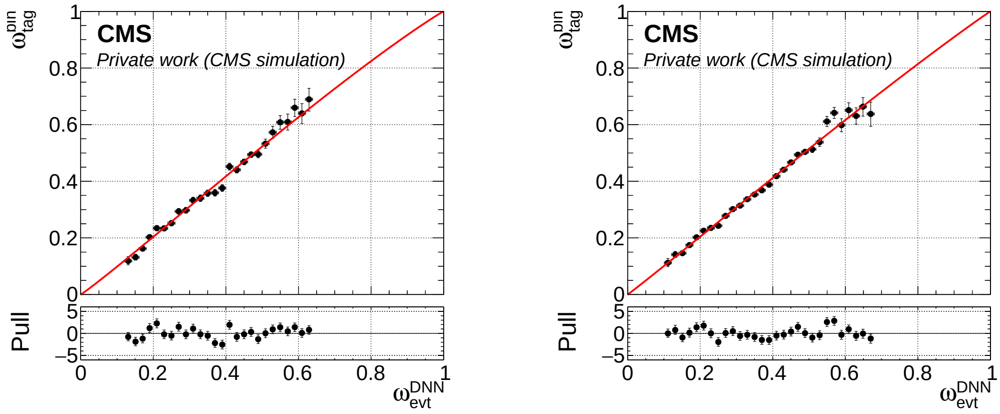
(b) *DeepMuonTagger_JpsiMuon* calibration plot for the 2018 data sample.



(c) *DeepMuonTagger_JpsiTrkTrk* calibration plot for the 2017 data sample.

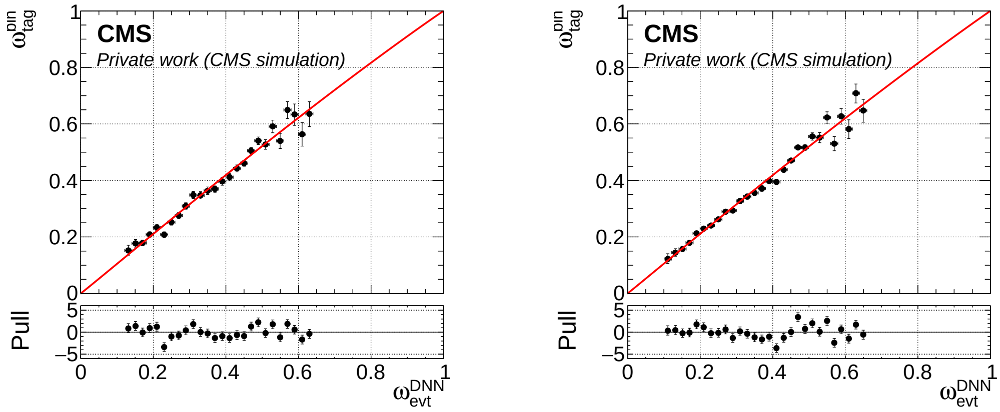
(d) *DeepMuonTagger_JpsiTrkTrk* calibration plot for the 2018 data sample.

Figure 7.8: Result of the mistag probability calibration fit for *DeepMuonTagger* on $B^+ \to J/\psi K^+$ decays from data. The solid red line shows the calibration fit to data (solid markers).



(a) *DeepMuonTagger_JpsiTrkTrk* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2017 data-taking conditions).

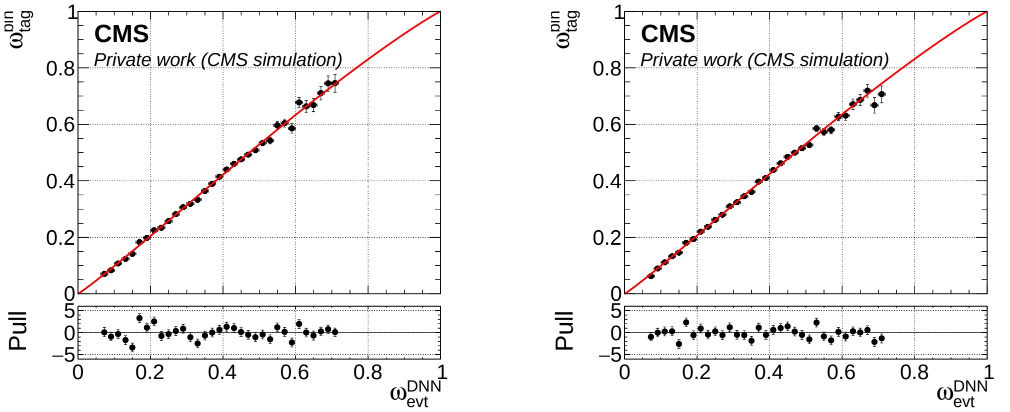
(b) *DeepMuonTagger_JpsiTrkTrk* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2018 data-taking conditions).



(c) *DeepMuonTagger_JpsiTrkTrk* calibration plot for $B^+ \to J/\psi\,K^+$ simulated events (2017 data-taking conditions).

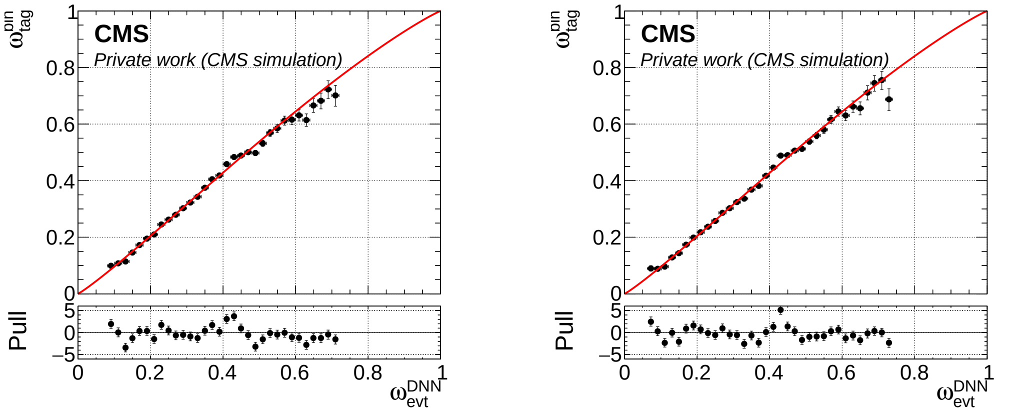
(d) *DeepMuonTagger_JpsiTrkTrk* calibration plot for $B^+ \to J/\psi\,K^+$ simulated events (2018 data-taking conditions).

Figure 7.9: Result of the mistag probability calibration fit for *DeepMuonTagger* on simulated samples. The solid red line shows the calibration fit to data (solid markers).



(a) *DeepMuonTagger_JpsiMuon* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2017 data-taking conditions).

(b) *DeepMuonTagger_JpsiMuon* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2018 data-taking conditions).



(c) *DeepMuonTagger_JpsiMuon* calibration plot for $B^+ \to J/\psi\,K^+$ simulated events (2017 data-taking conditions).

(d) *DeepMuonTagger_JpsiMuon* calibration plot for $B^+ \to J/\psi\,K^+$ simulated events (2018 data-taking conditions).

Figure 7.10: Result of the mistag probability calibration fit for *DeepMuonTagger* on simulated samples. The solid red line shows the calibration fit to data (solid markers).

Table 7.8: Result of the mistag probability calibration fit for *DeepMuonTagger* on simulated samples. The fitted model is $f(\omega_{\text{evt}}) = \texttt{sigmoid}(a + b \cdot \texttt{logit}(\omega_{\text{evt}}))$ as defined in Eq. 6.35.

Trigger	Process	Year	a	b	χ^2/d.o.f.
`HLT_JpsiMu`	$B_s^0 \to J/\psi\,\phi$	2017	0.1153 ± 0.0089	1.0618 ± 0.0088	2.0
		2018	0.1246 ± 0.0080	1.0611 ± 0.0079	1.5
	$B^+ \to J/\psi\,K^+$	2017	0.1533 ± 0.0094	1.0902 ± 0.0094	2.8
		2018	0.1508 ± 0.0086	1.0900 ± 0.0085	2.6
`HLT_JpsiTrkTrk`	$B_s^0 \to J/\psi\,\phi$	2017	0.091 ± 0.018	1.048 ± 0.022	1.6
		2018	0.058 ± 0.012	1.023 ± 0.015	1.4
	$B^+ \to J/\psi\,K^+$	2017	0.085 ± 0.019	1.015 ± 0.019	1.9
		2018	0.081 ± 0.014	1.010 ± 0.017	2.3

When taking into account the relative statistics of the various data sets[IX], the total tagging performance for the OS-muon tagger are $\hat{P}_{\text{tag}} = (1.535 \pm 0.006)\%$ with $\varepsilon_{\text{tag}} = (7.47 \pm 0.03)\%$ and $\widehat{\mathcal{D}}^2_{\text{tag}} = 0.205$.

Table 7.9: Calibrated opposite-side muon tagger performance evaluated in $B^+ \to J/\psi\,K^+$ events in the 2017 and 2018 data samples. The tagging efficiency ε_{tag} and the total tagging power $\hat{P}_{\text{tag}}$ are computed with Eq. 6.38 and Eq. 6.37, respectively. The effective dilution $\widehat{\mathcal{D}}_{\text{tag}}$ is computed as $\widehat{\mathcal{D}}_{\text{tag}} = \hat{P}_{\text{tag}}/\varepsilon_{\text{tag}}$. *DeepMuonTagger_JpsiMuon* is used in *dataset_JpsiMuon*, while *DeepMuonTagger_JpsiTrkTrk* is used in *dataset_JpsiTrkTrk*. The quoted uncertainties are statistical only. The evaluation of the uncertainty on $\widehat{\mathcal{D}}_{\text{tag}}$ is quite complex due to the correlations between $\sigma(\varepsilon_{\text{tag}})$ and $\sigma(\hat{P}_{\text{tag}})$, and is beyond the scope of this work. The last row reports the total performance averaged by taking into account the relative statistics of the dataset.

Data sample	Year	ε_{tag} [%]	$\widehat{\mathcal{D}}^2_{\text{tag}}$	$\hat{P}_{\text{tag}}$ [%]
dataset_JpsiMuon	2017	47.2 ± 0.3	0.217	10.23 ± 0.05
	2018	52.7 ± 0.2	0.219	11.54 ± 0.05
dataset_JpsiTrkTrk	2017	1.41 ± 0.01	0.142	0.2002 ± 0.0003
	2018	1.42 ± 0.01	0.142	0.2015 ± 0.0002
Total		7.47 ± 0.03	0.205	1.535 ± 0.006

[IX] $N($*dataset_JpsiMuon*$) : N($*dataset_JpsiTrkTrk*$) \approx 1 : 9$. $N(2017) : N(2018) \approx 3 : 7$.

8

OPPOSITE-SIDE ELECTRON TAGGING

In concept, the OS-electron and OS-muon taggers are identical apart from the type of lepton used, with the three core steps being the same: OS-lepton selection, tagging decision[I], and mistag probability estimation. As such, this section will not repeat concepts already described in Chapter 7 and contains mostly technical information about the algorithm.

8.1 OS-ELECTRON SELECTION

OS-electron candidates are searched in events with a reconstructed b-meson decay and a selected PV, as described in Chapter 5, with tracks associated with the reconstructed b-meson excluded from the pool of candidates[II]. The electron tagger is applied only to events in *dataset_JpsiTrkTrk*. Since semileptonic decays to muon and electron are mutually exclusive, the event is considered untagged by the OS-electron tagger if it contains an OS-muon candidate[III].

Candidates need to have $p_T \geq 2.5$ Gev, $|\eta| < 2.4$, a separation from the reconstructed b-meson decay of $\Delta R_{\eta,\phi} \geq 0.4$, and an impact parameter with respect to the PV of $|d_z| \leq 0.2$ cm and $\left|d_{xy}\right| \leq 0.08$ cm, in the longitudinal and transverse directions, respectively. d_{xy} refers to the transverse impact parameter, where the direction of the electron mother particle is approximated with the resulting momentum direction of all the charged PF candidates within $\Delta R_{\eta\phi} = 0.4$ from the electron momentum. In addition, candidates need to pass the quality cuts dictated by the CMS electron MVA discriminator, with a minimum score value of -0.9999.[IV] The selection requirements are optimised to maximise the tagging power of the tagging algorithm and are summarised in Tab. 8.1. As for the OS-muon, in the case of multiple OS-electron candidates, the one with the highest p_T is chosen.

About $\approx 6\%$ of the events in *dataset_JpsiTrkTrk* contain an OS-electron candidate.

8.2 DNN OS-ELECTRON TAGGER (*DEEPELECTRONTAGGER*)

The mistag probability of the OS-electron tagger is predicted with a DNN, hereafter called "*DeepElectronTagger*", that is trained to discriminate right- and wrong-

---

[I] Which exploits the charge of the OS lepton, e.g.: OS $\ell^- \rightarrow$ OS $b \xrightarrow{\text{tag}}$ signal B_s^0.

[II] Since the electron track is not a standard track in CMSSW, a $\Delta R_{\eta,\phi}$ and p_T matching is performed.

[III] The muon tagger is considered a more stable and performing algorithm, as such is preferred in the case of overlap.

[IV] In practice, only the value `eleIDNIV2Val` $= -1$ is excluded.

tag. The models have been developed using the KERAS machine learning library within the TMVA toolkit, and they use 14 input features related to electron kinematics and surrounding activity.

The training procedure and model-building process are very similar to those used for *DeepMuonTagger*.

8.2.1 *Training data set*

DeepElectronTagger is trained on $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ events simulated under both 2017 and 2018 data-taking conditions, randomly mixed. As for *DeepMuonTagger*, in all events the B_s^0 meson is identified using MC truth and selected with requirements to ensure reconstructed b-mesons in the same kinematic region of the B_s^0 candidates used in the CPV analysis. Selection requirements are reported in Tab. 7.4. The training candidates are selected with the same requirements on the OS-electron listed in Tab. 8.1. A total of 500k OS-electron candidates are selected for *DeepElectronTagger*, of which 200k are electrons that carry the wrong flavor correlation. The data set is then further divided into *training*, *validation* and *test* subsets with respective fractions of 60%, 20% and 20%.

8.2.2 *Input features*

DeepElectronTagger combines the following 14 variables, chosen for their power against various dilution sources:

- p_T: electron transverse momentum;
- η: electron pseudorapidity;
- d_{xy}: signed electron impact parameter in the transverse plane;
- d_{z}: electron longitudinal impact parameter;
- $\sigma(d_{\mathrm{xy}})$: uncertainty on d_{xy};
- $\sigma(d_{\mathrm{z}})$: uncertainty on d_{z};
- `eleIDNIV2Val`: the score of the electron MVA discriminator (developed by the CMS collaboration);
- $\Delta R_{\eta,\phi}(e, B)$: distance in the η-ϕ plane between the electron and the reconstructed b-meson momenta;
- Iso_e: electron isolation, defined similarly as the muon isolation;
- *Cone variables*, computed from a jet-like object (hereafter "electron cone") constructed from all the Particle Flow candidates contained in a $\Delta R_{\eta,\phi} < 0.4$

cone, centered in the electron momentum, with $p_T > 0.5$ GeV, $|\eta| < 3$ and $d_z < 1$ cm:

- p_T^{cone}: total transverse momentum of the electron cone, evaluated as the sum of the transverse momenta of its components,
- $p_T^{\text{rel}} = \vec{p}_T^{\,e} \cdot (\vec{p}^{\,\text{cone}} - \vec{p}^{\,e})/|\vec{p}^{\,\text{cone}} - \vec{p}^{\,e}|$: electron $\vec{p}_T$ projected on the cone axis, after subtracting the electron momentum;
- $\Delta R_{\eta,\phi}(e, \text{cone})$: distance in the η-ϕ plane between the direction of the electron momentum and the cone axis;
- E_e/E_{cone}: ratio of the electron energy to the cone energy;
- $Q_{\text{cone}} = Q_e \cdot \sum_i^{\text{cone}} p_{T,i} Q_i / \sum_i^{\text{cone}} p_{T,i}$: the cone charge, where Q_e, the electron charge is used as a multiplicative factor to remove the dependence of Q_{cone} on the tagging decision.

The properties of these variables are described in Section 7.2.2, the corresponding section for *DeepMuonTagger*. The distributions of the input features are shown in Fig. 8.1.

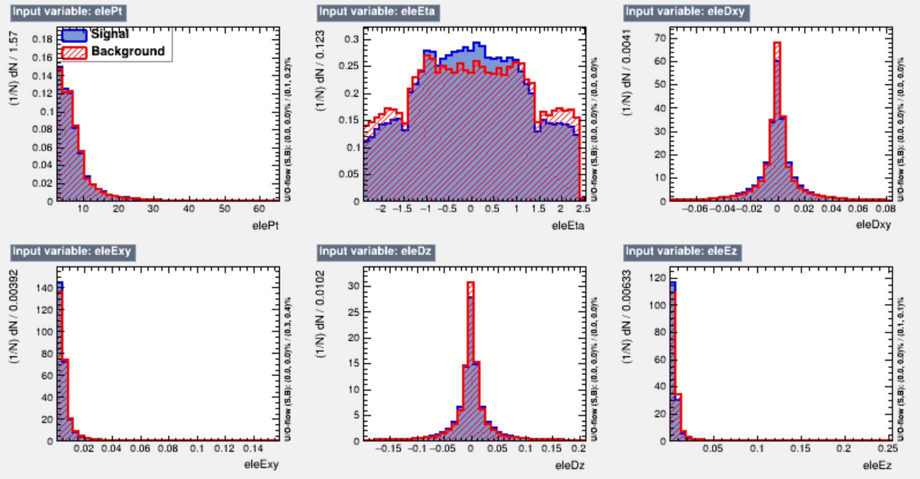
8.2.3 *Model structure and training*

The *DeepElectronTagger* models are constructed with fully connected `Dense` layers interlaced with `Dropout` layers to reduce overtraining and improve sparsity. To construct the network with the best right- vs wrong-tag separation, several hyperparameters related to the DNN structure are optimized with a grid scan: number of `Dense` layers, neuron for each layer, dropout probability, and batch size. The parameter configuration that maximizes the final tagging power and minimizes the normalized χ^2 of the calibration fit is chosen. The best parameter configuration chosen for the DNN is reported in Tab. 8.2, while other options relevant to the training are reported in Tab. 8.3.

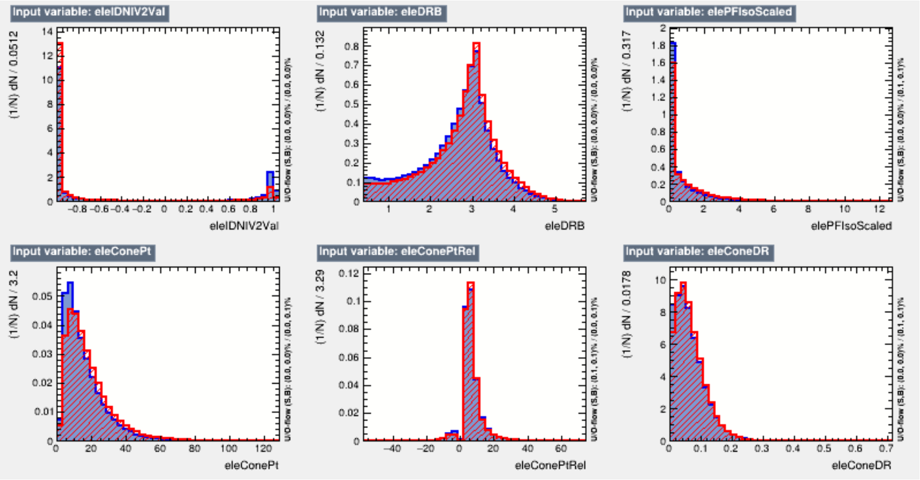
8.2.4 *Model performance*

The performance of trained models is evaluated in the test data subset, which is statistically independent of the training and validation samples. The output score distributions are shown in Fig. 8.2. No signs of overfitting are observed when comparing the output distributions obtained from the training and test data sets. The network score s_{DNN} is equivalent to the right-tag probability, therefore the mistag probability ω_{evt} is equal to

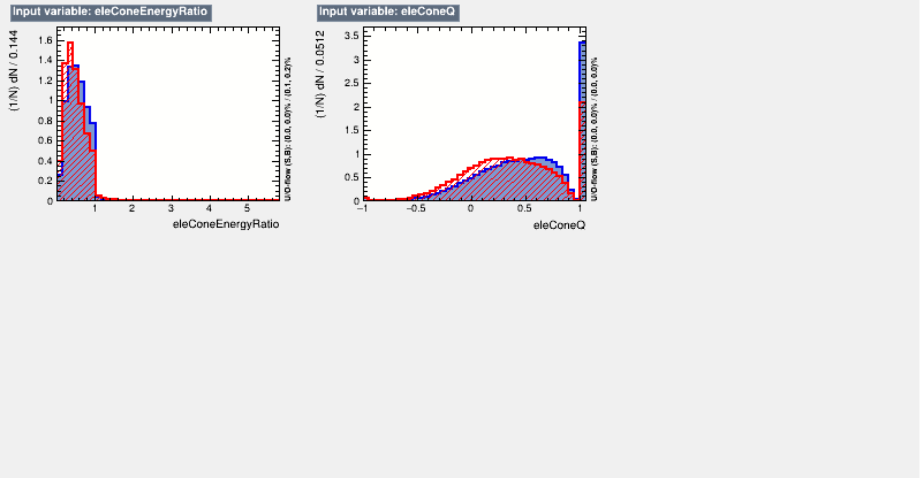
$$\omega_{\text{evt}} = 1 - s_{\text{DNN}} \,. \tag{8.1}$$



(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: `eleIDNIV2Val`, $\Delta R_{\eta,\phi}(e,B)$, Iso$_e$, p_T^{cone}, p_T^{rel} and $\Delta R_{\eta,\phi}(w,\text{cone})$.



(c) From left to right: E_μ/E_{cone} and Q_{cone}.

Figure 8.1: *DeepElectronTagger* input features distributions for right- (blue) and wrong-tag (red) OS-electron candidates. The distributions are normalised. The TMVA plotting tool automatically calls class-1 (in this case right-tag) "Signal" and class-2 (in this case wrong-tag) "Background".

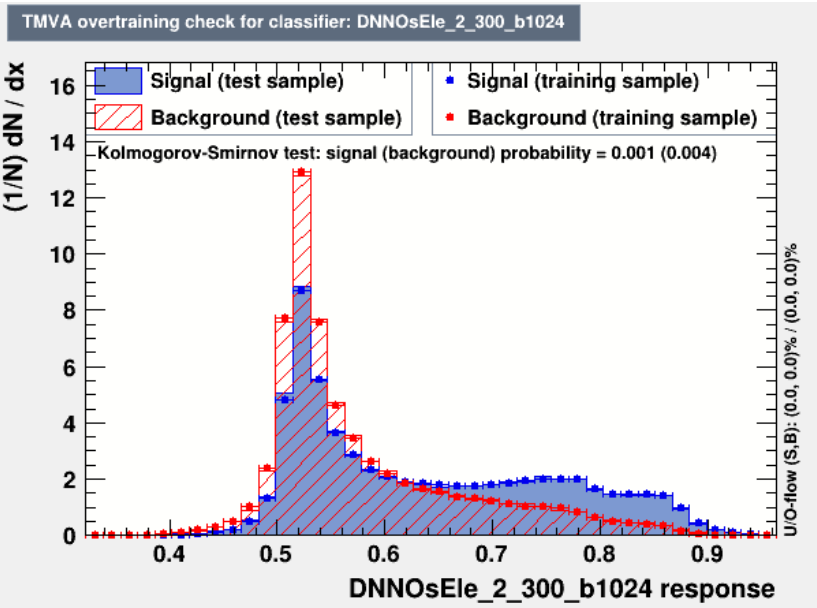


Figure 8.2: *DeepElectronTagger* normalised DNN output score for right- (blue) and wrong-tag (red) OS-electron for the training (markers) and in the test (histograms) data sets. The TMVA plotting tool automatically calls class-1 (in this case right-tag) “Signal” and class-2 (in this case wrong-tag) “Background”.

As for *DeepMuonTagger*, also *DeepElectronTagger* is able to partially identify tagging electrons that do not carry flavor information by clustering them around $\omega_{\text{evt}} \approx 50\%$.

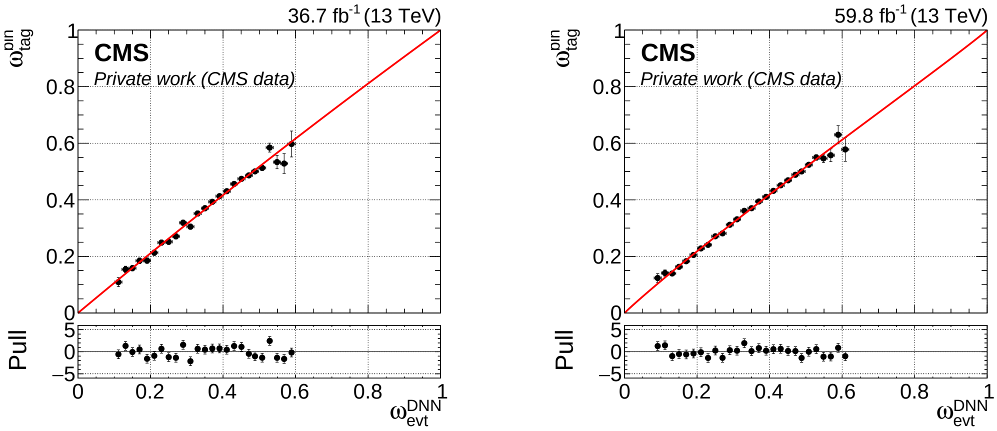
8.2.5 *Model calibration*

The predicted mistag probability ω_{evt} is calibrated in data using self-tagging $B^+ \to J/\psi K^+$ decays, following the procedure described in Section 6.6. The calibration is performed separately for the 2017 and 2018 data samples. The results are shown in Fig. 8.3 and Tab. 8.4.

The calibration of the DNN methods is also verified with the same procedure using independent samples of simulated $B^0_s \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ decays, where the reconstructed B^0_s and B^+ mesons are matched to the generated ones to find their true flavor at production time. The results are shown in Fig. 8.4 and Tab. 8.5.

In general, the measured mistag probability is accurately predicted by *DeepElectronTagger* throughout the entire range for all the simulated and data samples examined. Residual differences are well approximated by the calibration function.

Differences in the mistag probability calibration between the $B^0_s \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ samples and statistical uncertainties in the calibration fits are



(a) *DeepElectronTagger* calibration plot for the 2017 data sample.

(b) *DeepElectronTagger* calibration plot for the 2018 data sample.

Figure 8.3: Result of the mistag probability calibration fit for *DeepElectronTagger* on $B^+ \to J/\psi K^+$ decays from data. The solid red line shows the calibration fit to data (solid markers).

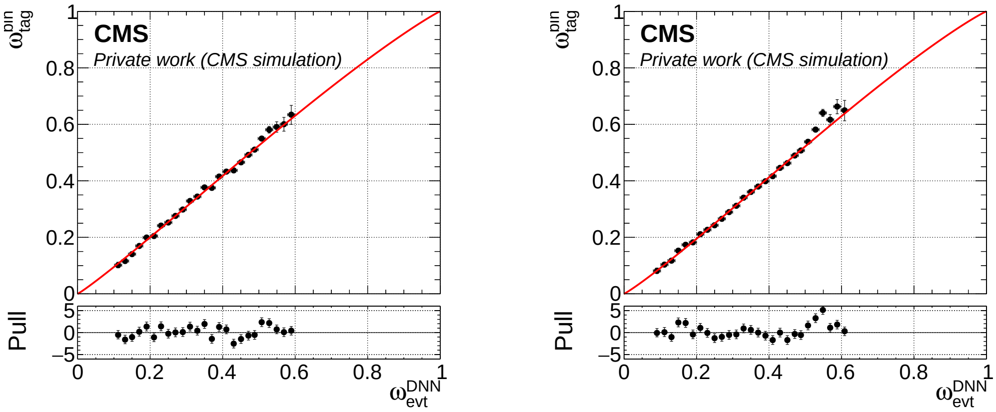
considered as systematic uncertainties in the CPV analysis and discussed in Chapter 14.

8.3 TAGGING PERFORMANCE

The tagging performance of the OS-electron tagger is evaluated on the same $B^+ \to J/\psi K^+$ data samples used for calibrating the mistag probability and with the same procedure described in Section 7.3 for the OS-muon tagger. The final tagging performance measured in the $B^+ \to J/\psi K^+$ data sample is reported in Tab. 8.6.

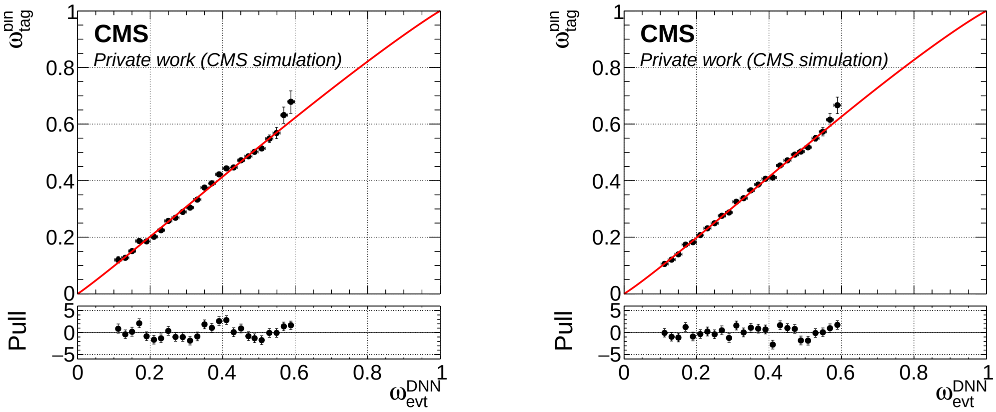
When taking into account the relative statistics of the various data sets and the fact that events in *dataset_JpsiMuon* are considered untagged (in practice the performance is set to $\varepsilon_{\text{tag}} \equiv 0\%$ and $\hat{P}_{\text{tag}} \equiv 0\%$), the total tagging performance for the OS-electron tagger is $\hat{P}_{\text{tag}} = (0.4096 \pm 0.0003)\%$ with $\varepsilon_{\text{tag}} = (5.25 \pm 0.01)\%$ and $\widehat{\mathcal{D}}^2_{\text{tag}} = 0.078$.

The resulting performance is approximately one-third of that of the OS-muon tagger. This is expected, as CMS is more suited to reconstruct and measure muons, and the usage of *dataset_JpsiMuon* triggered by `HLT_JpsiMu` greatly improves the tagging efficiency of muon techniques. Combined, the two lepton taggers have an efficiency of $\approx 13\%$ for a tagging power of $\approx 2\%$.



(a) *DeepElectronTagger* calibration plot for $B_s^0 \rightarrow J/\psi\,\phi(1020)$ simulated events (2017 data-taking conditions).

(b) *DeepElectronTagger* calibration plot for $B_s^0 \rightarrow J/\psi\,\phi(1020)$ simulated events (2018 data-taking conditions).



(c) *DeepElectronTagger* calibration plot for $B^+ \rightarrow J/\psi\,K^+$ simulated events (2017 data-taking conditions).

(d) *DeepElectronTagger* calibration plot for $B^+ \rightarrow J/\psi\,K^+$ simulated events (2018 data-taking conditions).

Figure 8.4: Result of the mistag probability calibration fit for *DeepElectronTagger* on simulated samples. The solid red line shows the calibration fit to data (solid markers).

Table 8.1: Selection requirements for OS-electron candidates. The term "Event veto" implies that the while event is discarded.

Variable	Requirement
Reconstructed b-meson tracks	Excluded
OS-muon candidate	Event veto
p_T	≥ 2.5 GeV
$\lvert\eta\rvert$	≤ 2.4
$\lvert d_z\rvert$ w.r.t. PV	≤ 0.2 cm
$\lvert d_{xy}\rvert$ w.r.t. PV	≤ 0.08 cm
$\Delta R_{\eta,\phi}(b\text{-meson})$	≥ 0.4
`eleIDNIV2Val`	> -0.9999

Table 8.2: Hyperparameters that define the structure of the *DeepElectronTagger* model.

Hyperparameter	Value
N. of full conn. layers	2
With of each layer	300
Dropout probability	0.5
Batch size	1024

Table 8.3: Training options for the *DeepElectronTagger* model.

Option/hyperparameter	Value
Optimiser	Adam, with $lr = 0.001$
Loss	Categorical cross-entropy
Metric	Accuracy
N. of training epochs	200
N. of epochs before early stopping	20
Feature preprocessing	Norm. → Gauss. → Norm.

Table 8.4: Result of the mistag probability calibration fit for *DeepElectronTagger* on $B^+ \to J/\psi K^+$ decays from data. The fitted model is $f(\omega_{\text{evt}}) = \texttt{sigmoid}(a + b \cdot \texttt{logit}(\omega_{\text{evt}}))$ as defined in Eq. 6.35.

Year (Data)	a	b	χ^2/d.o.f.
2017	0.069 ± 0.010	1.001 ± 0.017	1.5
2018	0.0644 ± 0.0063	0.968 ± 0.010	0.9

Table 8.5: Result of the mistag probability calibration fit for *DeepElectronTagger* on simulated samples. The fitted model is $f(\omega_{\mathrm{evt}}) = \mathtt{sigmoid}(a + b \cdot \mathtt{logit}(\omega_{\mathrm{evt}}))$ as defined in Eq. 6.35.

Process	Year (MC)	a	b	χ^2/d.o.f.
$B_s^0 \to J/\psi\,\phi(1020)$	2017	0.1000 ± 0.0076	1.071 ± 0.012	1.7
	2018	0.0869 ± 0.0054	1.0862 ± 0.0087	2.6
$B^+ \to J/\psi\,K^+$	2017	0.0763 ± 0.0083	1.046 ± 0.013	2.0
	2018	0.0826 ± 0.0063	1.068 ± 0.010	1.5

Table 8.6: Calibrated opposite-side electron tagger performance evaluated in $B^+ \to J/\psi\,K^+$ events in the 2017 and 2018 data samples. The tagging efficiency $\varepsilon_{\mathrm{tag}}$ and the total tagging power $\hat{P}_{\mathrm{tag}}$ are computed with Eq. 6.38 and Eq. 6.37, respectively. The effective dilution $\widehat{\mathcal{D}}_{\mathrm{tag}}$ is computed as $\widehat{\mathcal{D}}_{\mathrm{tag}} = \hat{P}_{\mathrm{tag}}/\varepsilon_{\mathrm{tag}}$. *DeepElectronTagger* is applied only in *dataset_JpsiTrkTrk*, while events in *dataset_JpsiMuon* are treated as *untagged*. The quoted uncertainties are statistical only. The evaluation of the uncertainty on $\widehat{\mathcal{D}}_{\mathrm{tag}}$ is quite complex due to the correlations between $\sigma(\varepsilon_{\mathrm{tag}})$ and $\sigma(\hat{P}_{\mathrm{tag}})$, and is beyond the scope of this work. The last row reports the total performance averaged by taking into account the relative statistics of the data samples (the performance in *dataset_JpsiMuon* are set to $\varepsilon_{\mathrm{tag}} \equiv 0\%$ and $\hat{P}_{\mathrm{tag}} \equiv 0\%$).

Year	$\varepsilon_{\mathrm{tag}}$ [%]	$\widehat{\mathcal{D}}^2_{\mathrm{tag}}$	$\hat{P}_{\mathrm{tag}}$ [%]
2017	6.20 ± 0.02	0.080	0.4978 ± 0.0007
2018	5.90 ± 0.01	0.077	0.4561 ± 0.0004
Total	5.25 ± 0.01	0.078	0.4096 ± 0.0003

9

OPPOSITE-SIDE JET TAGGING

Similarly to OS-lepton tagging, OS-jet tagging exploits the decay of the *other* b-hadron in the event to infer the flavor of the reconstructed b-meson. However, in this tagger no specific decay signature is searched. Instead, as a signature for the second b-hadron, a b-tagged jet is used.

Jet b-tagging, or heavy flavor tagging[I], is the identification of jets originating from bottom or charm quarks. Common identification strategies exploit hard fragmentation, long lifetimes, and relatively high masses of b and c hadrons. In CMS, b-tagging of jets is achieved with the *DeepJet* algorithm [127], a DNN-based method that takes as input all particles (both charged and neutral) associated with the jet and any decay vertex that may be identified within the jet region.

The OS-jet tagger follows the same three core steps as OS-lepton tagging: OS-jet selection, tagging decision, and mistag probability estimation. Since jets do not have a clear tagging variable, such as lepton charge, a DNN is used to both perform the tagging decision and estimate the mistag probability, following the logic already described in Eq. 6.31:

$$s_{\text{DNN}} > 0.5 + \epsilon \xrightarrow{\text{tag}} \text{signal } B_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = 1 - s_{\text{DNN}} \tag{9.1}$$

$$s_{\text{DNN}} < 0.5 - \epsilon \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = s_{\text{DNN}} \,, \tag{9.2}$$

where s_{DNN} is the network output score, which is calibrated to correspond to the probability that the reconstructed b -meson is a B_s^0, and ϵ is a small positive number that can be used to remove the region of the DNN output where the mistag probability is the highest ($\omega_{\text{evt}} \approx 0.5$). To avoid misunderstandings, here's a reminder of the differences in the DNN usage in lepton and lepton-less taggers:

- In the case of the OS-lepton taggers, the DNNs are trained to discriminate between right and wrong tags, and the tagging decision is given by the lepton charge. For these algorithms, ω_{evt} can assume any value between zero and one.

- In the case of the OS-jet and SS taggers, the DNNs are trained to discriminate objects that tag b-mesons containing a b-quark from those containing a $\overline{b}$-quark, and the tagging decision is given by the network score. For these algorithms, ω_{evt} can assume values only between zero and $0.5 - \epsilon$.

[I] Not to be confused with the b flavor tagging described in this thesis.

9.1 OS-JET SELECTION

OS-jet candidates are searched in events with a reconstructed b-meson decay and a selected PV, as described in Chapter 5, with tracks associated with the reconstructed b-meson excluded from the pool of candidates. The event is considered untagged by the OS-jet tagger if it contains an OS-lepton candidate (either electron or muon). This veto is imposed to make the three OS algorithms orthogonal. No significant performance is lost in the case of an OS lepton inside a b-jet, since the charged activity around the lepton is already taken into account with the *cone* variables. The jet tagger is applied only to *dataset_JpsiTrkTrk*.

Jet candidates are required to have $p_T > 10$ GeV, $|\eta| < 2.5$, a separation from the reconstructed b-meson decay of $\Delta R_{\eta,\phi} \geq 0.5$, and to contain at least two *High Purity* tracks (described in Section 4.2.1) with $|d_z| \leq 1$ cm with respect to the PV. As mentioned above, the *DeepJet* b-tagging algorithm is used to further select b-jets: candidates must have a probability of being identified as b-jets greater than 0.2, where the probability is defined as

$$P_{\text{DeepJet}}(b) = \texttt{probb} + \texttt{probbb} + \texttt{problepb}\,, \tag{9.3}$$

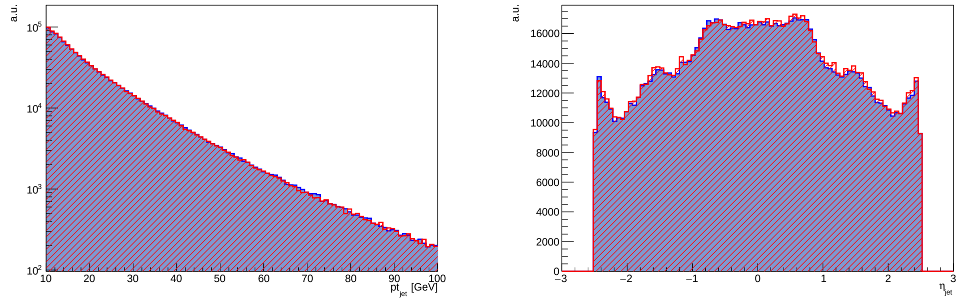
where `probb`, `probbb`, and `problepb` are the DeepJet estimated probabilities for a jet to contain a b-hadron, two b-hadrons, and a $b \to \ell X$ semileptonic decay[II], respectively. Selection requirements are optimized to maximize the tagging power and are summarized in Tab. 9.1. In the case of multiple OS-jet candidates, the one with the highest $P_{\text{DeepJet}}(b)$ is chosen.

About $\approx 20\%$ of the events in *dataset_JpsiTrkTrk* contain an OS-jet candidate that satisfies all the selection requirements. The input distributions of the candidates selected in the simulated samples are shown in Fig. 9.1. As expected, no significant differences between B_s^0 and $\overline{B}_s^0$ mesons are observed in these variables.

9.2 DNN OS-JET TAGGER (*DEEPJETCHARGE*)

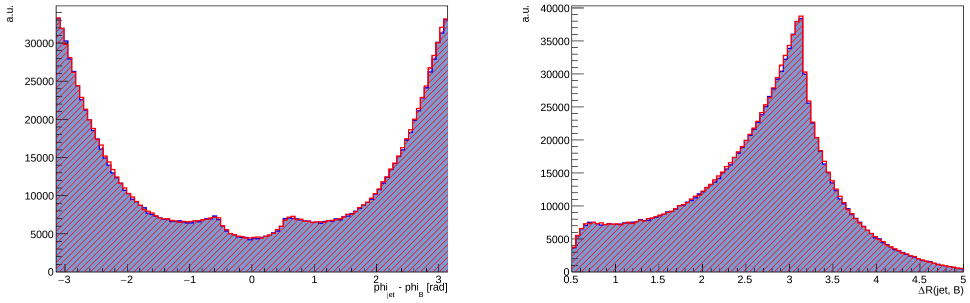
The tagging decision is performed using a DNN, hereafter called "*DeepJetCharge*", that is trained to discriminate OS-jets that tag b-mesons on the signal side containing a b-quark from those containing a $\overline{b}$-quark. The network output score is expected to approximate the probability of a B_s^0 on the signal side. The model uses the *DeepSets* architecture [128] and takes as input 16 features related to both the jet, its charged constituents, surrounding tracks, and the reconstructed b-meson. The model has been developed using the KERAS machine learning library.

[II] This considers cases in which a lepton from a real b-hadron decay did not pass the selections from the lepton based taggers.



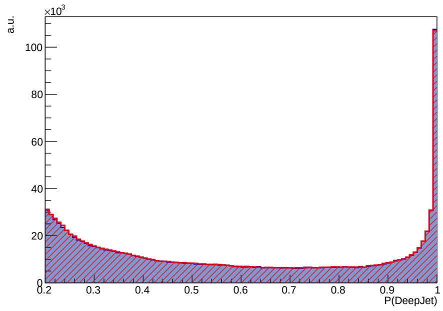
(a) p_T.

(b) η.



(c) $\Delta\varphi$ between the jet and the reconstructed B meson.

(d) ΔR between the jet and the reconstructed B meson. The cut on the left is due to the veto on the tracks used in the meson reconstruction, which are forbidden to be part of the jet.



(e) *DeepJet* probability.

Figure 9.1: Distribution of the selected jet input variables in simulated $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ events for B_s^0 (solid blue) and $\overline{B}_s^0$ (hatched red). All selection requirements reported in Tab. 9.1 are applied. Histograms are not stacked.

Table 9.1: Selection requirements for OS-jet candidates. The term ”Event veto” implies that the whole event is discarded.

Variable	Requirement
Reconstructed b-meson tracks	Excluded
OS-muon candidate	Event veto
OS-electron candidate	Event veto
p_T	≥ 10 GeV
$\lvert\eta\rvert$	≤ 2.5
N. of *High Purity* tracks with $d_z(PV) < 1$ cm	≥ 2
$\Delta R_{\eta,\phi}(b\text{-meson})$	≥ 0.5
$P_{\text{DeepJet}}(b)$	≥ 0.2

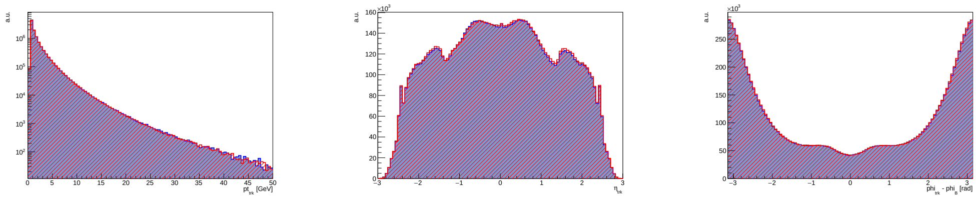
9.2.1 *Tracks selection for DeepJetCharge*

As already stated, *DeepJetCharge* exploits feature from the jet charged constituents. A charged constituent is defined as a track associated with the jet by the anti-k_t clustering algorithm. To be used in *DeepJetCharge*, the *High Purity* requirements need to be satisfied. Due to the poor clustering performance in the kinematic region where most OS-jet candidates reside (see Fig. 9.1a) additional tracks are searched within a cone of $\Delta R_{\eta,\phi} = 0.5$ from the jet axis. To limit contamination from pile-up tracks, a loose requirement of $|d_z| \leq 1$ cm is imposed for tracks not associated with the jet. These requirements are summarized in Tab. 9.2. The selected tracks are ordered by p_T and up to 15 will be used in *DeepJetCharge*.

The distribution of the track input variables can be found in Fig. 9.2. As for jet-related features, the track input variables do not show differences between B_s^0 and $\overline{B}_s^0$. The exception is the track charge, which is influenced by the initial charge (+2 from the pp collision) of the underlying event. This creates a bias that needs to be corrected in the output.

Table 9.2: Selection requirements for the track to be used in *DeepJetCharge*.

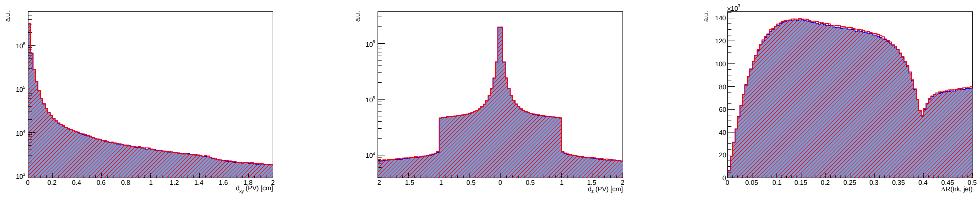
	Variable	Requirement
Assoc. with the OS-jet	Track quality	*High Purity*
	Signal b-meson tracks	Excluded
Not assoc. with the OS-jet	$\Delta R_{\eta,\phi}(\text{trk, jet})$	≤ 0.5
	$|d_z|$ w.r.t. PV	≤ 1 cm
	Track quality	*High Purity*
	Signal b-meson tracks	Excluded



(a) p_T.

(b) η.

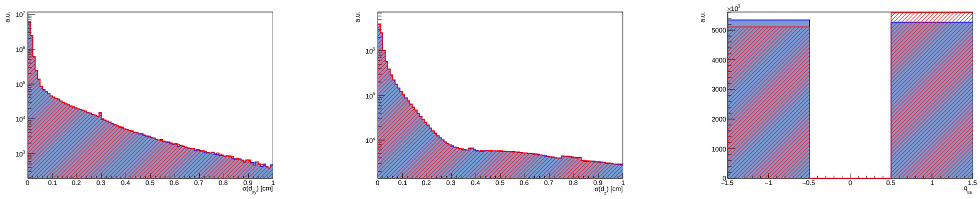
(c) $\Delta\varphi$ between the track and the reconstructed B hadron.



(d) d_{xy} of the track w.r.t. the primary vertex.

(e) d_z of the track w.r.t. the primary vertex.

(f) ΔR between the track and the tagged selected jet.



(g) $\sigma\left(d_{xy}\right)$ of the track.

(h) $\sigma(d_z)$ of the track.

(i) Track charge.

Figure 9.2: Distribution of the selected track input variables in simulated $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-$ events for B_s^0 (solid blue) and $\overline{B}_s^0$ (hatched red). All selection requirements reported in Tab. 9.1 are applied. Histograms are not stacked.

9.2.2 *Training data set*

DeepJetCharge is trained on $B_s^0 \rightarrow J/\psi\,\phi(1020)$ events simulated in the 2018 data-taking conditions. As for *DeepMuonTagger* and *DeepElectronTagger*, in all events the B_s^0 meson is identified using the MC truth and selected with requirements to ensure signal b-mesons in the same kinematic region of the B_s^0 candidates used in the CPV analysis. Selection requirements are reported in Tab. 7.4. Training candidates are selected with the same requirements listed in Tab. 9.1. A total of 2.5M OS-jet candidates are selected for training *DeepJetCharge*. The training data set is further split into *training*, *validation*, and *test* subsets with respective fractions of 60%, 20%, and 20%.

9.2.3 *Input features*

DeepJetCharge combines features from the jet candidate, its charged constituents, surrounding tracks, and the reconstructed b-meson to construct a powerful inference on the b-meson flavor. These are:

- Jet features:
  - p_T: jet transverse momentum;
  - η: jet pseudorapidity;
  - $\phi_{\text{jet}} - \phi_b$: difference in ϕ between the jet axis and the reconstructed b-meson momentum;
  - $P_{\text{DeepJet}}(b)$: DeepJet probability of being a b-jet, as defined in Eq. 9.3;
  - $\Delta R_{\eta,\phi}(\text{jet}, b\text{-meson})$: distance in the $\eta - \phi$ plane between the axis of the jet and the reconstructed b-meson momentum direction.
- Reconstructed b-meson features:
  - p_T: momentum of the reconstructed b-meson;
  - η: pseudorapidity of the reconstructed b-meson.
- Tracks features:
  - p_T: track transverse momentum;
  - η: track pseudorapidity;
  - $\phi_{\text{trk}} - \phi_b$: difference in ϕ between the track and the reconstructed b-meson momenta;
  - d_{xy}: track impact parameter with respect to the PV in the transverse plane;
  - d_z: track impact parameter with respect to the PV in the longitudinal direction;

- $\sigma(d_{xy})$: uncertainty on d_{xy};
- $\sigma(d_z)$: uncertainty on d_z;
- $\Delta R_{\eta,\phi}(\text{trk}, \text{jet})$: distance in the η, ϕ plane between the direction of the track momentum and the jet axis;
- q_{trk}: track charge.

Up to 15 tracks can be used for each event, and therefore up to 15 sets of these features.

9.2.4 *Model structure and training*

The *DeepJetCharge* model is, as mentioned above, based on the *DeepSets* [128] architecture. This architecture was selected because it is invariant with respect to the reordering of the tracks inside the event. In the *DeepSets* architecture, the output function can be written as

$$f(x) = \rho\left(\sum_{i \in S} \phi(\vec{x}_i)\right) \tag{9.4}$$

where $\vec{x}_i$ are the features of the i-*th* set element (a track in our case), ϕ is a submodel that acts on the features of each element and ρ maps the hidden, post-summation, features to the output score.

Both ϕ and ρ are implemented as a sequence of fully connected layers with `ELU` activation, interleaved with `Dropout` (with $p = 0.1$) layers. Since ϕ acts on each track independently, each of its layers is wrapped in a `TimeDistributed` modifier. The `ELU` (*Exponential Linear Unit*, [129]) activation function is defined as

$$\texttt{ELU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ e^x - 1 & \text{if } x < 0 \end{cases} . \tag{9.5}$$

There are 4 main blocks of layers:

1. 3 fully connected layers with width 16 act only on the flavor-invariant track features and on the jet and same side b-meson features (that is, everything but the charge)
2. the charge is added and 2 more fully connected layers of increasing width $(32, 48)$ are executed
3. all the hidden features are summed across all tracks, flattening the vector to a single dimension
4. 4 fully connected layers of width $[70, 100, 50, 1]$ convert the feature vector to a probability, the output of the model

Before entering the neural network, the sample is transformed so that the mean and variance for each of the input variables are, respectively, 0 and 1.

The parameters chosen for training can be found in Tab. 9.3

Table 9.3: Training options for the *DeepJetCharge* model.

Option/hyperparameter	Value
Optimiser	Adam, with $lr = 0.001$
Loss	Binary cross-entropy
Metric	Accuracy
Number of training epochs	100
Number of epochs before early stopping	10
Feature preprocessing	Normalisation

Since *DeepJetCharge* is trained to separate b-mesons from $\bar{b}$-mesons, an important property to preserve is the symmetry between the two classes. In particular, we want

$$P(s_{\text{DNN}}\,|\,b\,) = P\big(1 - s_{\text{DNN}}\,\big|\,\bar{b}\,\big) \tag{9.6}$$

Due to the charge imbalance in the LHC events (proton-proton), this property is not upheld, and an explicit symmetrization is necessary. Thus, we define

$$s_{\text{DNN}}^{\text{sym}}(x) = \frac{s_{\text{DNN}}(x) + [1 - s_{\text{DNN}}(\bar{x})]}{2} \tag{9.7}$$

as the symmetric output of the NN, where $\bar{x}$ is the neural network input with all the track charges reversed.

This transformation is applied only during evaluation after the training is completed. Hereafter, $s_{\text{DNN}}^{\text{sym}}$ will simply be referred to as s_{DNN}.

9.2.5 *Model performance*

The performance of the trained model is evaluated in the test data subset, which is statistically independent of the training and validation samples. The output score distribution is shown in Fig. 9.3. As discussed previously, the tagging decision follows the logic reported in Eq. 6.31:

$$s_{\text{DNN}} > 0.5 + \epsilon \xrightarrow{\text{tag}} \text{signal } B_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = 1 - s_{\text{DNN}} \tag{9.8}$$

$$s_{\text{DNN}} < 0.5 - \epsilon \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = s_{\text{DNN}}\,, \tag{9.9}$$

where ϵ is set to 0 for the calibration procedure and to 0.02 in the algorithm application.

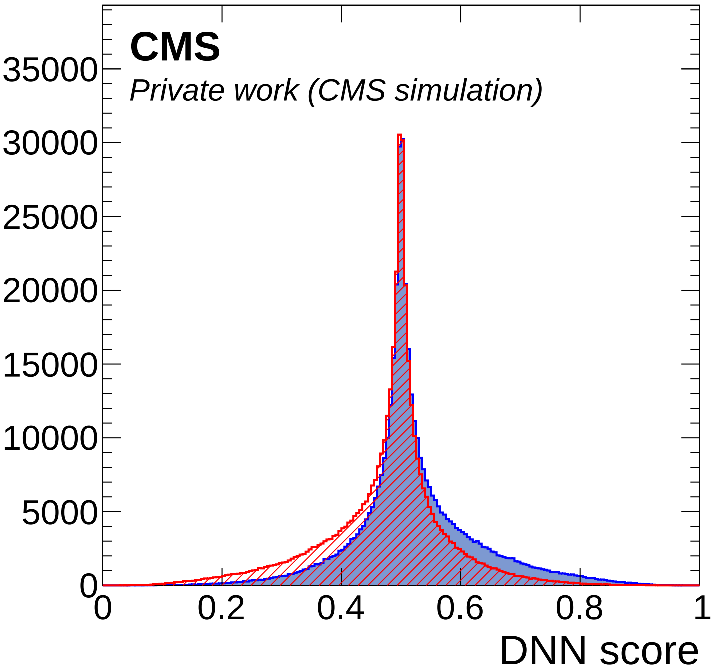


Figure 9.3: *DeepJetCharge* DNN output score for jets that tag B_s^0 (solid blue) and jets that tag $\overline{B}_s^0$ (hatched red).

9.2.6 *Model calibration*

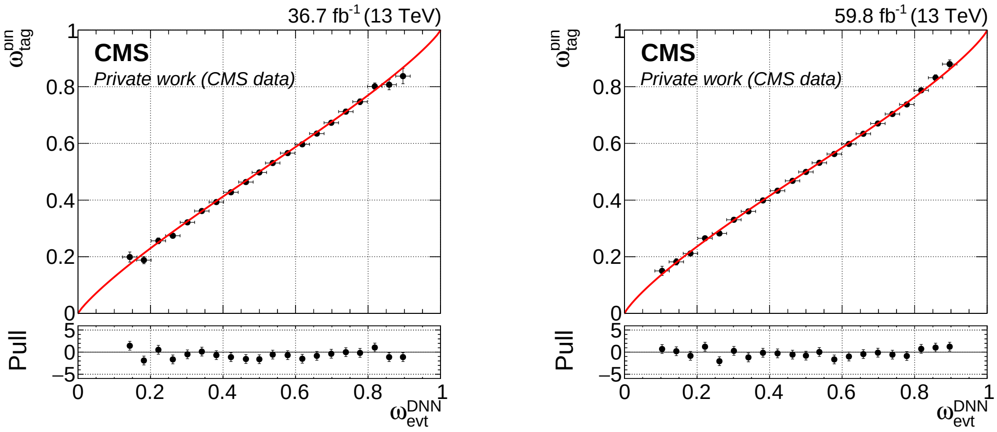
The predicted mistag probability is calibrated in data using self-tagging $B^+ \to J/\psi K^+$ decays, following the procedure described in Section 6.6. Bins of 0.04 in ω_{evt} are used, double the size of those used for the other taggers, due to the lack of statistics in the tails of the mistag probability distribution. Moreover, the a parameter (the constant term of the first-order polynomial fit function) is set to zero to force the function to pass through the point $(0.5, 0.5)$ to avoid flavor asymmetries caused by the calibration procedure[III]. Calibration was performed separately for the 2017 and 2018 data samples. The results are shown in Fig. 9.4 and Tab. 9.4.

The calibration of the DNN methods is also verified with the same procedure using independent samples of simulated $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ decays, where the reconstructed B_s^0 and B^+ mesons are matched to the generated ones to find their true flavor at production time. The results are shown in Fig. 9.5 and Tab. 9.5.

In general, the measured mistag probability is accurately predicted by *DeepJetCharge* over the entire range for all the simulated and data samples examined. Residual differences are well approximated by the calibration function. The nor-

---

[III] Reminder: $\omega_{\text{evt}} = 0.5$ marks the transition between the two flavor inferences.



(a) *DeepJetCharge* calibration plot for the 2017 data sample.

(b) *DeepJetCharge* calibration plot for the 2018 data sample.

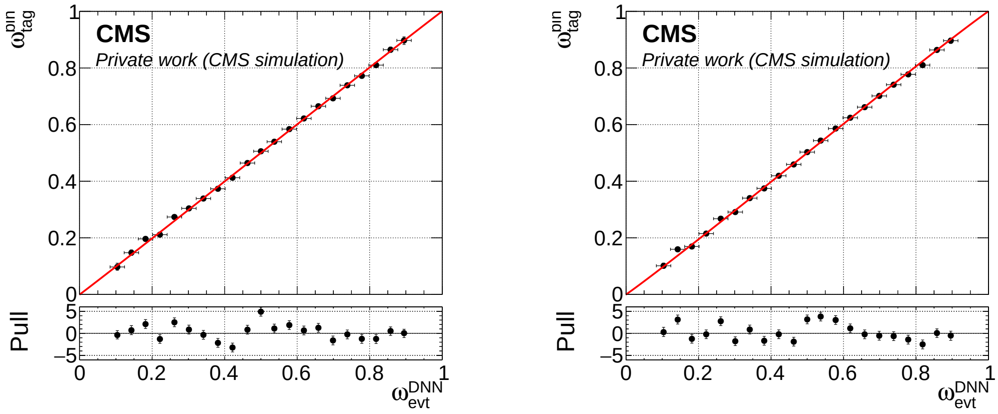
Figure 9.4: Result of the mistag probability calibration fit for *DeepJetCharge* on $B^+ \to J/\psi K^+$ decays from data. The solid red line shows the calibration fit to data (solid markers).

malized χ^2/d.o.f. values for all the fits are around 1 for data and around 1–4 for MC samples.

Differences in the mistag probability calibration between the $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ samples and the statistical uncertainties in the calibration fits are considered systematic uncertainties in the CPV analysis in Chapter 14.

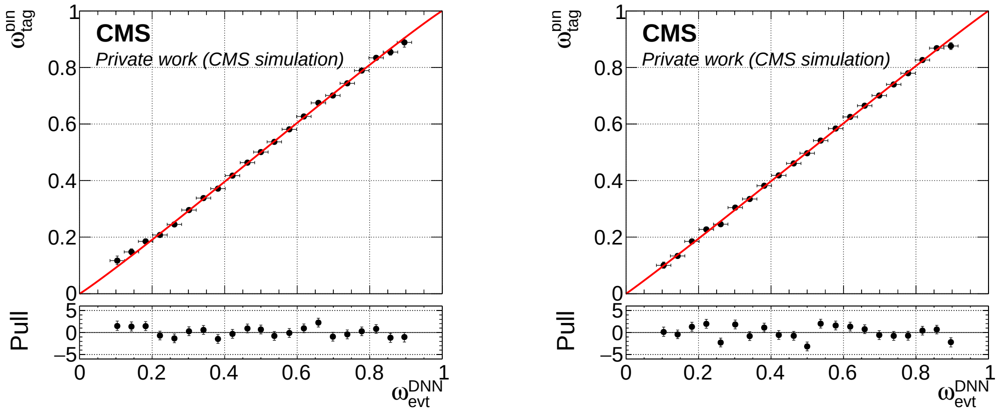
Table 9.4: Result of the mistag probability calibration fit for *DeepJetCharge* on $B^+ \to J/\psi K^+$ decays from data. The fitted model is $f(\omega_{\text{evt}}) = \texttt{sigmoid}(a + b \cdot \texttt{logit}(\omega_{\text{evt}}))$ (as defined in Eq. 6.25), where the a parameter is fixed to zero to force flavor symmetry in the calibration.

Data sample	a	b	χ^2/d.o.f.
2017	0	0.872 ± 0.011	1.2
2018	0	0.8475 ± 0.0067	0.9



(a) *DeepJetCharge* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2017 data-taking conditions).

(b) *DeepJetCharge* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2018 data-taking conditions).



(c) *DeepJetCharge* calibration plot for $B^+ \to J/\psi\,K^+$ simulated events (2017 data-taking conditions).

(d) *DeepJetCharge* calibration plot for $B^+ \to J/\psi\,K^+$ simulated events (2018 data-taking conditions).

Figure 9.5: Result of the mistag probability calibration fit for *DeepJetCharge* on simulated samples. The solid red line shows the calibration fit to data (solid markers).

Table 9.5: Result of the mistag probability calibration fit for *DeepJetCharge* on simulated samples. The fitted model is $f(\omega_{\text{evt}}) = \texttt{sigmoid}(b \cdot \texttt{logit}(\omega_{\text{evt}}))$ (as defined in Eq. 6.25), where the a parameter is removed to force flavor symmetry in the calibration.

Process	Year (MC)	b	χ^2/d.o.f.
$B_s^0 \to J/\psi\,\phi(1020)$	2017	1.0074 ± 0.0073	3.3
	2018	1.0243 ± 0.0053	3.8
$B^+ \to J/\psi\,K^+$	2017	1.0410 ± 0.0082	1.2
	2018	1.0243 ± 0.0063	2.2

9.3 TAGGING PERFORMANCE

The tagging performance of the OS-jet tagger is evaluated on the same $B^+ \to J/\psi K^+$ data samples used for calibrating the mistag probability and with the procedure very similar to the one described in Section 7.3 for the OS-muon tagger. The only difference is that events with a calibrated mistag probability between 0.48 and 0.50 are considered untagged ($\varepsilon_{\text{evt}} \equiv 0\% \to P_{\text{evt}} \equiv 0\%$). These are events for which the mistag probability is the highest and their contribution to the overall tagging performance is negligible. The final tagging performance measured in the $B^+ \to J/\psi K^+$ data sample is reported in Tab. 9.6.

When taking into account the relative statistics of the various data sets and the fact that events in *dataset_JpsiMuon* are considered untagged (in practice the performance is set to $\varepsilon_{\text{tag}} \equiv 0\%$ and $\hat{P}_{\text{tag}} \equiv 0\%$), the total tagging performance for the OS-jet tagger is $\hat{P}_{\text{tag}} = (0.4571 \pm 0.0004)\%$ with $\varepsilon_{\text{tag}} = (10.65 \pm 0.02)\%$ and $\widehat{\mathcal{D}}^2_{\text{tag}} = 0.043$.

Table 9.6: Calibrated opposite-side jet tagger performance evaluated in $B^+ \to J/\psi K^+$ events in the 2017 and 2018 data samples. The tagging efficiency ε_{tag} and the total tagging power $\hat{P}_{\text{tag}}$ are computed with Eq. 6.38 and Eq. 6.37, respectively. The effective dilution $\widehat{\mathcal{D}}_{\text{tag}}$ is computed as $\widehat{\mathcal{D}}_{\text{tag}} = \hat{P}_{\text{tag}}/\varepsilon_{\text{tag}}$. *DeepJetCharge* is applied only in *dataset_JpsiTrkTrk*, while events in *dataset_JpsiMuon* are treated as *untagged*. The quoted uncertainties are statistical only. The evaluation of the uncertainty on $\widehat{\mathcal{D}}_{\text{tag}}$ is quite complex due to the correlations between $\sigma(\varepsilon_{\text{tag}})$ and $\sigma(\hat{P}_{\text{tag}})$, and is beyond the scope of this work. The last row reports the total performance averaged by taking into account the relative statistics of the data samples (the performance in *dataset_JpsiMuon* are set to $\varepsilon_{\text{tag}} \equiv 0\%$ and $\hat{P}_{\text{tag}} \equiv 0\%$).

Year	ε_{tag} [%]	$\widehat{\mathcal{D}}^2_{\text{tag}}$	$\hat{P}_{\text{tag}}$ [%]
2017	12.64 ± 0.04	0.044	0.5573 ± 0.0007
2018	11.96 ± 0.02	0.042	0.5080 ± 0.0004
Total	10.65 ± 0.02	0.043	0.4571 ± 0.0004

10

SAME-SIDE TAGGING

The same-side tagging algorithm uses the charged *activity* surrounding the reconstructed b-meson to exploit the signatures present in the hadronization products. In principle, with hadron particle identification one can search for specific species of charged light hadrons (kaons or pions), to directly infer the flavor of the reconstructed b-meson[I]. Kaons and pions could be discriminated in CMS with dE/dx measurements in the silicon tracker [130], however, this is limited to low p_T ranges and it has not been studied in this work[II].

In this work, an inclusive approach is used, where (almost) all tracks around the reconstructed b-meson are used by a DNN method to perform a tagging inference. As for the OS-jet tagger, the DNN is used to perform both the tagging decision and estimate the mistag probability, following the logic already described in Eq. 6.31:

$$s_{\text{DNN}} > 0.5 + \epsilon \xrightarrow{\text{tag}} \text{signal } B_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = 1 - s_{\text{DNN}} \tag{10.1}$$

$$s_{\text{DNN}} < 0.5 - \epsilon \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0 \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = s_{\text{DNN}} \,, \tag{10.2}$$

where s_{DNN} is the network output score calibrated, to correspond to the probability that the reconstructed b-meson is a B_s^0 and ϵ is a small positive number that is used to remove the region of the DNN output where the mistag probability is the highest ($\omega_{\text{evt}} \sim 0.5$) either in the calibration process or in DNN deployment (or both).

10.1 TRACKS SELECTION

Candidate tracks are searched among all *High Purity* tracks within a $\Delta R_{\eta,\phi} = 0.8$ cone from the momentum direction of the reconstructed b-meson. To limit the contamination from pile-up tracks, a requirement of $|d_z| \leq 0.4$ cm with respect to the PV is imposed. Additionally, since in the ideal case all interesting tracks are produced in the proton-proton interaction vertex, we require that the significance of the transverse impact parameter with respect to the PV is $\left|d_{xy}\right|/\sigma\left(d_{xy}\right) \leq 1$.

[I] This is the strategy used by LHCb.

[II] This will change in a few years, when CMS will be upgraded with a timing detector (Mip Timing Detector, or MTD) which can give good PID information up to 3.5 GeV. This is discussed in Section 10.3.1.

Tracks associated with the reconstructed b-meson are excluded from the candidate pool. If any of the tracks in the cone is shared with any of the OS taggers[III], the event is considered untagged by the SS algorithm. This is done to avoid a correlation between the taggers which would break the tagging combination described later. Even without the explicit veto, the geometrical overlap in candidates between the SS tagger and the OS algorithms is expected to be very low since all OS objects are selected with a minimum distance from the reconstructed b-meson.

The selection requirements are summarized in Tab. 10.1. The selected tracks are ordered by $|d_z|$ and up to 20 tracks are used in each event. Events with no track candidates are considered untagged by the SS tagger.

Table 10.1: Selection requirements for the track to be used in the SS taggers.

Variable	Requirement
$\Delta R_{\eta,\phi}(\text{trk}, b\text{-meson})$	≤ 0.8
$\lvert d_z \rvert$ w.r.t. PV	≤ 0.4 cm
$\left\lvert d_{xy} \right\rvert / \sigma\left(d_{xy}\right)$ w.r.t. PV	≤ 1
Track quality	*High Purity*
Reconstructed b-meson tracks	Excluded
Common tracks with OS-muon	Event veto
Common tracks with OS-ele	Event veto
Common tracks with OS-jet	Event veto

10.2 DNN SS TAGGER (*DEEPSSTAGGER*)

The tagging decision is made using a DNN, hereafter called "*DeepSSTagger*", that is trained to discriminate *activity* that surrounds a b-meson containing a b-quark from those containing a $\overline{b}$-quark. The network output score is expected to approximate the probability of a B_s^0 on the signal side in $B_s^0 \to J/\psi\,\phi(1020)$ events. The model is very similar to *DeepJetCharge* and takes 20 tracks as input. The model has been developed using the KERAS machine learning library.

10.2.1 *Training data set*

DeepSSTagger is trained on a mixture of $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi\,K^+$ events simulated in the data-taking conditions of both 2017 and 2018. This is done to make the DNN unable to discriminate $B_s^0 \to J/\psi\,\phi(1020)$ from $B^+ \to J/\psi\,K^+$ to improve the calibration procedure, which is performed on $B^+ \to J/\psi\,K^+$ events.

[III] I.e., the tracks corresponding to the OS muon, OS-muon cone, OS electron, OS-electron cone, and the additional tracks searched around the OS jet.

The b-meson corresponding to the sample is identified using the MC truth and selected with requirements to ensure that the signal b-mesons are in the same kinematic region of the B_s^0 candidates used in the CPV analysis.

The SS tagger uses the charge of the particles produced in the hadronization for the discrimination, which depends on the secondary quark in the meson. Due to the difference in charge between the u and s quarks, it is expected that a B^+ meson has the same behavior as a $\overline{B}_s^0$ and not a B_s^0. During training, this is taken into account by switching the labels of the B^+ and B^- mesons to match the flavor of B_s^0-$\overline{B}_s^0$.

Training candidates are selected with the same requirements listed in Tab. 10.1. After selection, the sample contains 31M events (25M $B_s^0 \to J/\psi\,\phi(1020)$ and 6M $B^+ \to J/\psi K^+$), split into *training*, *validation*, and *test* subsets with respective fractions of 60%, 20%, and 20%.

Since the $B^+ \to J/\psi K^+$ sample is smaller than the $B_s^0 \to J/\psi\,\phi(1020)$ one, $B^+ \to J/\psi K^+$ events are assigned a higher weight during training.

10.2.2 *Input features*

DeepSSTagger combines features from the activity surrounding the reconstructed b-meson to construct a powerful flavor inference. These are:

- Reconstructed b-meson features:
  - p_T: momentum of the reconstructed b-meson;
  - η: pseudorapidity of the reconstructed b-meson.
- Tracks features:
  - p_T: track transverse momentum;
  - η: track pseudorapidity;
  - $\phi_{\text{trk}} - \phi_b$: difference in ϕ between the track and reconstructed b-meson momenta;
  - d_{xy}: track impact parameter with respect to the PV in the transverse plane;
  - d_z: track impact parameter with respect to the PV in the longitudinal direction;
  - $\sigma(d_{xy})$: uncertainty on d_{xy};
  - $\sigma(d_z)$: uncertainty on d_z;
  - $\sigma(p_T)$: uncertainty on p_T;
  - #hits: total number of hits in the track;
  - #pxhits: total number of pixel hits in the track;
  - χ^2/ndof: track reduced χ^2 from the track fit;

  - q_{trk}: track charge.

  Up to 20 tracks can be used for each event, and therefore up to 20 sets of these features.

As for *DeepJetCharge*, *DeepSSTagger* is designed to self-engineer complex features to exploit charge asymmetries in the activity surrounding the reconstructed b-meson.

10.2.3 *Model structure and training*

The model used in *DeepSSTagger* is extremely similar to that in *DeepJetCharge*, the main difference being a higher model width (number of neurons per layer) and the absence of jet information. The specific layers are:

1. 3 fully connected layers with width 64 act only on the flavor-invariant track features and same side b-meson features
2. the charge is added and 2 more fully connected layers of width 128 are executed
3. all the hidden features are summed across all tracks, flattening the vector to a single dimension
4. 5 fully connected layers of width $[128, 128, 64, 32, 1]$ convert the feature vector to the probability output

Before entering the neural network, the sample is transformed so that the mean and variance for each of the input variables are, respectively, 0 and 1.

The parameters chosen for training can be found in Tab. 10.2

Table 10.2: Training options for the *DeepSSTagger* model.

Option/hyperparameter	Value
Optimiser	Adam, with $lr = 0.001$
Loss	Binary cross-entropy
Metric	Accuracy
Number of training epochs	100
Number of epochs before early stopping	10
Feature preprocessing	Normalization

Like in the *DeepJetCharge* case, during evaluation, the model is made symmetric in B_s^0-$\overline{B}_s^0$ by applying Eq. 9.7 to the NN output.

10.2.4 *Model performance*

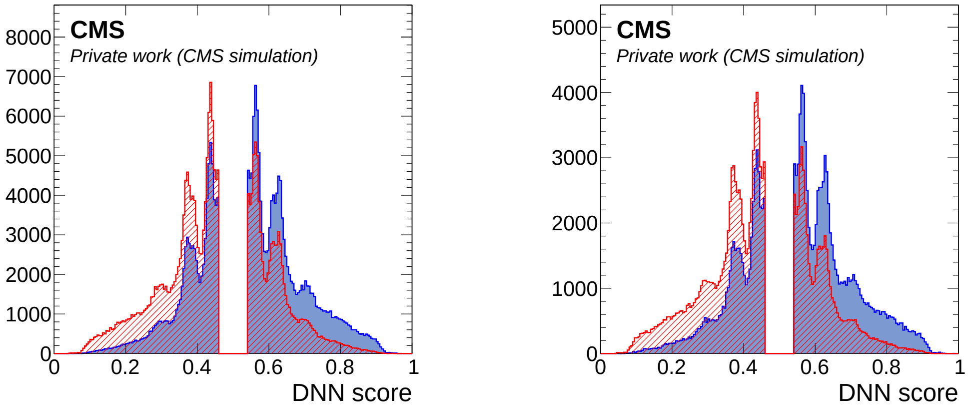
The performance of the trained model is evaluated in the *test* data subsets, which is statistically independent of the training and validation samples. The distribution of the output score is shown in Fig. 10.1. As discussed previously, the tagging decision follows the logic:

$$s_{\text{DNN}} > 0.5 + \epsilon \xrightarrow{\text{tag}} \text{signal } B_s^0/B^- \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = 1 - s_{\text{DNN}} \tag{10.3}$$

$$s_{\text{DNN}} < 0.5 - \epsilon \xrightarrow{\text{tag}} \text{signal } \overline{B}_s^0/B^+ \quad \text{with} \quad \omega_{\text{evt}}^{\text{DNN}} = s_{\text{DNN}}\,, \tag{10.4}$$

where

- $\epsilon = 0.04$ for the calibration procedure, since the core of the distribution was found to disrupt the calibration procedure;
- $\epsilon = 0.02$ for the algorithm application, which is looser than the value set in the calibration procedure since the calibration function distort and enlarge the s_{DNN} distribution[IV].



(a) *DeepSSTagger* DNN output score B_s^0 mesons (solid blue) and for $\overline{B}_s^0$ mesons (hatched red)

(b) *DeepSSTagger* DNN output score B$^-$ mesons (solid blue) and for B^+ mesons (hatched red)

Figure 10.1: *DeepSSTagger* DNN output score in $B_s^0 \rightarrow J/\psi\,\phi(1020)$ (left) and $B^+ \rightarrow J/\psi\,K^+$ (right) samples

---

[IV] To make an example: $s_{\text{DNN}} = 0.47$ is discarded, $s_{\text{DNN}} = 0.45$ is kept, but $f_{\text{cal}}(s_{\text{DNN}} = 0.45) = 0.47$ is also kept.

10.2.5 *Model calibration*

Since the hadronization of B_s^0 and B^+ mesons may be different, using self-tagging $B^+ \to J/\psi K^+$ decays in data is not an immediate solution for the calibration of *DeepSSTagger*, with the ideal way to calibrate the SS tagger being to directly study a *flavor-definite* decay of the B_s^0 meson. The following strategies have been studied, but only the last one succeeded.

$B_s^0 \to D_s^- \pi^+$ DECAYS Using $B_s^0 \to D_s^- \pi^+$ is a common choice for the calibration in B_s^0 decays, and is for example used by LHCb [131] for the calibration of their kaon based Same Side tagger. All particles produced in the decay are detectable (since no neutrinos are produced), making it a good candidate for reconstruction in CMS.

There are, however, a few significant problems with this approach. The first is that unlike B^+ mesons, B_s^0 mesons oscillate. For that reason, to calibrate the tagger a full oscillation analysis is required. Additionally, no trigger specific to this decay is present in the Run2 CMS trigger menu, forcing the use of a generic B-Physics trigger, which must necessarily apply substantial cuts to reduce the trigger rate. Finally, unlike the decays previously discussed (e.g. $B^+ \to J/\psi K^+$ or $B_s^0 \to J/\psi \phi(1020)$), this decay is fully hadronic, so the reconstruction would need to use exclusively tracks. Unlike muons, tracks are a very common object in a CMS event, leading to a large increase in combinatorial background. To reduce background, other significant cuts need to be applied to the candidates. These last two factors combined lead to a predicted number of candidates of no more than a few thousands, in the *largest* B-Physics related dataset in CMS, the BPH Parking dataset. As a consequence of the lack of statistics, this approach has not been studied in detail.

$B_s^{**} \to B^{+(*)} K^-$ DECAYS $B_s^{**} \to B^{+(*)} K^-$ decays, where B_s^{**} can be $B_{s2}^*(5840)$ or $B_{s1}(5830)$, are another option for the calibration of *DeepSSTagger*. This strategy too is used by LHCb [132]. The advantage of this method is that the B_s^{**} meson immediately decays to B^+ and therefore does not require an oscillation analysis. The vertex reconstruction can make use of the already available $B^+ \to J/\psi K^+$ vertices, and associate a track with the correct mass hypothesis[V].

The calibration curve in simulated samples was found to be similar to that of a B_s^0 meson. However, this decay requires a complex fit in data, due to the pollution from reflections of $B^{0**} \to B^+ \pi^-$ events (where the π is assigned the K mass) and a very large fraction of combinatorial background with a non-trivial shape (as can be seen in Fig. 10.2).

[V] The resonances can optionally decay into a B^{+*} meson instead of a B^+, which decays into the non-excited state by emitting a photon. The photon is not used in the reconstruction, which causes a shift to lower mass of the resonance.

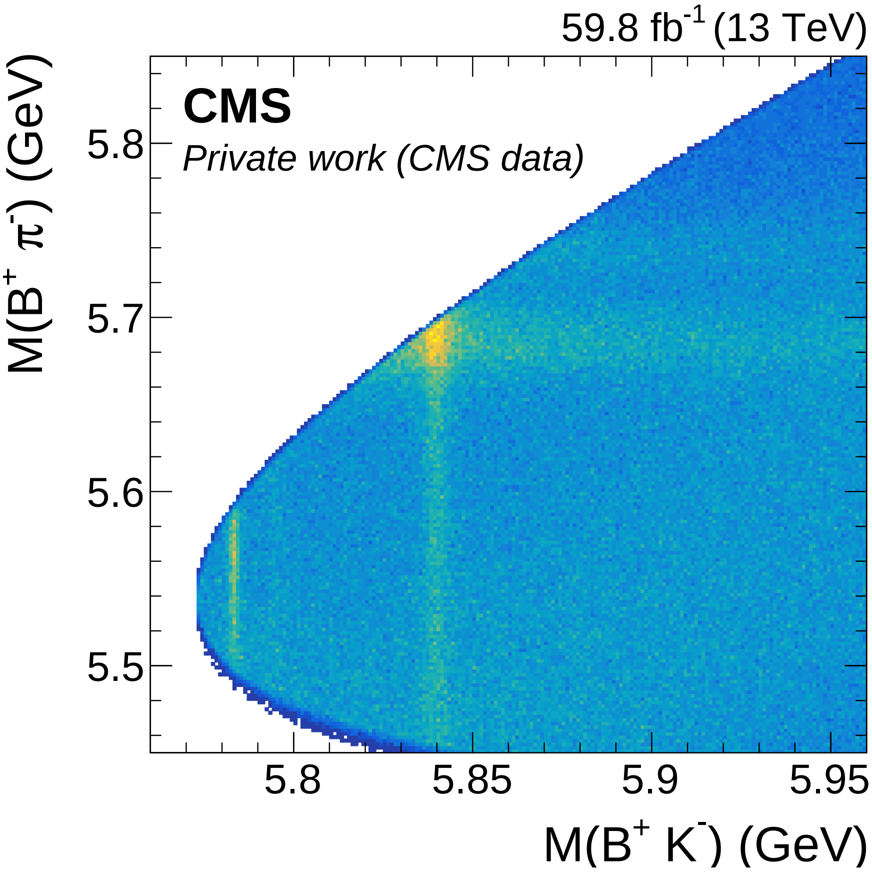


Figure 10.2: Two-dimensional mass distribution of the $B^{+}K^{-}$ vs $B^{+}\pi^{-}$ hypotheses, showing clearly the B_s^{**} (vertical) and B^{0**} (horizontal) resonances and the phase-space boundary.

Reflections are caused when, during reconstruction, a decay of B^{0**} is selected. B^{0**} decay into $B^+\pi^-$, but the masses are sufficiently close to be included in the B_s^{**} fit range. In the range we are interested in, two B^{0**} resonances are relevant: $B_2^*(5747)$ and $B_1(5721)$. For each of these resonances, a simulated sample with 1 M events was produced, each containing the respective resonance. Each resonance has a well defined mass distribution when correctly reconstructed, but when reconstructed as B_s^{**} the shape is mostly governed by the phase space boundary of the B^+K^- mass hypothesis as a function of the $B^+\pi^-$ mass. The templates assume that the B^+K^- mass is independent from $B^+\pi^-$ aside from border effects:

$$P^{\text{bulk}}(m_{\text{refl}}, m_d) = P^{\text{peak}}_{m_d}(m_d) \cdot P^{\text{bulk}}_{\text{refl}}(m_{\text{refl}}) \tag{10.5}$$

where m_{refl} and m_d represent the resonance mass when reconstructed respectively as B^+K^- and as $B^+\pi^-$. In this equation, $P(m_d)$ is a gaussian-like shape centered at the resonance mass, while $P_{\text{bulk}}(m_{\text{refl}})$ is taken to be an exponential distribution, which approximates well what is observed in the simulation. The phase-space boundary modifies the distribution to include a turn-on curve. The distribution including the boundary can be written as (assuming a single phase space boundary in range for simplicity)

$$\begin{aligned} P_{\text{refl}}(m_{\text{refl}}) &= \frac{\mathrm{d}}{\mathrm{d}m_{\text{refl}}} \text{Prob}(m < m_{\text{refl}}) \\ &= \frac{\mathrm{d}}{\mathrm{d}m_{\text{refl}}} \int_{m_{\text{refl}}^{\text{min}}}^{m_{\text{refl}}} \mathrm{d}m \int_{m_d^{\text{min}}}^{\phi(m)} \mathrm{d}m_d \, P^{\text{bulk}}(m, m_d) \\ &= \frac{\mathrm{d}}{\mathrm{d}m_{\text{refl}}} \int_{m_{\text{refl}}^{\text{min}}}^{m_{\text{refl}}} \mathrm{d}m \, P^{\text{bulk}}_{\text{refl}}(m) \int_{m_d^{\text{min}}}^{\phi(m)} \mathrm{d}m_d \, P^{\text{peak}}_{m_d}(m_d) \\ &= P^{\text{bulk}}_{\text{refl}}(m_{\text{refl}}) \cdot \text{CDF}^{\text{peak}}_{m_d}(\phi(m_{\text{refl}})) \,, \end{aligned} \tag{10.6}$$

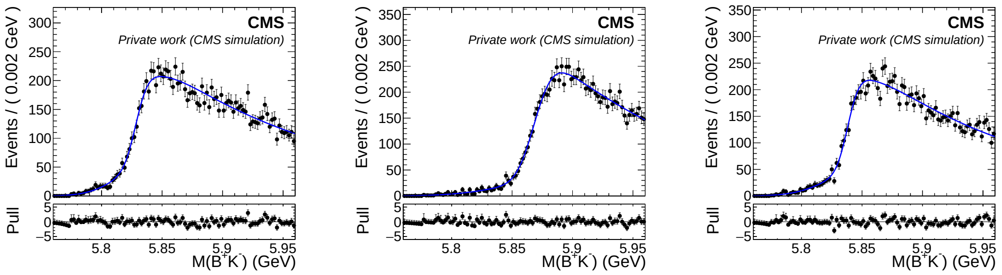
where $\phi(m)$ indicates the phase-space boundary of m_d as function of m and CDF_{m_d} is the cumulative distribution function of m_d. Since $P(m_d)$ is gaussian-like, CDF_{m_d} takes a shape similar to an `Erf` function (the integral of a gaussian).

$$\texttt{Erf}(x \,|\, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{x} \mathrm{d}y \, \mathrm{e}^{\frac{(y-\mu)^2}{2\sigma^2}} \tag{10.7}$$

A similar shape is obtained if a second boundary is included on the other side of the m_{refl} spectrum. This distribution can be fitted to the simulated samples, obtaining estimates for $P_{\text{bulk}}(m_{\text{refl}})$ (an exponential) and the parameters for the CDF (modeled thes a sum of two `Erf` with the same center μ). If $\phi(m)$ is assumed to be a linear function (a good approximation in the limited range of the m_d resonance), its parameters can be reabsorbed in the `Erf`.

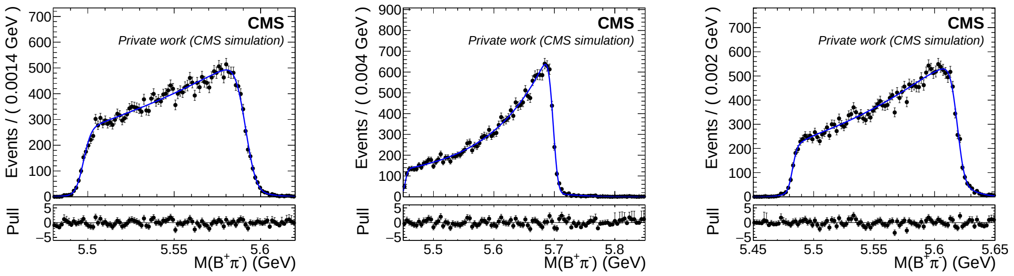
The m_d distributions fitted to the simulated samples can be found in Fig. 10.3.

Estimation of the reflection yields is a more complex task. The yields cannot be left free to float in the fit, as the reflection shape can easily be approximated by



(a) $B_1 \to B^+\pi^-$ (b) $B_2^* \to B^+\pi^-$ (c) $B_2^* \to B^{+*}\pi^-$

Figure 10.3: Templates for the B^{0**} reflections in the B_s^{**} fit, obtained using MC samples.



(a) $B_{s1} \to B^+K^-$ (b) $B_{s2}^* \to B^+K^-$ (c) $B_{s2}^* \to B^{+*}K^-$

Figure 10.4: Templates for the B_s^{0**} reflections in the B^{0**} fit, obtained using MC samples.

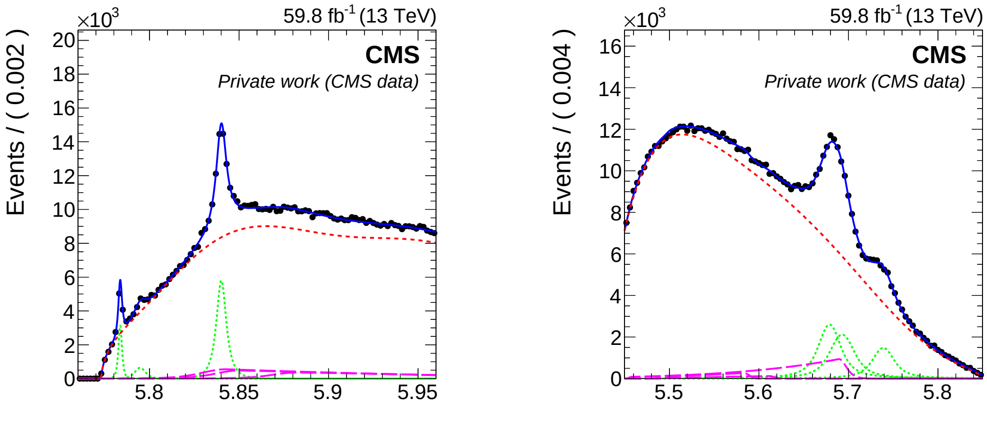
the combinatorial background. For the estimation, a secondary fit of the $B^+\pi^-$ hypothesis is done simultaneously. In the $B^+\pi^-$, the B_s^{**} are instead seen as reflections. Their shape is estimated similarly to the B^{0**} in simulated samples (1 M events per resonance), and is found in Fig. 10.4.

In the ideal case, a two-dimensional fit of the two mass hypotheses would be done. However, the 2D shapes of the resonance masses (peak × reflection) and especially of the combinatorial background are not easily estimated. To avoid the 2D fit, a different method is used: the sample is randomly split into two subsamples of the same size and each is fitted with a different mass hypothesis. The two subsamples are fitted simultaneously, sharing only the yield parameters. After the fit, the yield of each resonance is multiplied by two, thus recovering the full sample yield.

In the fit, the combinatorial background of both mass hypotheses is taken as a degree 6 polynomial, multiplied by a power law with a threshold

$$P_{\text{comb}}(m) = \theta(m - m_{\text{min}}) \cdot (m - m_{\text{min}})^{\alpha} \cdot \texttt{Pol}_6(m)\,, \tag{10.8}$$

where θ represents the Heavyside function (1 if its argument is greater than 0, 0 otherwise).



(a) $M(B^+K^-)$

(b) $M(B^+\pi^-)$

Figure 10.5: Final result of the fit to the mass of the B_s^{**} and B^{0**} resonances. The short-dashed green curves represent the correctly reconstructed resonances, depending on the plot, while the long-dashed purple represents the wrong reconstructed resonance. In both plots the red curve represents the combinatorial background.

The resulting fit can be found in Fig. 10.5.

The combinatorial background represents the biggest challenge in the fit, as the complex shape from Eq. 10.8 is required to follow its shape, but at the same time is too flexible, leading to an unstable fit. To constrain the combinatorial background, an attempt was made using "same-charge" combinations B^+K^+. The resulting distribution presents, as expected, a smooth shape with no resonances, but a different distribution compared to that of the B^+K^-.

For the mistag calibration, the mass fit shown in Fig. 10.5 needs to be executed in each bin of the mistag distribution, following the method of Section 6.6, which introduces additional challenges since the combinatorial background shape was found to depend on the bin (specifically, the distribution was significantly different for $\omega_{\text{evt}} > 0.5$ and $\omega_{\text{evt}} < 0.5$).

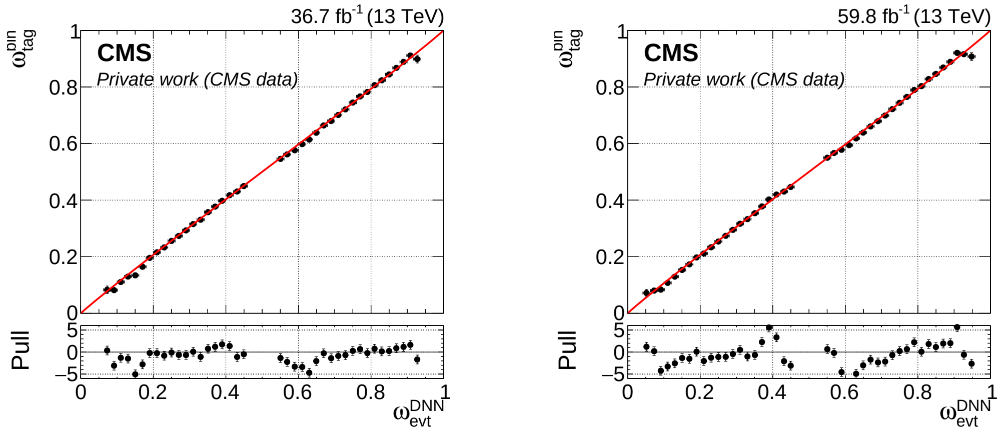
In the end, the fit proved to be far too unstable: the large statistical uncertainty led to a variation on the estimated tagging power of the same side tagger ranging from 0.3% (comparable to the OS taggers) to 5% (the current *total* estimated tagging power), and simple variations on the combinatorial background model further increased this. All these uncertainties would be propagated, as systematic uncertainties, to any analysis making use of the tagger (as will be shown in Chapter 14), so in the end, this calibration method was deemed unsuitable.

FACTORIZATION OF THE DIFFERENCES USING MC SAMPLES Since the usage of B_s^0 decays in data is not feasible, $B^+ \to J/\psi K^+$ decays are used instead, with simulations to account for differences between the two hadronization processes. Possible biases introduced by this technique will be checked and eventual systematic uncertainties will be evaluated in the CPV analysis.

The four samples of interest are $B^+ \to J/\psi K^+$ MC and data, and $B_s^0 \to J/\psi\,\phi(1020)$ MC and data, where the standard calibration fit can be performed in all samples but the last one. The procedure is similar to that of the other taggers, with some differences:

- events with a predicted DNN mistag probability (before calibration) $0.46 < \omega_{\mathrm{evt}}^{DNN} < 0.5$ are considered untagged,
- events with a *calibrated* mistag probability $0.48 < \omega_{\mathrm{evt}} < 0.5$ are also considered untagged[VI],
- as done for the jet tagger, the intercept parameter of the fit model is fixed to zero to force the symmetry between the two flavors.

The calibration plots are shown in Fig. 10.6 and Fig. 10.7, for data and MC respectively, with the corresponding fit results reported in Tab. 10.3.



(a) *DeepSSTagger* calibration plot for the 2017 $B^+ \to J/\psi K^+$ data sample.

(b) *DeepSSTagger* calibration plot for the 2018 $B^+ \to J/\psi K^+$ data sample.

Figure 10.6: Result of the mistag probability calibration fit for *DeepSSTagger* on $B^+ \to J/\psi K^+$ decays from data. The solid red line shows the result of the calibration fit to data (solid markers).

The resulting calibration function is then used to produce the $B_s^0 \to J/\psi\,\phi(1020)$ calibration for data with the following scheme:

$$f_{B_s^0,\mathrm{Data}} = f_{B_s^0,\mathrm{MC}} \circ f^{-1}_{B^+,\mathrm{MC}} \circ f_{B^+,\mathrm{Data}} \tag{10.9}$$

[VI] Even if before the calibration a hard cut at 0.46 is imposed, the calibration function can produce mistag probability higher than this number. For this reason, an additional more relaxed cut is imposed to avoid values around 0.5.

Table 10.3: Result of the mistag probability calibration fits for *DeepSSTagger*. The fitted model is $f(\omega_{\text{evt}}) = \texttt{sigmoid}(a + b \cdot \texttt{logit}(\omega_{\text{evt}}))$ (as defined in Eq. 6.25), where the a parameter is fixed to zero to force flavor symmetry in the calibration.

Data/MC	Process	Year	b	χ^2/d.o.f.
Data	$B^+ \to J/\psi K^+$	2017	0.9706 ± 0.0037	3.1
		2018	0.9690 ± 0.0023	7.1
MC	$B_s^0 \to J/\psi\,\phi(1020)$	2017	0.8684 ± 0.0067	2.1
		2018	0.8780 ± 0.0048	3.7
	$B^+ \to J/\psi K^+$	2017	0.9423 ± 0.0078	1.7
		2018	0.9511 ± 0.0058	2.5

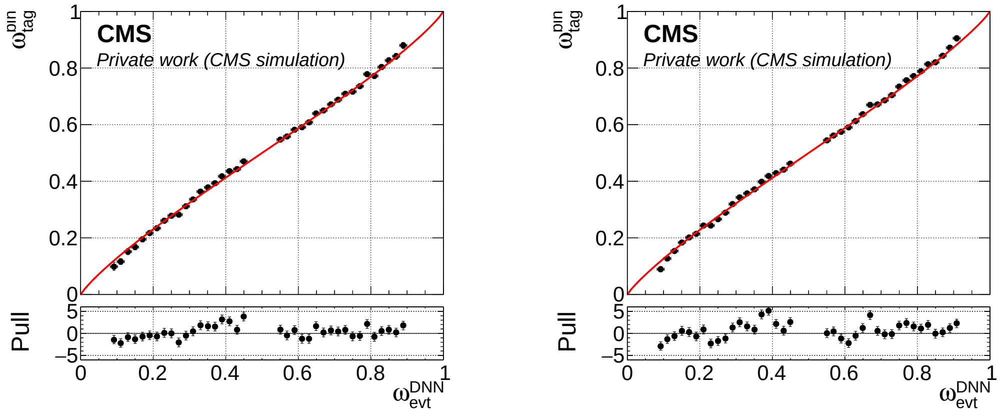
where f are the calibration functions and B_s^0 and B^+ refers to the $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ samples. Eq. 10.9 is implemented as:

$$\omega_{\text{evt}} = f_{B_s^0,\text{Data}}(\omega_{\text{evt}}^{\text{DNN}}) = f_{B_s^0,\text{MC}}\Big(f^{-1}_{B^+,\text{MC}}\big(f_{B^+,\text{Data}}(\omega_{\text{evt}}^{\text{DNN}})\big)\Big) \tag{10.10}$$

In practice, the calibration obtained in data from $B^+ \to J/\psi K^+$ decays is corrected with the observed differences between the calibrations of $B_s^0 \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$ in the simulated segments. The resulting functions are shown in Fig. 10.8 and reported in Tab. 10.4.

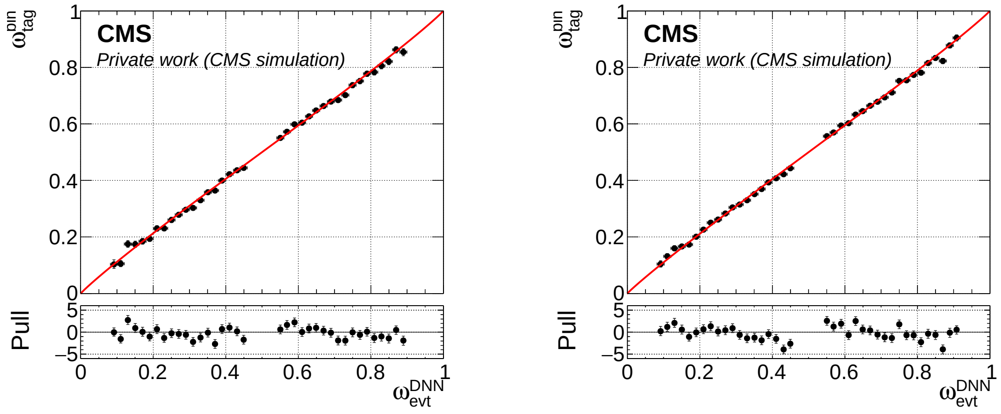
Table 10.4: Resulting mistag probability calibration functions for *DeepSSTagger* for $B_s^0 \to J/\psi\,\phi(1020)$ events in data, obtained with the procedure in Eq. 10.9.

Year	b
2017	0.8946 ± 0.0034
2018	0.8945 ± 0.0023



(a) *DeepSSTagger* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2017 data-taking conditions).

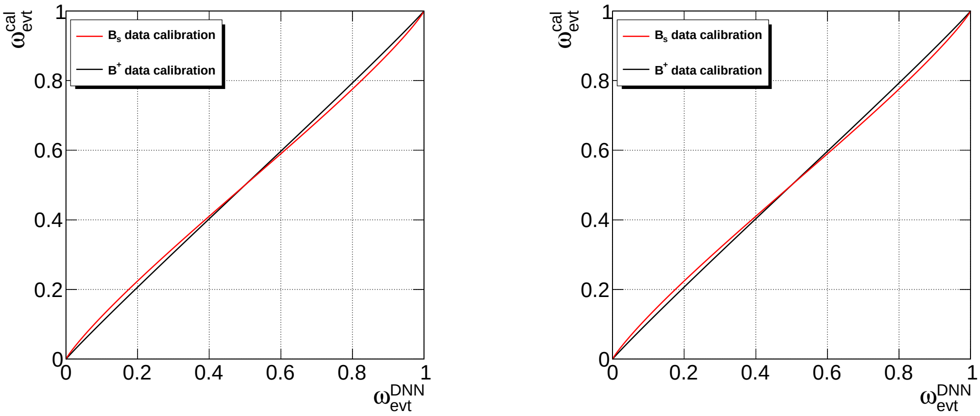
(b) *DeepSSTagger* calibration plot for $B_s^0 \to J/\psi\,\phi(1020)$ simulated events (2018 data-taking conditions).



(c) *DeepSSTagger* calibration plot for $B^+ \to J/\psi K^+$ simulated events (2017 data-taking conditions).

(d) *DeepSSTagger* calibration plot for $B^+ \to J/\psi K^+$ simulated events (2018 data-taking conditions).

Figure 10.7: Result of the mistag probability calibration fit for *DeepSSTagger* on simulated samples. The solid red line shows the result of the calibration fit (solid markers).



(a) Comparison of B^+ and B_s^0 calibration curves in 2017 data.

(b) Comparison of B^+ and B_s^0 calibration curves in 2018 data.

Figure 10.8: Comparison of the calibration curves before (black, B^+) and after (red, B_s) the application of Eq. 10.9. The red function is the one used as calibration for the analysis sample.

10.3 TAGGING PERFORMANCE

Evaluating the tagging performance for the SS tagger is not straightforward, since the $B^+ \to J/\psi K^+$ data sample may not be a suitable proxy. The standard ($\sum P_{\text{bin}}$) method, however, can only work with $B^+ \to J/\psi K^+$ events, as it needs a flavor-defined final state to distinguish the right and wrong tags. The method used is instead based on the per-event dilution ($\sum P_{\text{evt}}$):

1. first, the $\mathcal{D}^2_{\text{evt}} = (1 - 2\omega_{\text{evt}})^2$ distribution for the signal is obtained using the sideband background subtraction method;

2. then, the final performance is obtained with

$$\hat{P}_{\text{tag}} = \frac{\sum_{\text{evt}\in\text{tagged}} w^{\text{sb}}_{\text{evt}} \mathcal{D}^2_{\text{evt}}}{\sum_{\text{evt}} w^{\text{sb}}_{\text{evt}}} , \tag{10.11}$$

   where $w^{\text{sb}}_{\text{evt}}$ is the event weight obtained from sideband subtraction.

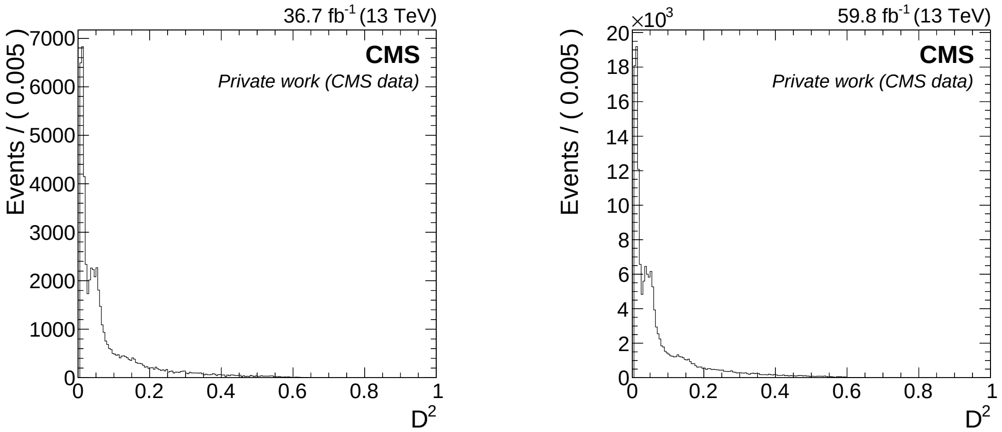
This strategy can be applied to both $B^0_s \to J/\psi\,\phi(1020)$ and $B^+ \to J/\psi K^+$, as it relies on the predicted mistag probability.

$B^+ \to J/\psi K^+$ samples are used to validate the $\sum P_{\text{evt}}$ method, by comparing its results with the $\sum P_{\text{bin}}$ one. The two approaches give similar results within $1-2\%$. The results are reported in Tab. 10.5, while the background-subtracted $\mathcal{D}^2_{\text{evt}}$ distributions for $B^0_s \to J/\psi\,\phi(1020)$ are shown in Fig. 10.9. It is worth recalling that these performances are just figures of merits of the methods and are not used in the analysis, where a per-event dilution is applied.

The values obtained from the $B^0_s \to J/\psi\,\phi(1020)$ samples are used to evaluate the following total performance of the SS tagger when taking into account the relative statistics of the data sets: $\hat{P}_{\text{tag}} = (3.39 \pm 0.01)\%$ with $\varepsilon_{\text{tag}} = (41.91 \pm 0.07)\%$ and $\widehat{\mathcal{D}}^2_{\text{tag}} = 0.081$. The combined performance of the entire tagging framework is presented in the next section.

10.3.1 *Tagging performance extrapolation with the future timing detector*

LHC schedule alternates periods of data taking with periods of shutdown. During the next LHC shutdown, expected to start in 2026, both LHC and CMS are expected to receive significant upgrades. The MTD (MIP Timing Detector) is a new subdetector in CMS, which adds the ability to measure the production time of minimum ionizing particles [133]. The explicit goal of the subdetector is the disentanglement of the signals from pileup collisions, that due to the LHC upgrade are expected to increase to nearly 200. However, the timing information can also be used for the identification of the particle flavor.



(a) Background-subtracted $\mathcal{D}^2_{\text{evt}}$ distribution for the 2017 $B^0_s \to J/\psi\,\phi(1020)$ data sample.

(b) Background-subtracted $\mathcal{D}^2_{\text{evt}}$ distribution for the 2018 $B^0_s \to J/\psi\,\phi(1020)$ data sample.

Figure 10.9: Background-subtracted $\mathcal{D}^2_{\text{evt}}$ distributions for the SS tagger applied to the $B^0_s \to J/\psi\,\phi(1020)$ data samples.

Particle identification (PID) relies on the difference in velocity β for particles with the same momentum p and different masses. The velocity of a particle can be estimated as

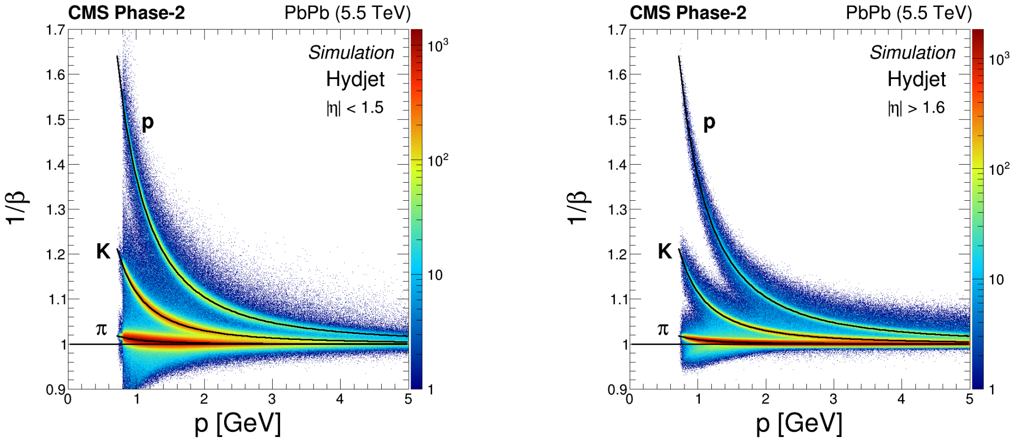
$$\frac{1}{\beta} = \frac{c\,(t_{\text{mtd}} - t_0)}{L}\,, \tag{10.12}$$

where L is the flight distance, t_0 is the event start time and t_{mtd} is the time of arrival of the particle in the MTD. The event start time t_0 can be measured by clustering the primary vertices as described in paragraph 4.2.1 while taking into account the newly added timing information and averaging the start time of the different particles contained in the vertex. Fig. 10.10 shows the relation between β and p in the barrel and endcaps of the timing detector.

The use of Eq. 10.12 for particle identification is only valid for particles that have been produced directly in a pp interaction. *DeepSSTagger* is required in the ideal case to use only the products of the B^0_s meson hadronization, and is thus a prime candidate for evaluating the predicted performance of the MTD.

This study is described in [134]. The network developed in Section 10.2 was modified to include flavor information from the PID and tested on simulated samples augmented with a "simulated" PID. Flavor was passed as a sequence of three numbers, representing respectively the probability of being a pion/muon[VII], a kaon, or a proton. The PID was simulated using the MC truth and generating the effects of the MTD to match the official performance measurements, which refer to a time resolution in the barrel timing detector of 40 ps. Specifically, the simulation makes use of the PID efficiency, defined as the probability Prob(PID $= A|$flav. $= B$)

[VII] Pions and muons are too close in mass to be resolved by looking at the time of flight.



(a) $|\eta| < 1.5$ (b) $|\eta| > 1.6$

Figure 10.10: Inverse velocity as a function of momentum in barrel ($|\eta| < 1.5$) and endcap ($|\eta| > 1.6$) in PbPb simulation at 5.5 TeV.

of assigning PID A (pion, kaon, or proton) to a particle with real flavor B, and the PID purity, defined as the probability $\mathrm{Prob}(\text{flav.} = A \,|\, \text{PID} = B)$ that a particle with assigned PID B actually has flavor A^{VIII}, both given as a function of the momentum p and pseudorapidity η. Starting from the MC truth, the efficiency is used to give a simulated PID response given the generated particle flavor. The three flavor probabilities are then computed using the purity, and passed as input to the network.

Using this simulation, the total performance of the suite of taggers presented in this thesis was increased by $\sim 24\%$. An additional test was performed, using a significantly degraded timing resolution (70 ps), corresponding to a worst case scenario of performance degradation in the barrel subdetector due to radiation damage. In this test the performance increase was $\sim 14\%$.

[VIII] As can be seen, the two probabilities are related by the rules of joint probability.

Table 10.5: Calibrated same-side tagger performance evaluated in the 2017 and 2018 *dataset_JpsiTrkTrk* samples. The effective dilution $\widehat{\mathscr{D}}_{\text{tag}}$ is computed as $\widehat{\mathscr{D}}_{\text{tag}} = \hat{P}_{\text{tag}}/\varepsilon_{\text{tag}}$. The last row reports the total performance measured in $B_s^0 \to J/\psi\,\phi(1020)$ using the $\sum P_{\text{evt}}$ method and averaged by taking into account the relative statistics of the data samples (the performance in *dataset_JpsiMuon* are set to $\varepsilon_{\text{tag}} \equiv 0\%$ and $\hat{P}_{\text{tag}} \equiv 0\%$). The quoted uncertainties are statistical only.

Data sample	Year	ε_{tag} [%]	$\widehat{\mathscr{D}}^2_{\text{tag}}$	$\hat{P}_{\text{tag}}$ [%]
$B^+ \to J/\psi K^+$ ($\sum P_{\text{bin}}$)	2017	50.65 ± 0.09	0.126	6.376 ± 0.008
	2018	50.70 ± 0.06	0.122	6.167 ± 0.005
$B^+ \to J/\psi K^+$ ($\sum P_{\text{evt}}$)	2017	51.34 ± 0.04	0.126	6.479 ± 0.010
	2018	51.26 ± 0.03	0.121	6.226 ± 0.006
$B_s^0 \to J/\psi\,\phi(1020)$ ($\sum P_{\text{evt}}$)	2017	47.0 ± 0.2	0.081	3.83 ± 0.02
	2018	47.0 ± 0.1	0.081	3.80 ± 0.01
$B_s^0 \to J/\psi\,\phi(1020)$ ($\sum P_{\text{evt}}$)	Total	41.91 ± 0.07	0.081	3.39 ± 0.01

11

TAGGING ALGORITHMS COMBINATION AND FINAL PERFORMANCE

The performances of the four discussed taggers are combined in this chapter. All performances are measured on the samples used for the CPV analysis that is discussed in Part IV, one of the key analyses for which the taggers were initially developed. The selections described in Chapter 5 are applied.

A comparison with the performances quoted by other LHC experiments in the same channel, as well as previous CMS results, is also provided.

11.1 COMBINED OS PERFORMANCE

The three opposite-side tagging algorithms are mutually exclusive by design[I] and their combined performance can therefore be obtained by a simple sum. The total efficiency of the OS algorithms is $\varepsilon_{\text{tag}} = (23.37 \pm 0.03)\%$ with a total tagging power of $P_{\text{tag}} = (2.402 \pm 0.006)\%$, which is $\approx 90\%$ higher than the tagging power available in the Run-1 CMS analysis [4]. The results are summarized in Tab. 11.1.

Table 11.1: Calibrated opposite-side tagging performance evaluated in $B^+ \to J/\psi K^+$ events data samples with the "$\sum P_{\text{bin}}$" method, averaged between *dataset_JpsiMuon* and *dataset_JpsiTrkTrk*. The quoted uncertainties are statistical only.

Year	ε_{tag} [%]	$\widehat{\mathcal{D}}^2_{\text{tag}}$	$\hat{P}_{\text{tag}}$ [%]
OS muon	7.47 ± 0.03	0.205	1.535 ± 0.006
OS electron	5.25 ± 0.01	0.078	0.4096 ± 0.0003
OS jet	10.65 ± 0.02	0.043	0.4571 ± 0.0004
Total OS algorithms	23.37 ± 0.03	0.103	2.402 ± 0.006

11.2 TAGGERS OVERLAP COMBINATION

The SS algorithm, however, can overlap with any of the OS ones with the logic summarized in Tab. 11.2. This happens in $\approx 20\%$ of the signal events tagged by the SS tagger. In this case, both tagging inferences are used to improve performance. The combination of a generic number i of tagging algorithms can be performed with the following procedure:

[I] With hierarchy: OS-muon > OS-electron > OS-jet.

1. First, *probabilities*[II] of the reconstructed b-meson to contain a b or a $\bar{b}$ are recomputed taking into account information from multiple taggers:

$$p(\bar{b}) = \prod_i \left(\frac{1-\xi_i}{2} + \xi_i (1-\omega_i) \right), \tag{11.1}$$

$$p(b) = \prod_i \left(\frac{1+\xi_i}{2} - \xi_i (1-\omega_i) \right) \tag{11.2}$$

   where $\xi_i = \pm 1$ and ω_i are the tagging decision (defined in Eq. 6.1) and the calibrated mistag probability of the i-th tagger, respectively.

2. Then, the new probabilities are normalized as

$$P(\bar{b}) = \frac{p(\bar{b})}{p(b) + p(\bar{b})}, \qquad P(b) = \frac{p(b)}{p(b) + p(\bar{b})}. \tag{11.3}$$

3. Finally, the tag decision is performed as:

$$P(\bar{b}) > P(b) \longrightarrow \qquad \xi_{\rm tag} = +1, \quad \omega_{\rm evt} = 1 - P(\bar{b}) \tag{11.4}$$

$$P(b) > P(\bar{b}) \longrightarrow \qquad \xi_{\rm tag} = -1, \quad \omega_{\rm evt} = 1 - P(b) \tag{11.5}$$

If only the information from two taggers is taken into consideration, as in our case, the combined tagging decision is always the one corresponding to the tagger that provided the lowest mistag probability, that is:

$$\xi(\xi_1, \xi_2, \omega_1, \omega_2) = \begin{cases} \xi_1 & \text{if} \quad \omega_1 \leq \omega_2 \\ \xi_2 & \text{if} \quad \omega_2 < \omega_1 \end{cases} . \tag{11.6}$$

Table 11.2: Tagging overlap logic.

Overlap	OS muon	OS electron	OS jet	SS
OS muon		X	X	✓
OS electron	X		X	✓
OS jet	X	X		✓
SS	✓	✓	✓	

[II] These *probabilities* may not be normalized.

11.3 FINAL TAGGING PERFORMANCE

The tagging performance is evaluated in the $B_s^0 \to J/\psi\,\phi(1020)$ data sample with the "$\sum P_{\text{evt}}$" method and taking into consideration the SS/OS combination. The resulting performance is reported in Tab. 11.3, while the background-subtracted $\mathcal{D}^2_{\text{evt}}$ distribution for $B_s^0 \to J/\psi\,\phi(1020)$ is shown in Fig. 11.1.

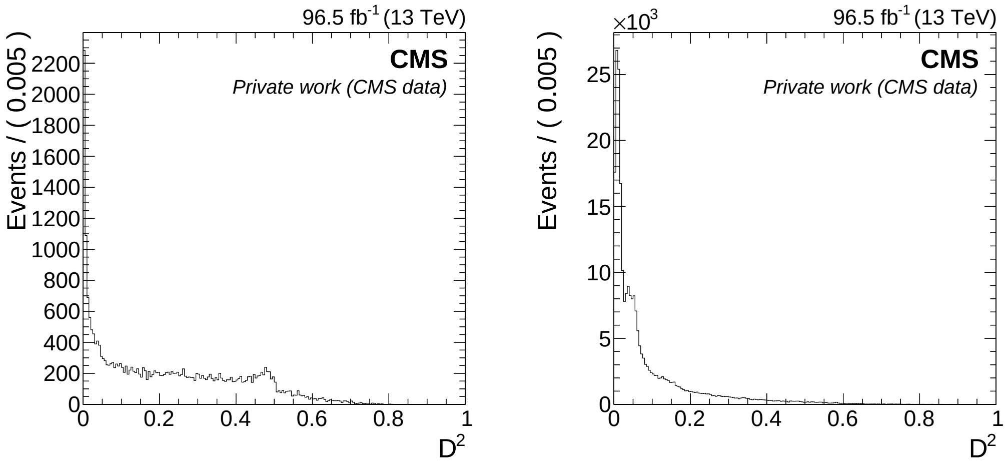
The combined performance of the tagging framework achieves a tagging power of $\hat{P}_{\text{tag}} = (5.59 \pm 0.02)\%$ with a tagging efficiency of $\varepsilon_{\text{tag}} = (55.9 \pm 0.1)\%$ and $\widehat{\mathcal{D}}^2_{\text{tag}} = 0.100$, which is approximately four times the tagging power available in the Run-1 CMS result [4] and among the highest tagging performances ever achieved in a hadronic collider experiment in the context of neutral B-mesons tagging. A comparison with the performance achieved by the ATLAS and LHCb collaboration in the same physics case (CPV in $B_s^0 \to J/\psi\,\phi(1020)$) is presented in Tab. 11.4.

Table 11.3: Calibrated tagging performance evaluated in the $B_s^0 \to J/\psi\,\phi(1020)$ data samples with the "$\sum P_{\text{evt}}$" method, averaged between *dataset_JpsiMuon* and *dataset_JpsiTrkTrk*. The "*only*" performance refers to the cases where only one tagger provides the tagging inference. The quoted uncertainties are statistical only. Note that the sum of the "*only* X" and "SS + X" entries are not expected to match the performances quoted in Tab. 11.1 due to the combination with the SS tagger which improves the tagging power.

Category	ε_{tag} [%]	$\widehat{\mathcal{D}}^2_{\text{tag}}$	$\hat{P}_{\text{tag}}$ [%]
Only OS muon	6.07 ± 0.05	0.212	1.29 ± 0.07
Only OS electron	2.72 ± 0.02	0.079	0.214 ± 0.004
Only OS jet	5.16 ± 0.03	0.045	0.235 ± 0.003
Only SS	33.12 ± 0.07	0.080	2.64 ± 0.01
SS+OS muon	0.62 ± 0.01	0.202	0.125 ± 0.003
SS+OS electron	2.77 ± 0.02	0.150	0.416 ± 0.005
SS+OS jet	5.40 ± 0.03	0.124	0.671 ± 0.006
Total	55.9 ± 0.1	0.100	5.59 ± 0.02

Table 11.4: Tagging performance comparison between what is presented in this work, ATLAS and LHCb in the context of CPV measurements in $B_s^0 \to J/\psi\,\phi(1020)$. Values are taken from the latest measurements [131, 135]. Uncertainties are omitted.

Result	ε_{tag} [%]	$\widehat{\mathcal{D}}^2_{\text{tag}}$	$\hat{P}_{\text{tag}}$ [%]
ATLAS [135]	21.2	0.082	1.75
this work	55.9	0.100	5.59
LHCb [131]	77.8	0.061	4.73



(a) Background-subtracted $\mathcal{D}^2_{\text{evt}}$ distribution in *dataset_JpsiMuon*.

(b) Background-subtracted $\mathcal{D}^2_{\text{evt}}$ distribution in *dataset_JpsiTrkTrk*.

Figure 11.1: Background-subtracted $\mathcal{D}^2_{\text{evt}}$ distributions as measured in the $B^0_s \to J/\psi\,\phi(1020)$ data sample.

Part IV

CPV ANALYSIS

The measurement of the CP violating phase ϕ_s in the $B_s^0 \to J/\psi\,\phi(1020)$ decay is one of the main CP violation analyses in CMS. The final state of the decay is accessible both to B_s^0 and $\overline{B}_s^0$ meson, and the process can thus violate CP in the interference between decay and mixing. The analysis is presented here not as a standalone measurement, but as a benchmark for the performance of the taggers described in the previous chapters. As such, only the tagging related systematic uncertainties are evaluated and compared to the statistical ones, to provide an estimate of the impact of the assumptions taken during the development of the taggers. The following chapters present an overview of the measurement, starting from the theory specific to the system, continuing with a description of the fit model used in the time-dependent angular analysis fit, and ending with the estimate of the tagging systematic uncertainties.

12

ANALYSIS OVERVIEW AND THEORETICAL BACKGROUND

The following chapters present one of the flagship analyses for CP violation in CMS, the measurement of the ϕ_s CP violating phase in $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-$ decays. While the analysis is not the focus of this dissertation, it was presented here as a benchmark of a possible use of the tagging framework described in Part III. The results presented in this chapters are part of an ongoing measurement in CMS, which, to avoid possible biases, are still blinded. The samples used have thus been reduced in statistics: only 20% of *dataset_JpsiTrkTrk* is used, where the number was chosen to equalize $N_{\mathrm{evt}} \cdot P_{\mathrm{tag}}$ in the two samples, as measured in Chapter 11[I]. *dataset_JpsiMuon* was not skimmed, since the corresponding result in CMS has already been published, as Ref. [4].

12.1 OVERVIEW

The $B_s^0 \to J/\psi\,\phi(1020)$ channel is particularly interesting for the study of CP violation. This process is characterized by a final state $(J/\psi\,\phi(1020))$ accessible to both B_s^0 and $\overline{B}_s^0$, with and without mixing, as schematically represented in Fig. 12.1. The decay presents interference between the direct and mixing CP violation. The predicted complex phase governing the process, ϕ_s (more precisely called $\phi_s^{c\bar{c}s}$), is one of the most theoretically clean CPV observables in the Standard Model. Fig. 12.2 shows the Feynman diagrams of the dominant tree-level amplitude and the higher-order penguin contribution. The $B_s^0 \to J/\psi\,\phi(1020)$ decay is relatively rare, with a branching ratio of $\mathcal{B}(B_s^0 \to J/\psi\,\phi(1020)) \approx 10^{-3}$ [24]. In the measurement presented in the following chapters, the decay is, in particular, reconstructed in the final state where the J/ψ decays to two muons, and ϕ mesons into two charged kaons. The branching ratio for these final states is $\mathcal{B}(J/\psi \to \mu^+\mu^-) \cdot \mathcal{B}(\phi \to K^+K^-) \approx 0.06 \cdot 0.5 \approx 0.03$ [24], leading to a total branching ratio of:

$$\mathcal{B}(B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-) \approx 3 \cdot 10^{-5}\,. \qquad (12.1)$$

[I] As a reminder, P_{tag} behaves like an effective efficiency for CP violation measurements, so the equalized number represents an effective statistic available in the sample.

$$\begin{tikzcd} B_s^0 \arrow[dd, leftrightarrow, "\frac{q}{p}"'] \arrow[dr, "A_f"] & \\ & J/\psi\,\phi \\ \overline{B}_s^0 \arrow[ur, "\overline{A}_f"'] & \end{tikzcd}$$

Figure 12.1: Schematic representation of the possible paths of B_s^0 mesons to the $J/\psi\,\phi(1020)$ final state.

12.2 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 [24]:

$$\begin{aligned} A_{f_{\text{CP}}} &= (V_{cb}^* V_{cs})t + \sum_{q=u,c,t} \left(V_{qb}^* V_{qs}\right) p_q \\ &= (V_{cb}^* V_{cs})(t + p_c - p_t) + (V_{ub}^* V_{us})(p_u - p_t) \\ &= (V_{cb}^* V_{cs})T + (V_{ub}^* V_{us})P\,, \end{aligned} \tag{12.2}$$

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. 12.2, the T and P contributions carry different weak phases, and the latter is 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 λ is the Wolfenstein expansion parameter. Given $\lambda \simeq 0.23$, the suppression factor is $\mathcal{O}(0.01)$.

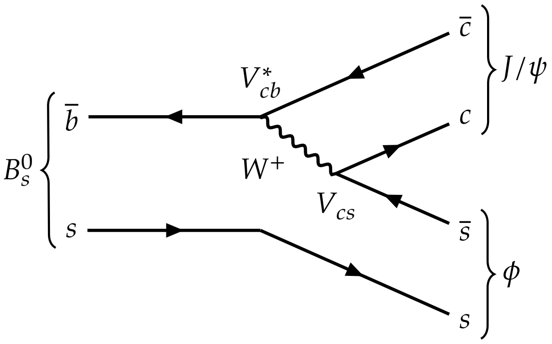
In the decay, penguin contributions can be neglected, as they are predicted [136] to cause a shift orders of magnitude smaller than the current experimental precision on ϕ_s [25]. Under this assumption only one amplitude contributes to the decay

$$\left|A_{f_{\text{CP}}}\right| = \left|\overline{A}_{\bar{f}_{\text{CP}}}\right|, \quad \text{with } f_{\text{CP}} \equiv J/\psi\,\phi(1020)\,. \tag{12.3}$$

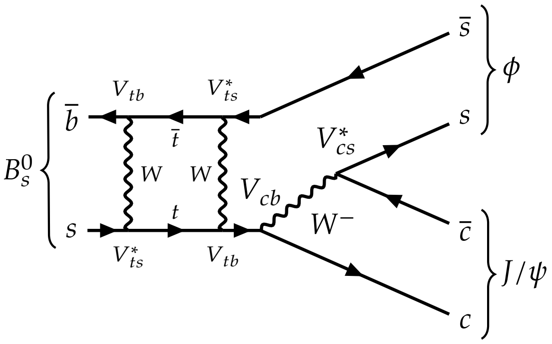
thus no direct CP violation is expected. This makes the $B_s^0 \to J/\psi\,\phi(1020)$ decay a *golden mode* decay. In golden mode decays, Eq. 2.54 can be simplified as

$$a_{\text{CP}}(t) = \frac{\text{Im}(\lambda_f)\sin(\Delta m t)}{\cosh\left(\frac{\Delta\Gamma t}{2}\right) + Re(\lambda_f)\sinh\left(\frac{\Delta\Gamma t}{2}\right)}\,. \tag{12.4}$$

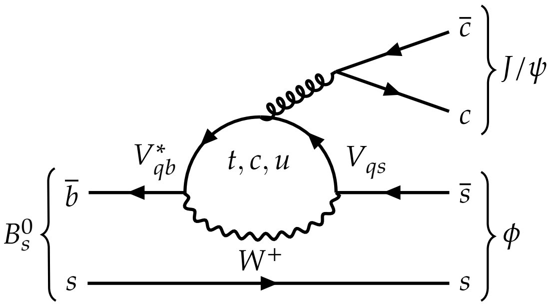
According to Eq. 12.4, the CP asymmetry in the golden modes is completely characterized by the real and imaginary components of $\lambda_{f_{\text{CP}}}$. It is worth noting that, for neutral B meson decays, the golden mode assumption is almost equivalent to



(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 is negligible in this dissertation.

Figure 12.2: Relevant Feynman diagrams for the $B_s^0 \to J/\psi\,\phi(1020)$ decay. The process is dominated by the tree amplitude with or without mixing.

$\left|\lambda_{f_{\mathrm{CP}}}\right| = 1$ since $|q/p| \approx 1$ in very good approximation[II]. To simplify the reading, the following definitions will be used hereafter:

$$\lambda_{J/\psi\phi} \equiv \lambda\,, \quad f_{J/\psi\,\phi(1020)} \equiv f\,, \tag{12.5}$$

unless otherwise specified.

Using Equations 2.46 and 2.48, the tree-amplitudes A_f and $\overline{A}_f$ can be expressed as:

$$\begin{aligned} A_f &= \quad \left|A_f\right| e^{i\theta_D} e^{i\phi_D}\,, \\ \overline{A}_f &= \eta_f \left|A_f\right| e^{i\theta_D} e^{-i\phi_D}\,. \end{aligned} \tag{12.6}$$

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

However, the CP eigenvalue η_f in the $B_s^0 \to J/\psi\,\phi(1020)$ decay is not uniquely determined, as the final state is a mixture of CP-even and CP-odd eigenstates. This is caused by the decay of the pseudoscalar meson with spin 0 into two vector mesons with spin 1, which allows three values of the final-state orbital momentum. To conserve the spin 0 of the initial state, the orbital momentum can take values of $\ell = 2$, 1 and 0, to balance, respectively, the final state spins pointing parallel, orthogonal, or opposite. Knowing this, the CP eigenvalue of the final state can be calculated from the value of the orbital momentum ℓ as:

$$CP\,|J/\psi\,\phi(1020)\rangle_\ell = \eta_f\,|\,J/\psi\,\phi(1020)\,\rangle_\ell = (-1)^\ell\,|\,J/\psi\,\phi(1020)\,\rangle_\ell\,. \tag{12.7}$$

The amplitudes can be expressed in terms of the associated CKM contributions to the tree-level transition as

$$\begin{aligned} A_f &= \quad V_{cs} V_{cb}^*\,, \\ \overline{A}_f &= \eta_f\, V_{cs}^* V_{cb}\,. \end{aligned} \tag{12.8}$$

and their ratio becomes

$$\frac{\overline{A}_f}{A_f} = \eta_f\, e^{-2i\phi_D} = \eta_f \frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*}\,, \tag{12.9}$$

Using Eq. 2.34 and 12.9, λ can be written as:

$$\lambda = \frac{q}{p}\frac{\overline{A}_f}{A_f} = \eta_f \left(\frac{V_{ts} V_{tb}^*}{V_{ts}^* V_{tb}}\right)\left(\frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*}\right) = \eta_f\, e^{-i\phi_s}\,, \tag{12.10}$$

[II] See Section 2.5.2.

where $\phi_s \equiv \phi_M + 2\phi_D$ is the phase associated with CP violation in interference between decay and mixing of $B_s^0 \to J/\psi\,\phi(1020)$. Using the properties of the complex argument, the CP-violating phase ϕ_s can be computed as[III]:

$$\begin{aligned} \phi_s &= -\arg\left(\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}\frac{V_{cs}^*V_{cb}}{V_{cs}V_{cb}^*}\right) \\ &= -2\arg\left(\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) \\ &= -2\left[\arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) - \pi\right] \\ &= -2\beta_s - 2\pi \\ &= -2\beta_s\,, \end{aligned} \tag{12.11}$$

where β_s is the B_s^0 unitary triangle angle defined in Eq. 2.16.

Equation 12.11 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 by a CKM global fit to experimental data. The current best determinations of $-2\beta_s$ are from the CKMfitter Group and UTfit Collaboration:

$$\phi_s = -2\beta_s = \begin{cases} -36.82^{+0.60}_{-0.86}\ \text{mrad} & \text{(CKMfitter [33, 137])} \quad (12.12) \\ -37 \pm 2\ \ \text{mrad} & \text{(UTfit [138])} \quad (12.13) \end{cases}$$

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

Combining the λ value from Eq. 12.10 into Eq. 12.4, the time-dependent CP asymmetry for the decay of a B_s^0 meson to a $J/\psi\,\phi(1020)$ final state is given by :

$$\begin{aligned} a_{\mathrm{CP}}(t) &= \frac{\Gamma_{\overline{B}_s^0 \to J/\psi\,\phi(1020)}(t) - \Gamma_{B_s^0 \to J/\psi\,\phi(1020)}(t)}{\Gamma_{\overline{B}_s^0 \to J/\psi\,\phi(1020)}(t) + \Gamma_{B_s^0 \to J/\psi\,\phi(1020)}(t)} \\ &= \frac{-\eta_f \sin\phi_s \sin(\Delta m_s t)}{\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f \cos\phi_s \sinh\left(\frac{\Delta\Gamma_s t}{2}\right)}\,. \end{aligned} \tag{12.14}$$

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

In this section, the differential decay rate used in the analysis of $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ events is described.

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

From Eq. 2.41 the time dependent decay rates are

$$\begin{aligned}\frac{\mathrm{d}\Gamma(B_s^0\to J/\psi\,\phi(1020))}{\mathrm{d}t} &\propto \left|A_f\right|^2\frac{1}{1+C}e^{-\Gamma_s t}\left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right)+\eta_f D\sinh\left(\frac{\Delta\Gamma_s t}{2}\right)\right.\\ &\qquad\left.+\,C\cos(\Delta m_s t)-\eta_f S\sin(\Delta m_s t)\right],\\ \frac{\mathrm{d}\Gamma(\overline{B}_s^0\to J/\psi\,\phi(1020))}{\mathrm{d}t} &\propto \left|A_f\right|^2\frac{1}{1+C}e^{-\Gamma_s t}\left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right)+\eta_f D\sinh\left(\frac{\Delta\Gamma_s t}{2}\right)\right.\\ &\qquad\left.-\,C\cos(\Delta m_s t)+\eta_f S\sin(\Delta m_s t)\right].\end{aligned}\tag{12.15}$$

Substituting Eq. 12.10 into Eq. 2.43 the coefficients C, S and D become

$$C=\frac{1-|\lambda|^2}{1+|\lambda|^2}\,,\quad S=-\frac{2|\lambda|\sin\phi_s}{1+|\lambda|^2}\quad\text{and}\quad D=-\frac{2|\lambda|\cos\phi_s}{1+|\lambda|^2}\,,\tag{12.16}$$

where η_f was factorized out into Eq. 12.15. The decay rates in Eq. 12.15 are invariant under the following simultaneous transformations:

$$\begin{cases}\phi_s & \leftrightarrow \quad \pi-\phi_s\,,\\ \Delta\Gamma_s & \leftrightarrow \quad -\Delta\Gamma_s\,.\end{cases}\tag{12.17}$$

In this thesis $\Delta\Gamma_s$ is assumed to be positive to follow the convention set by similar measurements in other experiments[139].

As shown in Eq. 12.7, η_f depends on the final-state orbital momentum as:

$$\eta_f=(-1)^\ell\quad\text{for}\quad\ell=0,\,1,\,2\,.\tag{12.18}$$

The decay rate from Eq. 12.15, which contains η_f, is then only valid for a single value of orbital momentum, and for a proper measurement the different CP eigenstates need to be separated statistically by performing an angular analysis on the final-state particle system.

12.3.1 *The transversity basis*

The three different CP eigestates of the final have different angular distributions for the final-state particles. In the analysis, the full $B_s^0\to J/\psi\,\phi(1020)$ decay amplitude is decomposed using the transversity basis [140]. The decomposition uses three polarization amplitudes(A_0, $A_\parallel$ and $A_\perp$) matched to three strong phases (δ_0, $\delta_\parallel$ and $\delta_\perp$). The amplitudes, shown in Fig. 12.3, are:

- the CP-even amplitude A_0, corresponding to the longitudinal polarization, where the spins are both aligned to the vector mesons momenta;

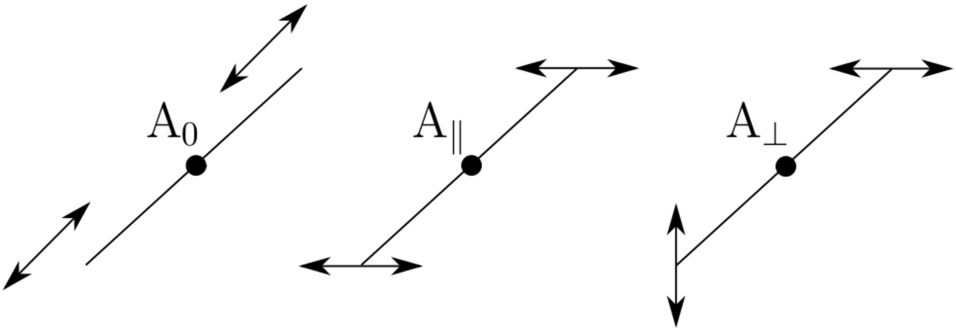


Figure 12.3: Schematic representation of the polarization amplitudes of the $J/\psi\,\phi(1020)$ system in the transversity basis. The black circle represents the B_s^0 meson, while the straight lines and the short arrows represent the final-state vector meson momenta and spin orientations, respectively.

- 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.

The three amplitudes obey the following normalization condition:

$$|A_0|^2 + |A_\perp|^2 + |A_\parallel|^2 = 1\,. \tag{12.19}$$

Each of the amplitudes of the transversity basis model the resonant $B_s^0 \to J/\psi\,\phi(1020)$ decays. In a real sample, however, there are also contributions 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, which needs to be taken into account with an additional S-wave amplitude A_S and an associated strong phase δ_S.

For disentangling the eigenstates, the transversity bases uses the angles between the decay products, with a parametrization defined in the products rest frame as $\Theta = (\theta_T, \psi_T, \varphi_T)$ [140], shown in Fig. 12.4. 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.

12.3.2 *The $B_s^0 \to J/\psi\,\phi(1020)$ differential decay rate*

The full decay rate, with all CP eigestates accounted for, can be written in the transversity basis as

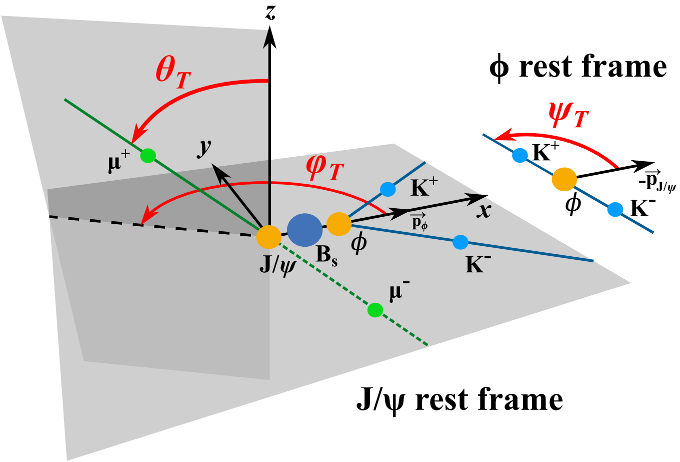


Figure 12.4: Definition of the three angles of the transversity basis θ_T, ψ_T, and φ_T describing the topology of the $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-$ decay. Figure from Ref. [141].

$$\frac{\mathrm{d}^4\Gamma(B_s^0)}{\mathrm{d}\Theta\,\mathrm{d}t} = f(\Theta, t\,|\,\alpha) \propto \sum_{i=1}^{10} O_i(\alpha, t)\cdot g_i(\Theta)\,, \tag{12.20}$$

where:

- O_i are time-dependent functions:

$$\begin{aligned} O_i(\alpha, t) = N_i e^{-\Gamma_s t}\bigg[& a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ & + c_i\xi_i \cos(\Delta m_s t) + d_i\xi_i \sin(\Delta m_s t)\bigg], \end{aligned} \tag{12.21}$$

  with N_i, a_i, b_i, c_i, and d_i defined in Tab. 12.1 and ξ_i is a number representing the flavor of the meson ($+1$ for B_s^0 and -1 for $\overline{B}_s^0$),

- g_i are the angular functions defined in Tab. 12.2,

- $\alpha = (\phi_s, \Delta\Gamma_s, \Gamma_s, \Delta m_s, |\lambda|, |A_0|, |A_\perp|, |A_\parallel|, |A_S|, \delta_0, \delta_\parallel, \delta_\perp, \delta_S)$ is the set of physics parameters,

- the coefficients C, S, and D in Tab. 12.1 are defined as in Eq. 12.16:

$$C = \frac{1-|\lambda|^2}{1+|\lambda|^2}\,, \quad S = -\frac{2|\lambda|\sin\phi_s}{1+|\lambda|^2} \quad \text{and} \quad D = -\frac{2|\lambda|\cos\phi_s}{1+|\lambda|^2}\,. \tag{12.22}$$

C only depends on the amount of direct CP violation, and is equal to zero if, as expected, $|\lambda| = 1$. The coefficients S and D depend respectively on the sine and cosine of ϕ_s, and are then sensitive to the CP violation in the interference. Given

Table 12.1: Time-dependent terms of the $B_s^0 \to J/\psi\,\phi(1020)$ 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. 12.22.

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 12.2: Angular terms of the $B_s^0 \to J/\psi\,\phi(1020)$ differential decay rate model. The transversity basis angles are defined in Fig. 12.4.

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	$\frac{4}{3}\sqrt{3}\cos\psi_T(1 - \sin^2\theta_T\cos^2\varphi_T)$

that ϕ_s is expected to be small ($\mathcal{O}(10\,\text{mrad})$) in the Standard Model, only S is expected to show a significant dependence on ϕ_s.

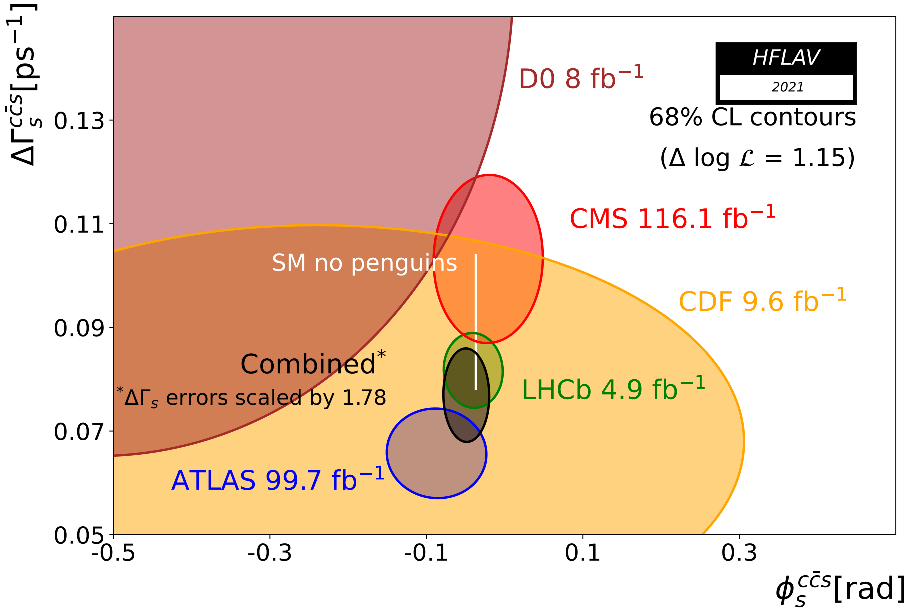
12.4 STATE-OF-THE-ART

The CP-violating phase ϕ_s has been measured at the Fermilab Tevatron experiments [142–146], and at the LHC by the ATLAS, CMS, and LHCb experiments [4, 131, 135, 147–154], using $B_s^0 \to J/\psi\,\phi(1020)$, $B_s^0 \to J/\psi f_0(980)$, 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 [155, 156].

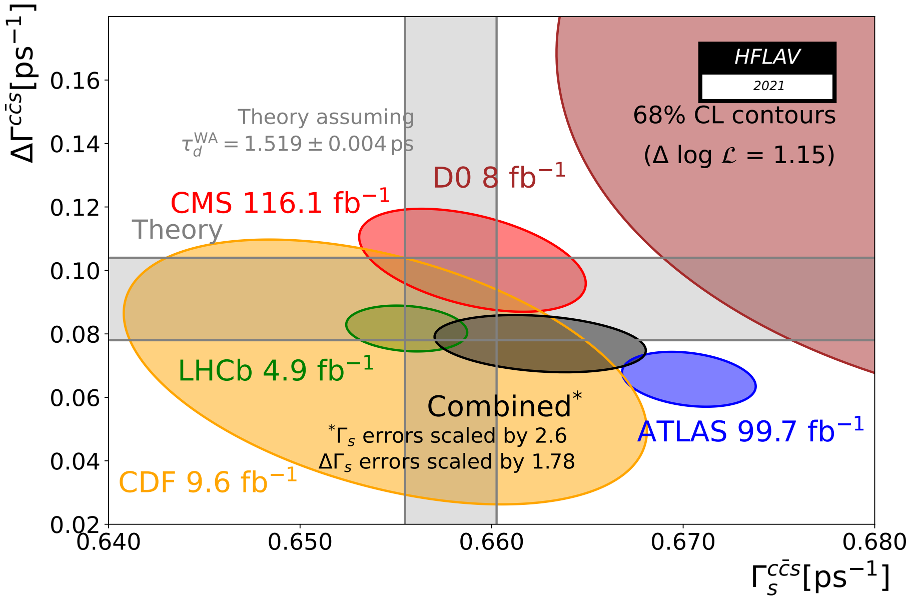
With the same analysis used to extract ϕ_s, several other parameters of the differential decay rate model can be measured, with Γ_s, $\Delta\Gamma_s$ being among the most interesting ones. The available theoretical predictions for the physics parameters of Eq. 12.20 are reported in Tab. 12.3, while the most recent world-average values are reported in Tab. 12.4. The strong phase δ_0 is not reported, as the general convention [143, 145] is to set it at zero, since only phase differences are physically meaningful (see Tab. 12.1). The two-dimensional world average contours for ϕ_s and $\Delta\Gamma_s$, and $\Delta\Gamma_s$ and Γ_s are shown in Fig. 12.5.

Table 12.3: Available theoretical predictions for the physics parameters describing the $B_s^0 \to J/\psi\,\phi(1020)$ decay rate.

Parameter	Prediction	Ref.
ϕ_s	$-36.82\,^{+0.60}_{-0.86}$ mrad	[33]
$\Delta\Gamma_s$	$0.091 \pm 0.013\,\text{ps}^{-1}$	[157]
Δm_s	$18.77 \pm 0.86\,\hbar\text{ps}^{-1}$	[157]
$\lvert\lambda\rvert$	≈ 1	SM



(a)



(b)

Figure 12.5: Two-dimensional (ϕ_s, $\Delta\Gamma_s$) (a) and (Γ_s, $\Delta\Gamma_s$) (b) 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. [137, 157]. Figures from[25].

Table 12.4: World-average values for the physics parameters describing the $B_s^0 \to J/\psi\,\phi(1020)$ decay rate. The strong phase δ_S, related to the S-wave, is given as δ_S- $\delta_\perp$, since it is usually fitted in this way to reduce correlations. The strong phase δ_0 is not given as it is usually put to zero as a general convention.

Parameter	World-average	Ref.
ϕ_s	-49 ± 19 mrad	[25]
$\Delta\Gamma_s$	0.084 ± 0.005 ps^{-1}	[25]
Γ_s	0.6578 ± 0.0024 ps^{-1}	[25]
Δm_s	17.765 ± 0.006 $\hbar$ps^{-1}	[25]
$\lvert\lambda\rvert$	1.001 ± 0.018	[158]
$\lvert A_0\rvert^2$	0.520 ± 0.003	[25]
$\lvert A_\parallel\rvert^2$	0.2222 ± 0.0027	[25]
$\lvert A_\perp\rvert^2$	0.253 ± 0.006	[25]
$\lvert A_S\rvert^2$	0.030 ± 0.005	[25]
$\delta_\parallel$	3.18 ± 0.06 rad	[25]
$\delta_\perp$	3.08 ± 0.12 rad	[25]
δ_S- $\delta_\perp$	0.23 ± 0.05 rad	[25]

12.4.1 *New-physics contributions*

The presence of new physics (NP) can modify ϕ_s via the addition of new, beyond Standard Model (BSM), particles which can contribute to the B_s^0–$\overline{B}_s^0$ mixing diagrams shown in Fig. 2.3 [159, 160]. The effect of these particles can be parametrized by adding an arbitrary complex parameter Δ_s in the off-diagonal matrix element of the mass matrix M_{12}:

$$M_{12} = M_{12}^{\mathrm{SM}} |\Delta_s| e^{i\phi_s^{\Delta}} , \tag{12.23}$$

written here in exponential form to separate the contributions from the module and argument. Any $\Delta_s \neq 1$ would correspond to new-physics effects, where differences in absolute value would affect Δm_s, and different in phase would affect ϕ_s

$$\Delta m_s = 2|M_{12}^{\mathrm{SM}}||\Delta_s| , \tag{12.24}$$

$$\phi_s = \phi_s^{\mathrm{SM}} + \phi_s^{\Delta} . \tag{12.25}$$

Comparing the experimental results with the SM predictions, one can obtain bounds on the complex value of Δ_s. In the most recent result of such an investigation, shown in Fig. 12.6, a good agreement was found between the SM global fit and $\Delta_s = 1 + 0i$. In Fig. 12.6, the orange ring denotes the constraint to $|\Delta_s|$ given by the mass difference Δm_s, while the purple region represents the contraint to ϕ_s^{Δ} given by the measurement of ϕ_s. The size of the orange region is governed mostly by theoretical uncertainties, while the purple one is limited by the precision of the measurements. In this formalism, new-physics effects are still allowed to modify ϕ_s by several degrees [160, 161], 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.

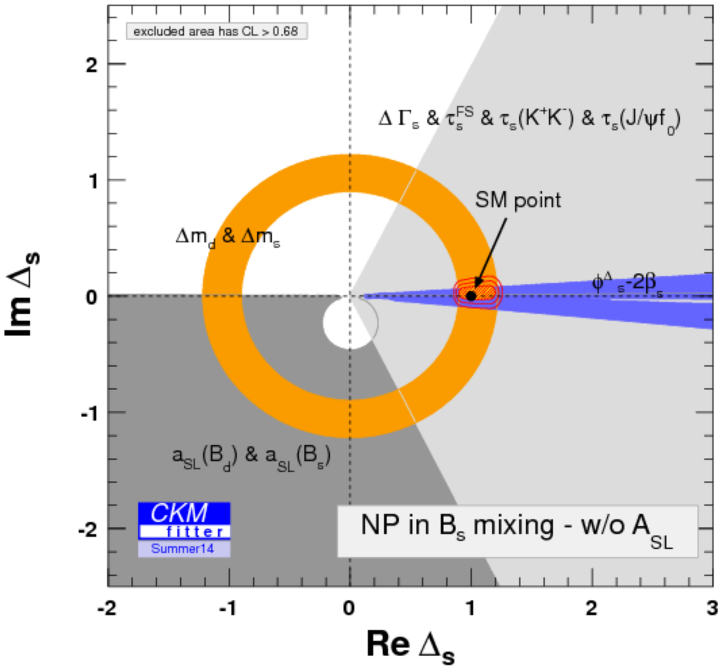


Figure 12.6: Constraint of the NP parameter Δ_s. The constraint due to the mass difference Δm_s is denoted by the orange ring, while the constraint due to ϕ_s by the violet triangular region. The red area shows the combined fit to experimental data. Figure from Ref. [161].

13

MAXIMUM LIKELIHOOD FIT

The physics parameters of interest that describe the $B_s^0 \to J/\psi\,\phi(1020)$ decay are extracted with an unbinned multidimensional extended maximum likelihood (UML) fit to the combined 2017+2018 *dataset_JpsiMuon* and *dataset_JpsiTrkTrk* data samples, where *dataset_JpsiTrkTrk* was skimmed randomly to keep only 20% of the events. In order to properly model the decay rate described in Eq. 12.20, the following observables are used as input:

- the three angular variables $\Theta = (\cos\theta_T, \cos\psi_T, \varphi_T)$ describing the event topology in the transversity basis[I];
- the flavor decision ξ_{tag} and the mistag probability ω_{tag} [II];
- the proper decay length of the B_s^0 candidate ct and its uncertainty σ_{ct};
- the B_s^0 candidate invariant mass $m_{B_s^0}$.

From the UML fit, the following physics parameters are extracted: ϕ_s, $\Delta\Gamma_s$, Γ_s, Δm_s, $|\lambda|$, $|A_0|^2$, $|A_\perp|^2$, $|A_S|^2$, $\delta_\parallel$, $\delta_\perp$, and $\delta_{S\perp}$, where $\delta_{S\perp}$ is defined as $\delta_S - \delta_\perp$.

With respect to the decay rate described by Eq. 12.20, the number of free parameters is reduced by two. The value of δ_0 is set to zero since only phase differences are physically meaningful, and the normalization relationship $|A_0|^2 + |A_\perp|^2 + |A_\parallel|^2 = 1$ is used to redefine $|A_\parallel|^2 = 1 - |A_0|^2 - |A_\perp|^2$. The value of $\Delta\Gamma_s$ is also limited to be positive, based on the LHCb measurement in Ref. [139]. The fit input observables, the main physics parameters extracted, and the conventions used are summarized in Tab. 13.1.

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.

13.1 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 the continuous equivalent of the value probability in discrete-valued

---

[I] Θ is here redefined from $(\theta_T, \psi_t, \phi_T)$ to $(\cos\theta_T, \cos\psi_T, \varphi_T)$ since these are the observables used in the fit.

[II] The event mistag probability ω_{evt} is here redefined to ω_{tag} to ease the reading and match ξ_{tag}.

Table 13.1: Summary of the UML fit input observables, main physics parameters to be determined and used assumption.

Input observables	Fitted parameters	Constraints
$\theta_T, \psi_T, \varphi_T$	$\phi_s, \Delta\Gamma_s, \Gamma_s, \Delta m_s, \lvert\lambda\rvert$	$\Delta\Gamma_s > 0$
$ct, \sigma_{ct}, m_{B_s^0}$	$\lvert A_\perp\rvert^2, \lvert A_0\rvert^2, \lvert A_S\rvert^2$	$\lvert A_\parallel\rvert^2 = 1 - \lvert A_0\rvert^2 - \lvert A_\perp\rvert^2$
$\xi_{\text{tag}}, \omega_{\text{evt}}$	$\delta_\parallel, \delta_\perp, \delta_{S\perp}$	$\delta_0 = 0, \delta_{S\perp} = \delta_S - \delta_\perp$

random variables. In a continuous variable, the probability of a single real value is meaningless, but the probability of a range of values is not. PDFs are defined so that the integral of the function in a range is equal to the probability of a value falling in that range.

$$\int_a^b = \text{Prob}(a < x < b) \tag{13.1}$$

$$P(a) = \frac{\mathrm{d}}{\mathrm{d}x}\text{Prob}(x < a) \tag{13.2}$$

By definition, probability density functions can only assume positive values and are normalized to unity in the definition range.

The *likelihood function* measures how likely a specific set of values sampled from a statistical model (with possibly unknown parameters) is. Assuming that the points in the sample are independent from each other, the likelihood of n observations $\vec{x} = (x_1, \ldots, x_n)$ distributed following a PDF $f(\vec{x}\,|\,\vec{\theta})$ is the product of the PDF values evaluated at each event

$$\mathcal{L}(\vec{\theta}\,|\,\vec{x}) = \prod_{i=1}^{n} f(\vec{x}_i\,|\,\vec{\theta})\,. \tag{13.3}$$

Since the $\vec{x}$ are fixed in Eq. 13.3, the likelihood behaves as a function[III] of the parameters $\vec{\theta}$. The likelihood can be used to estimate the values of the parameters, since it is maximized by the combination of model parameters values for which the joint probability of all the observables is maximal [162][IV]. The approach of maximizing $\mathcal{L}$ to estimate the parameters $\vec{\theta}$ is called *maximum likelihood estimation* (MLE):

$$\hat{\theta} = \arg\max \mathcal{L}(\vec{\theta}\,|\,\vec{x}) \quad \leftrightarrow \quad \left.\frac{\mathrm{d}\mathcal{L}(\vec{\theta}\,|\,\vec{x})}{\mathrm{d}\vec{\theta}}\right|_{\hat{\theta}} = 0\,. \tag{13.4}$$

The product in Eq. 13.3 can be performed using either the single data points, in which case the MLE is referred to as *unbinned*, or binning the sample and using the Poisson probabilities of the bin contents. Unbinned MLEs are more precise, but also computationally heavier than their binned counterpart, since the evaluation time scales with the sample size and not with the number of bins.

[III] *Not* as PDF, since it is not normalized.

[IV] It is not guaranteed that the likelihood function has a maximum.

The maximization of Eq. 13.3 is complex, so it is usually preferred to minimize the negative log-likelihood (NLL) function, defined as

$$-\ln \mathcal{L}(\vec{\theta}\,|\,\vec{x}) = -\sum_{i=1}^{n} \ln f(x_i\,|\,\vec{\theta})\,. \tag{13.5}$$

Because the natural logarithm is a monotonic function, minimizing the NLL is equivalent to maximizing the likelihood, with the advantages of dealing with sums instead of products in the derivative in Eq. 13.4.

Eq. 13.5 can also be modified to include event weights, whereas the logarithmic sum over the points can be weighted by an arbitrary real number. The minimization procedure is the same, but this has important consequences in the computation of the parameter uncertainties.

In this work, NLL functions are constructed with the RooFit library [115] and minimized with the Minuit minimization framework [163].

13.1.1 *Extended term*

Regular ML estimation uses only shape information, since the PDF of each event is normalized. If the number of events in the sample is of interest, an *extended* maximum likelihood (EML) term [164] 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.

The extended term is simply a Poisson distribution of the observed number of events given the expected one, and is multiplied in Eq. 13.3

$$\mathcal{L}(\nu, \vec{\theta}\,|\,\vec{x}) = \frac{e^{-\nu}\nu^n}{n!}\mathcal{L}(\vec{\theta}\,|\,\vec{x}) = \frac{e^{-\nu}\nu^n}{n!}\prod_{i=1}^{n} f(x_i\,|\,\vec{\theta})\,, \tag{13.6}$$

$$-\ln \mathcal{L}(\nu, \vec{\theta}\,|\,\vec{x}) = \nu - n\ln\nu - \sum_{i=1}^{n} \ln f(x_i, \vec{\theta}) + \text{const.}\,, \tag{13.7}$$

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:

$$\frac{\mathrm{d}[-\ln \mathcal{L}(\nu, \vec{\theta}\,|\,\vec{x})]}{\mathrm{d}\nu} = 1 - \frac{n}{\nu} = 0 \quad \rightarrow \quad \hat{\nu} = n\,. \tag{13.8}$$

13.1.2 *Uncertainties evaluation*

In an unweighted MLE, the uncertainties on the best values of the model parameters can be estimated by expanding the NLL around its minimum $\vec{\theta} = \hat{\theta}$, 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[V]:

$$-\ln \mathcal{L}(\theta) = f(\theta) \approx f(\hat{\theta}) + \underbrace{\left.\frac{\mathrm{d}f(\theta)}{\mathrm{d}\theta}\right|_{\hat{\theta}} (\theta - \hat{\theta})}_{=0} + \frac{1}{2} \left.\frac{\mathrm{d}^2 f(\theta)}{\mathrm{d}\theta^2}\right|_{\hat{\theta}} (\theta - \hat{\theta})^2 + \mathcal{O}\big((\theta - \hat{\theta})^3\big) . \tag{13.9}$$

Using Eq. 13.9, the likelihood function at the minimum can be evaluated as:

$$\mathcal{L}(\theta) = e^{-f(\theta)} \approx e^{-f(\hat{\theta})} \cdot e^{-\frac{1}{2} \left.\frac{\mathrm{d}^2 f(\theta)}{\mathrm{d}\theta^2}\right|_{\hat{\theta}} (\theta - \hat{\theta})^2} \cdot e^{-\mathcal{O}\left((\theta - \hat{\theta})^3\right)} . \tag{13.10}$$

If the shape of the likelihood is *sufficiently* Gaussian around its maximum, that is, if the higher order terms can be neglected, Eq. 13.10 can be expressed as:

$$\mathcal{L}(\theta) \approx \mathcal{L}_{\mathrm{max}} \cdot e^{-\frac{(\theta - \hat{\theta})^2}{2\sigma_\theta^2}} , \tag{13.11}$$

where σ_θ^2 is the variance of θ, estimated as:

$$\sigma_\theta^2 = \left(\left.\frac{\partial^2 - \ln \mathcal{L}(\theta)}{\partial \theta^2}\right|_{\hat{\theta}} \right)^{-1} , \tag{13.12}$$

and $\mathcal{L}_{\mathrm{max}} = e^{\ln \mathcal{L}(\hat{\theta})} = \mathcal{L}(\hat{\theta})$.

From Eq. 13.11 it immediately follows that the NLL near its minimum is approximated by the parabolic expression:

$$-\ln \mathcal{L}(\theta) \approx -\ln \mathcal{L}_{\mathrm{max}} + \frac{(\theta - \hat{\theta})^2}{2\sigma_\theta^2} , \tag{13.13}$$

which satisfies the following important relationships:

$$-\ln \mathcal{L}(\hat{\theta} \pm \sigma_\theta) = -\ln \mathcal{L}_{\mathrm{max}} + 0.5 , \tag{13.14}$$

$$-\ln \mathcal{L}(\hat{\theta} \pm 2\sigma_\theta) = -\ln \mathcal{L}_{\mathrm{max}} + 2.0 . \tag{13.15}$$

If the NLL is not parabolic, then $\hat{\theta} \pm \sigma_\theta$ is not a good estimate of 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 13.14 and 13.15, as illustrated in Fig. 13.1[VI]. 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 its

---

[V] The dependence on the observables x is dropped from the following equations for simplicity.

[VI] This construction does not guarantee adequate coverage but is not used in this dissertation.

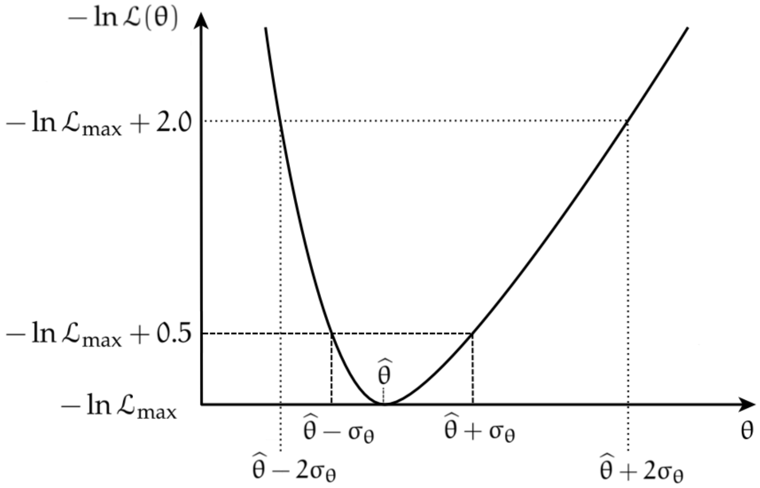


Figure 13.1: Confidence intervals (CI) construction for ML estimators with non-parabolic NLL function. The construction procedure for the 68% and 95% CIs is indicated by the dashed and dotted lines, respectively.

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[VII]:

$$\boldsymbol{H}_{ij} = \left.\frac{\partial^2 \ln \mathcal{L}(\vec{\theta})}{\partial\theta_i \partial\theta_j}\right|_{\hat{\theta}}, \tag{13.16}$$

which is then used to calculate both the parameter uncertainty and the covariance, as:

$$\mathrm{cov}(\theta_i, \theta_j) = \left(\boldsymbol{H}^{-1}\right)_{ij}, \tag{13.17}$$

where $(\boldsymbol{H}^{-1})$ 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[VIII].

PARAMETER UNCERTAINTIES IN WEIGHTED FITS In case of a weighted MLE, the formulas defined in Eq. 13.12 and Eq. 13.17 for parameter uncertainties stop working. This can be easily verified by adding a constant weight w to the NLL: w does not affect the estimation of the parameters, but behaves as a scale factor for the uncertainties.

[VII] H_{ij} indicates the elements of the Hessian matrix H.

[VIII] More details on the exact procedures are contained in the MINUIT documentation [165].

Several methods are available for correcting the Hessian in presence of weights, but not all of them are guaranteed to be asymptotically correct. In this thesis, two main methods were considered.

- The asymptotically correct method from Ref. [166], where the covariance matrix can be written as

  $$\boldsymbol{C}_{ij} = \sum_{k,l}^{N_{\text{par}}} \boldsymbol{H}^{-1}_{ik} \boldsymbol{D}_{kl} \boldsymbol{H}^{-1}_{lj} \quad \text{with} \quad \boldsymbol{D}_{kl} = \sum_{i=1}^{n} w_i^2 \, \frac{\partial \ln f(x_i \,|\, \vec{\theta})}{\partial \theta_k} \frac{\partial \ln f(x_i \,|\, \vec{\theta})}{\partial \theta_l} \,. \tag{13.18}$$

  This method is directly available in the used RooFit framework, but was found to be extremely slow, inflating the fitting time of a factor 20.

- The parametric bootstrap method from Ref. [167]. In this method, once the parameters $\hat{\theta}$ are estimated, the resulting PDF $\hat{f}$ is sampled n times, obtaining n "replicas". The minimization procedure is repeated in each replica $i \in 1..n$, obtaining a new estimate $\hat{\theta}^*_i$. The set of $\hat{\theta}^*$, called the bootstrap sample, can then be used to produce confidence intervals for $\hat{\theta}$. A simple method for confidence intervals is based on normal intervals, which at 1σ are

  $$[a, b] = \hat{\theta} - \hat{b} \pm s_{\text{sample}}\left(\hat{\theta}^*\right) , \tag{13.19}$$

  where s^2_{sample} is the unbiased sample variance of the bootstrap sample and $\hat{b}$ is the estimated bias in the bootstrap sample (the average). This method assumes that the distribution of $\hat{\theta}^*_i$ is a gaussian, and is inaccurate otherwise. Another method, unaffected by the gaussian assumption, is the percentile interval method, where the confidence interval bounds are

  $$\begin{aligned} a &= \hat{\theta}^*_{\frac{\alpha}{2}} \quad - \hat{b} \\ b &= \hat{\theta}^*_{1-\frac{\alpha}{2}} - \hat{b} \,, \end{aligned} \tag{13.20}$$

  where $\hat{\theta}^*_{\frac{\alpha}{2}}$ and $\hat{\theta}^*_{1-\frac{\alpha}{2}}$ are, respectively, the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles in the bootstrap sample, with α being the requested confidence level ($\sim 1 - 0.683$ for 1σ intervals). The quoted confidence intervals are statistical only, and a specific systematic uncertainty needs to be associated with the measured bias $\hat{b}$.

13.2 FIT MODEL

The fit model is made up of probability density functions (PDF) describing the B^0_s signal (P_{sig}), the main (hereafter, "combinatorial") background (P_{bkg}), and the background from $B^0 \to J/\psi K^*(892)^0$ (hereafter B^0 background) where the pion is misidentified as a kaon ($P_{\text{bkg}B^0}$). 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 angular efficiency function is included in the signal model, whereas the proper decay length efficiency function is not and is instead used as an event weight.

The event PDF is defined as:

$$P = f_{\text{sig}} P_{\text{sig}} + f_{\text{bkg}} P_{\text{bkg}} + f_{\text{bkg}B^0} P_{\text{bkg}B^0} \,, \tag{13.21}$$

where

$$P_{\text{sig}} = \varepsilon(\Theta) \left[\tilde{f}(\theta,\, ct \,|\, \alpha,\, \xi_{\text{tag}},\, \omega_{\text{tag}}) \otimes G(ct,\, \sigma_{ct}) \right] P_{\text{sig}}(m_{B_s^0})\, P_{\text{sig}}(\sigma_{ct})\, P_{\text{sig}}(\omega_{\text{tag}}) \,, \tag{13.22}$$

$$P_{\text{bkg}} = \left[P_{\text{bkg}}(ct) \otimes G(ct,\, \sigma_{ct}) \right] P_{\text{bkg}}(\Theta)\, P_{\text{bkg}}(m_{B_s^0})\, P_{\text{bkg}}(\sigma_{ct})\, P_{\text{bkg}}(\omega_{\text{tag}}) \,, \tag{13.23}$$

$$P_{\text{bkg}B^0} = \left[P_{\text{bkg}B^0}(ct) \otimes G(ct,\, \sigma_{ct}) \right] P_{\text{bkg}B^0}(\Theta)\, P_{\text{bkg}B^0}(m_{B_s^0})\, P_{\text{bkg}B^0}(\sigma_{ct})\, P_{\text{bkg}B^0}(\omega_{\text{tag}}) \,. \tag{13.24}$$

The individual terms of the products are explained in the following sections. The corresponding negative log-likelihood is equal to:

$$-\ln \mathcal{L} = -\sum_{i=0}^{N_{\text{evt}}} \ln P_i + N_{\text{tot}} - N_{\text{evt}} \ln N_{\text{tot}} \,, \tag{13.25}$$

where the last two terms represent the extended term, used in the estimation of the yields.

The *effective* fractions of the three fit components are defined as:

$$f_l = \frac{N_l}{N_{\text{tot}}} \,, \quad \text{with } l = \left(\text{sig, bkg, bkg}B^0\right) . \tag{13.26}$$

N_{sig} and N_{bkg} are free-floating parameters and can be interpreted as the sum of the weights $w_{\text{evt}} = 1/\varepsilon(ct)_{\text{evt}}$ belonging to a given category, while $N_{\text{tot}} = \sum_l N_l$ is their sum. The total weight of candidates selected in data is indicated by N_{evt} and is constant. The parameter $N_{\text{bkg}B^0}$ related to the effective yield of the B^0 background is fixed to the value estimated from a pre-fit, as it will be explained in Section 13.2.2. Note that, due to the weighted fit, the parameters N_l and N_{tot} do not correspond to real event yields. This is the reason why effective fractions f_l are preferred as a presentation tool. Also note that the usage of weighted events will inevitably change the scale of the y-axis in all fit-related plots.

The PDF $\tilde{f}(\theta,\, ct \,|\, \alpha,\, \xi_{\text{tag}},\, \omega_{\text{tag}})$ is the differential decay rate function $f(\theta, ct|\alpha)$ defined in Eq. 12.20, modified to include the tagging information from the taggers described in Part III and the S-P wave interference described later in Section 13.2.1. All the assumptions in Tab. 13.1 are used.

The efficiency function $\varepsilon(\Theta)$ and the ct resolution function $G(ct,\, \sigma_{ct})$ will be defined in Section 13.2.7 and Section 13.2.5, respectively. The other PDFs are described in the following sections and summarized in Tab. 13.2. Unless otherwise specified, parameters in the individual PDFs do not have any relationship with parameters sharing the same name used in PDFs or part of this dissertation. Unless otherwise specified, all parameters are left free to float.

Note that in Eq. 13.22, Eq. 13.23, and Eq. 13.24 no probability for ξ_{tag} is provided. The fractions f_{sig}, f_{bkg} and $f_{\text{bkg}B^0}$ are instead split for each possible value of ξ_{tag} (0 for untagged values, ± 1 for events tagged as B_s^0 or $\overline{B}_s^0$).

The model presented in Eq. 13.21 is replicated in each of the four data sets, with the likelihood being summed to obtain the full likelihoods:

$$-\ln \mathcal{L}_{\text{tot}} = \sum_d (-\ln \mathcal{L}_d)\,, \tag{13.27}$$

where d is an index that runs over the four dataset (*dataset_JpsiMuon* 2017, *dataset_JpsiMuon* 2018, *dataset_JpsiTrkTrk* 2017, *dataset_JpsiTrkTrk* 2018). The only PDF with shared parameters between the four data sets is the decay rate model $\tilde{f}(\theta, ct\,|\,\alpha, \xi_{\text{tag}}, \omega_{\text{tag}})$.

Table 13.2: Summary of the PDFs used to parametrize the signal and background distributions in Eq. 13.21. The sum of two distributions is indicated with "2×". "Decay rate model" indicates that the variable is described by the $B_s^0 \to J/\psi\,\phi(1020)$ differential decay rate $\tilde{f}(\theta, ct\,|\,\alpha, \xi_{\text{tag}}, \omega_{\text{tag}})$. Unless stated otherwise all the distributions are independent and do not share any parameter. "Shared with signal" in the B^0 background column means that in the final fit the two pdfs are constrained to be the same and share the parameters.

Variable	Signal	Comb. background	B^0 background
$m_{B_s^0}$	Johnson's S_U	Exponential	Johnson's S_U
ct	Decay rate model	2×Exponential	Exponential
Θ	Decay rate model	Bernstein polynomials	Bernstein polynomials
σ_{ct}	2×Gamma	2×Gamma	Shared with signal

13.2.1 *Signal model*

The signal model has been developed in $B_s^0 \to J/\psi\,\phi(1020)$ simulated samples. Several PDF choices have been tested for all variables. The chosen distributions are the ones found to best model each variable in the simulated events.

The PDFs used to describe the signal are:

- **Decay rate -** $\tilde{f}(\theta, ct\,|\,\alpha, \xi_{\text{tag}}, \omega_{\text{tag}}) \otimes G(ct, \sigma_{ct})$, described in Section 12.3.2 and corrected for the effects of time resolution, tag dilution, and time and angular acceptances

- **Invariant mass -** The mass distribution is described by a single Johnson's S_U distribution. The Johnson's S_U distribution is obtained by transforming a normally distributed variable x into

$$x \to z = \gamma + \delta \sinh^{-1}\left(\frac{x-\mu}{\lambda}\right), \tag{13.28}$$

  where γ, λ, and μ are parameters left free to float. With respect to the Gaussian distribution, the Johnson's S_U distribution can be asymmetric, if the γ is different from zero, and can accommodate different tail widths, using different values for the parameter δ. A Johnson's S_U distribution converges to a gaussian when $\delta, \lambda \to \infty$ and $\gamma = 0$, with $\sigma = \lambda/\delta$.

- **PDL uncertainty -** The proper decay length uncertainty σ_{ct} is described by the sum of two gamma distributions, each defined as:

$$P_\Gamma = \frac{(\sigma_{ct}-\mu)^{\gamma-1} e^{\frac{-(\sigma_{ct}-\mu)}{\beta}}}{\Gamma(\gamma)\beta^\gamma}, \tag{13.29}$$

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

- **Mistag probability -** The mistag probability ω_{tag} is described using an adaptative KDE[X] in the range $[0, 0.5]$ with gaussian cores and mirrored boundary conditions. As a reminder, ω_{tag} is constrained between 0 and 0.5 by inverting the tagging decision for $\omega_{\text{tag}} > 0.5$.

The result of the UML fit to the background subtracted data samples is shown in Fig. 13.2 (*dataset_JpsiMuon*) and Fig. 13.3 (*dataset_JpsiTrkTrk*).

S-P WAVE INTERFERENCE The invariant mass of the kaon system m_{KK} is not fitted. This, if not accounted for, may cause a bias in the physics results due to the differences in the m_{KK} dependence between the P-wave from $B_s^0 \to J/\psi\,\phi(1020)$ and the S-wave from the non-resonant $B_s^0 \to J/\psi K^+K^-$ decay and the $B_s^0 \to J/\psi f_0(980) \to B_s^0 \to \mu^+\mu^- K^+K^-$ decay.

The amplitude dependence on m_{KK} can be expressed as:

$$A_i(m_{KK}) = A_i \pi(m_{KK}), \tag{13.30}$$

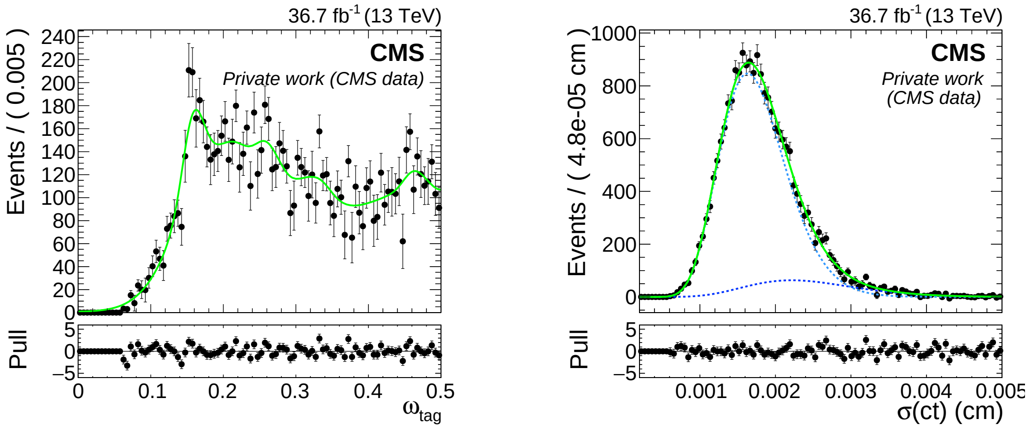
$$A_S(m_{KK}) = A_S \sigma(m_{KK}), \tag{13.31}$$

where $i = (0, \parallel, \perp)$, and $\pi(m_{KK})$ and $\sigma(m_{KK})$ are the mass dependence parametrization of the P- and S-wave amplitudes, respectively, assumed to be normalized to unity in the m_{KK} range of the measurement. To properly model

---

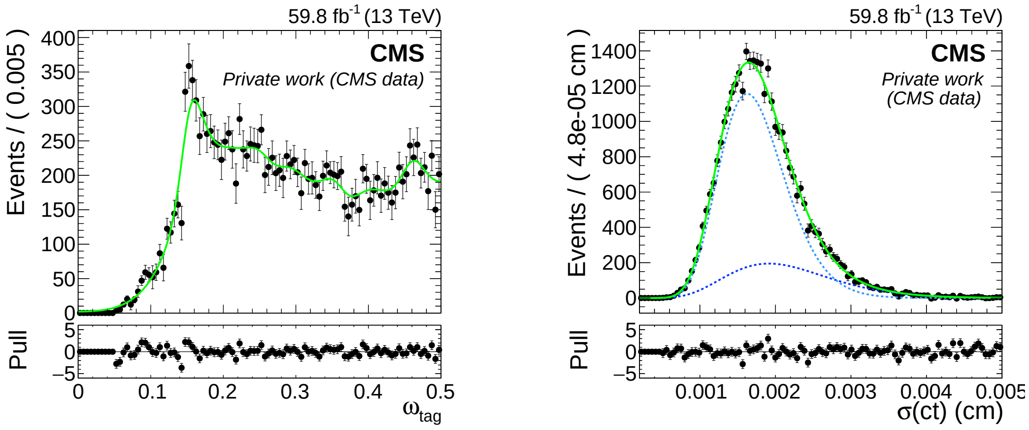
[IX] The gamma *distribution* and gamma *function* are different objects.

[X] More accurately described in Section 13.2.7.1.



(a) ω_{tag} (2017).

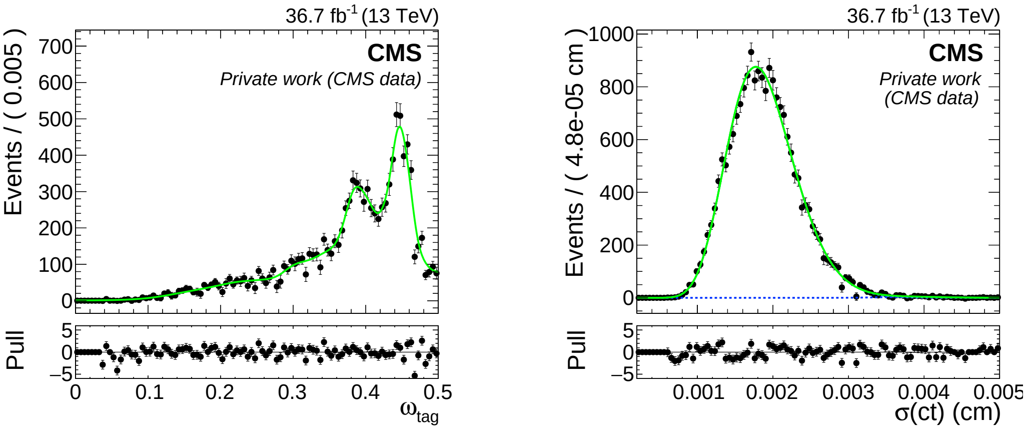
(b) PDL uncertainty (2017).



(c) ω_{tag} (2018).

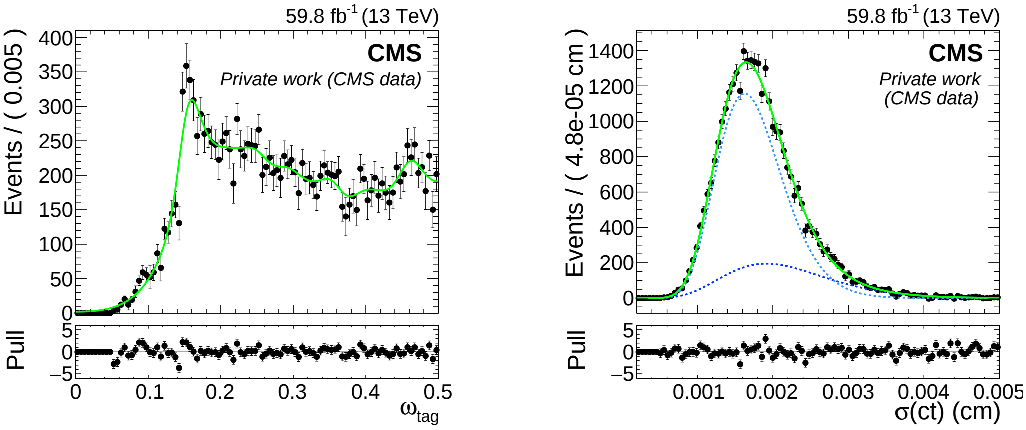
(d) PDL uncertainty (2018).

Figure 13.2: Distribution of the input observables in *dataset_JpsiMuon* after the subtraction of both backgrounds. The solid green line represents the 1D projection of the P_{sig} fit, which will be used as pre-fit in the final fit. The dashed blue lines in the PDL uncertainty fit are the two gamma components of the signal fit.



(a) ω_{tag} (2017).

(b) PDL uncertainty (2017).



(c) ω_{tag} (2018).

(d) PDL uncertainty (2018).

Figure 13.3: Distribution of the input observables in *dataset_JpsiTrkTrk* after the subtraction of both backgrounds. The solid green line represents the 1D projection of the P_{sig} fit, which will be used as pre-fit in the final fit. The dashed blue lines in the PDL uncertainty fit are the two gamma components of the signal fit.

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, as in this work, the decay rate terms related to the interference between the P and S-wave, i. e., the terms in Tab. 12.1 with $i = 8, 9, 10$, need to be modified by an effective coupling factor hereafter called k_{SP}. The value of k_{SP} can be evaluated by integrating the amplitudes over the selected m_{KK} range $[m_1, m2] = [m^{\text{w.a.}}_{\phi(1020)} - 10,\ m^{\text{w.a.}}_{\phi(1020)} + 10]$ MeV (corresponding to the analysis selection for the $\phi(1020)$ mass), as:

$$\int_{m_1}^{m_2} A_i^* A_S\, \pi(m_{KK})^* \sigma(m_{KK})\, dm_{KK} = k_{SP}|A_i A_S|\,. \tag{13.32}$$

Since $\pi(m_{KK})$ and $\sigma(m_{KK})$ are normalized in $[m_1, m2]$, the $|A_i|^2$, $\left|A_i A_j\right|$ and $|A_S|^2$ terms do not require any correction:

$$\int_{m_1}^{m_2} A_i^* A_j\, \pi(m_{KK})^* \pi(m_{KK})\, dm_{KK} = \left|A_i A_j\right|, \tag{13.33}$$

$$\int_{m_1}^{m_2} A_S^* A_S\, \sigma(m_{KK})^* \sigma(m_{KK})\, dm_{KK} = |A_S|^2\,. \tag{13.34}$$

The effective coupling factor is computed to be $k_{SP} = 0.538 \pm 0.004$, assuming that the P-wave is described by a relativistic Breit–Wigner distribution and the S-wave amplitude by a constant (for the $f_0(980)$ too, since we are in the tails far outside the resonant peak). The uncertainty is estimated by trying different assumptions for the S-wave lineshape and taking the largest difference from the nominal k_{SP} value as the uncertainty value. Polynomials up to the second order and the Flatté distribution have been tested.

13.2.2 $B^0 \to J/\psi K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ *background* (B^0 *background*)

This background includes $B^0 \to J/\psi K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ events where the pion is mistaken as a kaon candidate. The parameterization of this background has been developed mainly in $B^0 \to J/\psi K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ simulated events, reconstructed as $B_s^0 \to J/\psi\, \phi(1020) \to \mu^+\mu^- K^+K^-$ with the same framework for the signal and reweighed for the PDF efficiency as the data. This background is *not* free to float in the final fit. The methods used to parameterise the PDF are described below and summarized in Tab. 13.3.

The PDFs used to describe the B^0 background are:

- **Invariant mass -** The invariant mass is described by single Johnson's S_U distribution fitted on $B^0 \to J/\psi K^*(892)^0$ simulated events and then fixed in the final fit. This distribution is able to model well even the distorted mass distribution of the misreconstructed B^0 peak.

- **Angular variables -** The distribution for the angular variables $\Theta = (\cos\theta_T, \cos\psi_T, \varphi_T)$ is described by a linear combination of Bernstein poly-

Table 13.3: Summary of the methods used to derive the parametrization of the B^0 background. Simulation stands for $B^0 \to J/\psi\, K^*(892)^0$ simulated sample corrected for the PDL efficiency.

Component	Parametrization method
$P_{\mathrm{bkg}B^0}(m_{B_s^0})$	Simulation weighted for $1/\varepsilon(ct)$
$P_{\mathrm{bkg}B^0}(\Theta)$	Simulation weighted for $1/\varepsilon(ct)$
$P_{\mathrm{bkg}B^0}(ct)$	Simulation weighted for $1/\varepsilon(ct)$
$P_{\mathrm{bkg}B^0}(\sigma_{ct})$	Same as $P_{\mathrm{sig}}(\sigma_{ct})$
$P_{\mathrm{bkg}B^0}(\omega_{\mathrm{tag}})$	Simulation weighted for $1/\varepsilon(ct)$
$f_{\mathrm{bkg}B^0}$	Data

nomials fitted on $B^0 \to J/\psi\, K^*(892)^0$ simulated events and then fixed in the final fit. The model is defined as:

$$P(\cos\theta_T, \cos\psi_T, \varphi_T) = \sum_{i=0}^{n} \sum_{j=0}^{m} \sum_{k=0}^{r} c_{ijk}\, B_i^n(\cos\theta_T)\, B_j^m(\cos\psi_T)\, B_k^r(\sin\varphi_T)\,, \tag{13.35}$$

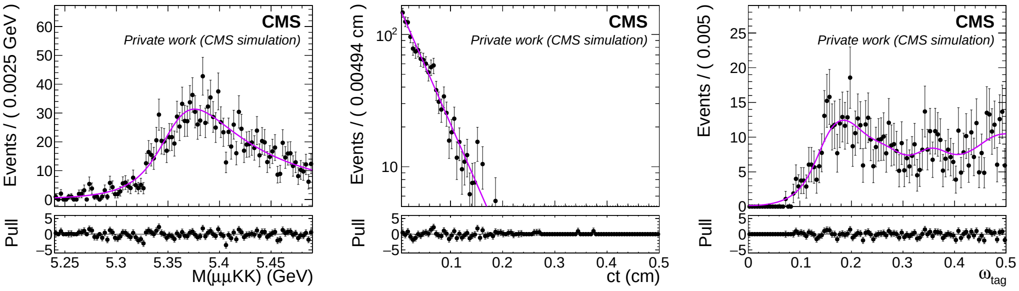
where

$$B_i^n(z) = \binom{n}{i} z^i\,(1-z)^{n-i} \tag{13.36}$$

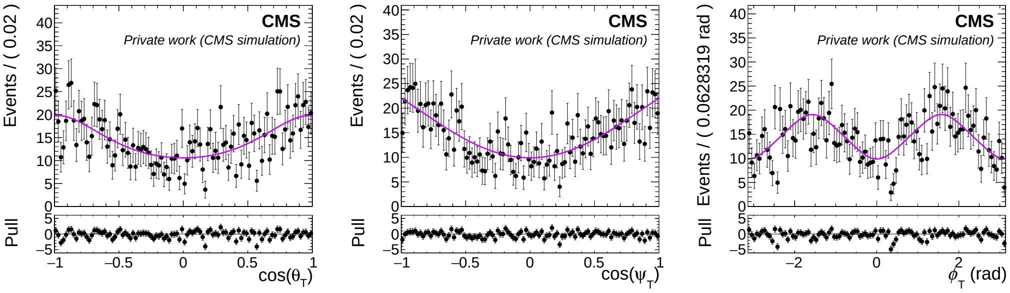
is the Bernstein basis of order n. The Bernstein polynomials were chosen because they are positive definite and form a basis for the polynomials up to degree n. In the fit, the following values are used: $n = 4$, $m = 4$, and $r = 5$.

- **Proper decay length -** The ct distribution is described by a single exponential distribution fitted on $B^0 \to J/\psi\, K^*(892)^0$ simulated events and then fixed in the final fit. However, σ_{ct} was observed to be underestimated in simulated samples, and a different method needs to be used for its distribution. Since the $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^-\, K^+\pi^-$ decay is topologically very similar to $B_s^0 \to J/\psi\, \phi(1020) \to \mu^+\mu^-\, K^+K^-$ (displaced secondary vertex fitted from two muons and two tracks), the B^0 background PDL uncertainty distribution is approximated with the one fitted for the signal, i. e. $P_{\mathrm{bkg}B^0}(\sigma_{ct}) \equiv P_{\mathrm{sig}}(\sigma_{ct})$.
- **Mistag probability -** The mistag probability ω_{tag} is described using KDE in the range [0, 0.5] with gaussian cores and mirrored boundary conditions. The KDE is applied to $B^0 \to J/\psi\, K^*(892)^0$ simulated events and then fixed in the final fit.

The result of the UML fit to the $B^0 \to J/\psi\, K^*(892)^0$ simulated samples is shown in Fig. 13.4, Fig. 13.5, Fig. 13.6, and Fig. 13.7.

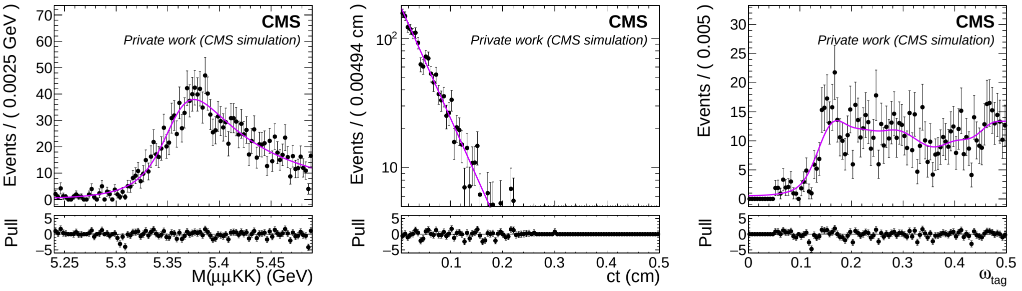


(a) Invariant mass. (b) Proper decay length. (c) Mistag probability.

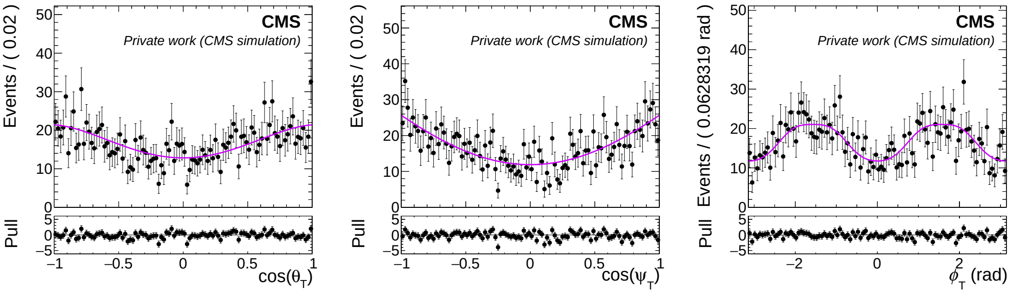


(d) $\cos\theta_T$. (e) $\cos\psi_T$. (f) φ_T.

Figure 13.4: Physics observables distributions for $B_s^0 \rightarrow J/\psi\,\phi(1020)$ candidates reconstructed in $B^0 \rightarrow J/\psi\,K^*(892)^0$ simulated events for *dataset_JpsiMuon* (2017). The PDL uncertainty is not shown as it is not used. The solid line represents the 1D projection of the $P_{\mathrm{bkg}B^0}$ fit.

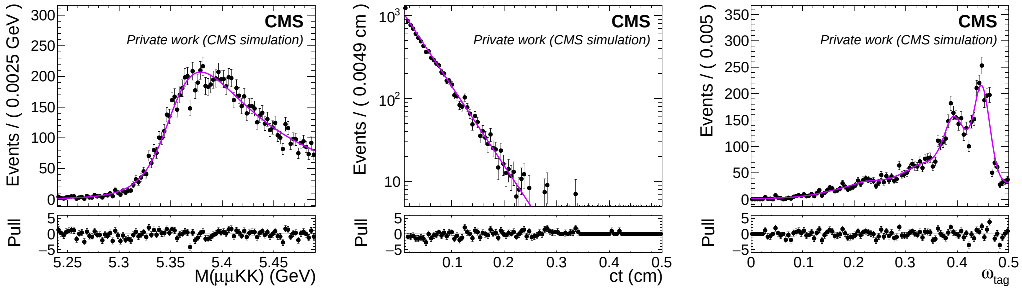


(a) Invariant mass. (b) Proper decay length. (c) Mistag probability.

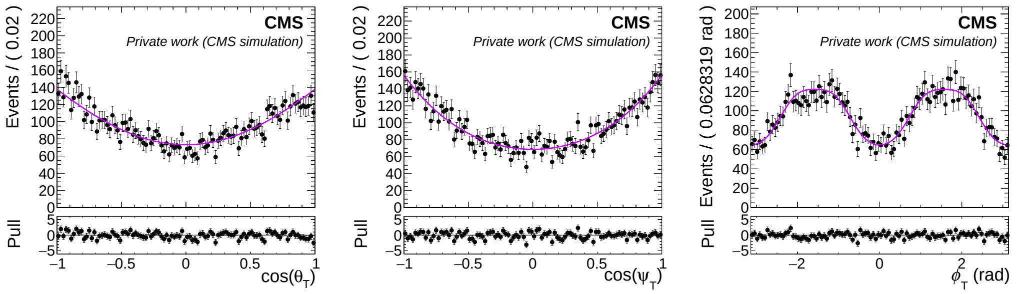


(d) $\cos\theta_T$. (e) $\cos\psi_T$. (f) φ_T.

Figure 13.5: Physics observables distributions for $B_s^0 \to J/\psi\,\phi(1020)$ candidates reconstructed in $B^0 \to J/\psi\,K^{*}(892)^0$ simulated events for *dataset_JpsiMuon* (2018). The PDL uncertainty is not shown as it is not used. The solid line represents the 1D projection of the $P_{\mathrm{bkg}B^0}$ fit.

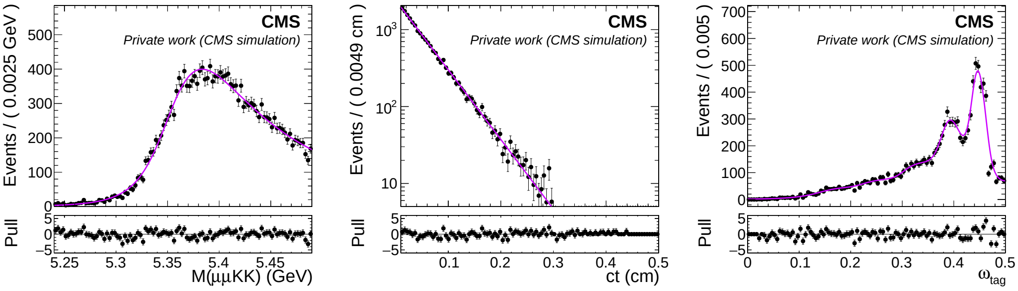


(a) Invariant mass. (b) Proper decay length. (c) Mistag probability.

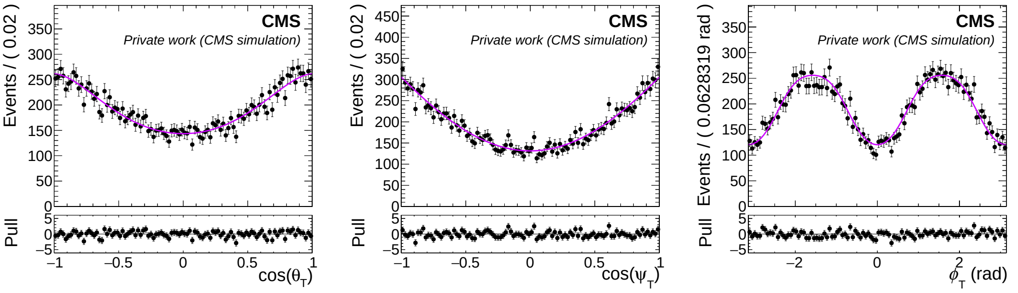


(d) $\cos\theta_T$. (e) $\cos\psi_T$. (f) φ_T.

Figure 13.6: Physics observables distributions for $B_s^0 \to J/\psi\,\phi(1020)$ candidates reconstructed in $B^0 \to J/\psi\,K^*(892)^0$ simulated events for *dataset_JpsiTrkTrk* (2017). The PDL uncertainty is not shown as it is not used. The solid line represents the 1D projection of the $P_{\mathrm{bkg}B^0}$ fit.



(a) Invariant mass. (b) Proper decay length. (c) Mistag probability.



(d) $\cos\theta_T$. (e) $\cos\psi_T$. (f) φ_T.

Figure 13.7: Physics observables distributions for $B_s^0 \rightarrow J/\psi\,\phi(1020)$ candidates reconstructed in $B^0 \rightarrow J/\psi\,K^{*}(892)^0$ simulated events for *dataset_JpsiTrkTrk* (2018). The PDL uncertainty is not shown as it is not used. The solid line represents the 1D projection of the $P_{\mathrm{bkg}B^0}$ fit.

EFFECTIVE EVENT FRACTION The effective event fraction for the B^0 background is measured in data with a simultaneous fit to the B_s^0 and B^0 mesons invariant mass distributions in both the analysis data sample of $B_s^0 \to J/\psi\,\phi(1020)$ and the corresponding "B^0 reflection" sample, using the procedure that was developed for the fit of B_s^{**} events for the same side tagger calibration (described in paragraph 10.2.5). The reflection sample is created by reconstructing each $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ candidate as a $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ one. This is achieved by changing the mass hypothesis on one of the two tracks to build a $K^*(892)^0 \to K^+\pi^-$ candidate[XI]. Both charged tracks are tested as the pion candidate and the choice that produces the $K^*(892)^0$ candidate with mass closer to the world-average value is chosen. In both samples, both the signal (B_s^0 for $B_s^0 \to J/\psi\,\phi(1020)$ and B^0 for $B^0 \to J/\psi\,K^*(892)^0$) and the reflection of the *other* meson species (B^0 for $B_s^0 \to J/\psi\,\phi(1020)$ and B_s^0 for $B^0 \to J/\psi\,K^*(892)^0$) are fitted to extract the B^0 effective fraction in the analysis data sample. The shape parameters of the B^0 distributions and the shape parameters of the B_s^0 distribution in the B^0 reflection sample are fixed to the values extracted from fit to simulations. Due to the very limited fit range, the complex reflection shape described in paragraph 10.2.5 is not necessary, and a simple Johnson's S_U distribution is used instead. The effective fractions and shape parameters of the combinatorial background and the B_s^0 distribution in the B_s^0 sample are left free to float. The results are reported in Tab. 13.4 shown in Fig. 13.8. The B^0 background makes up around $4-5\%$ of the total event weight and around 20% of the total background.

Table 13.4: Measured effective event (sum of weights) yield of $B^0 \to J/\psi\,K^*(892)^0 \to \mu^+\mu^- K^+\pi^-$ decays in the various data sets.

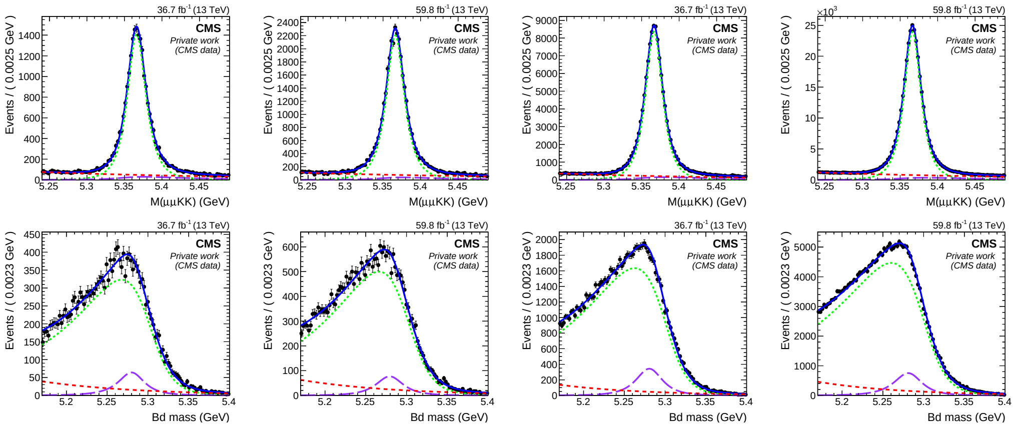
Data set	B^0 bkg. eff. fraction
dataset_JpsiMuon (2017)	0.051 ± 0.006
dataset_JpsiMuon (2018)	0.041 ± 0.005
dataset_JpsiTrkTrk (2017)	0.047 ± 0.002
dataset_JpsiTrkTrk (2018)	0.039 ± 0.001

13.2.3 *Combinatorial background*

The main component of the background, assumed to be mostly combinatorial, has been developed in the analysis data sample after the subtraction of the signal and the B^0 background. The chosen distributions are the ones found to best model each variable in the simulated events.

- **Invariant mass -** The mass distribution is described by an exponential distribution.

[XI] Charge conjugation is implied.



(a) *dataset_JpsiMuon* (2017). (b) *dataset_JpsiMuon* (2018). (c) *dataset_JpsiTrkTrk* (2017). (d) *dataset_JpsiTrkTrk* (2018).

Figure 13.8: Fit to the B_s^0 (upper row) and B^0 (lower row) invariant mass distributions used to extract the yield of the $B^0 \to J/\psi K^*(892)^0$ background. The solid line represents the projection of the fit. The dashed green, red, and magenta lines represent the B_s^0 signal, combinatorial background, and B^0 background, respectively.

- **Angular variables -** As for the B^0 background, the distribution for the angular variables $\Theta = (\cos\theta_T, \cos\psi_T, \varphi_T)$ is described by a linear combination of Bernstein polynomials.
- **Proper decay length -** The ct distribution is described by the sum of two decreasing exponential distributions with different time constants. Its uncertainty (σ_{ct}) is described by two Gamma distributions (defined in Eq. 13.29).
- **Mistag probability -** As for the signal and the B^0 background, the mistag probability is described using KDE in the range $[0,\ 0.5]$ with gaussian cores and mirrored boundary conditions.

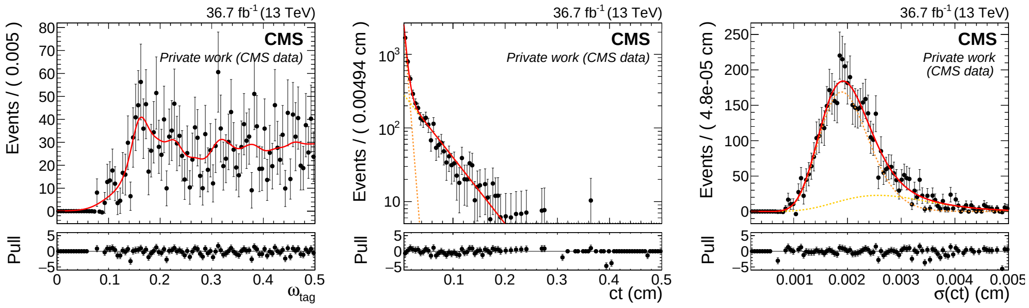
The result of the UML fit to the signal and B^0 background subtracted data samples is shown in Fig. 13.9 (*dataset_JpsiMuon*) and Fig. 13.10 (*dataset_JpsiTrkTrk*).

13.2.4 *Flavor tagging dilution*

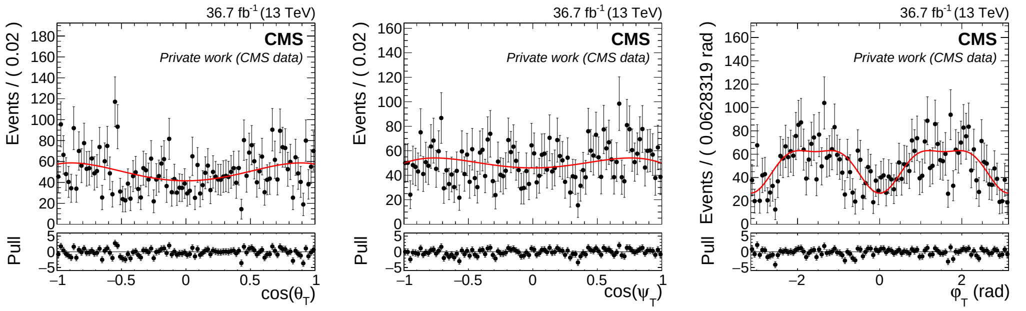
Mistagged events directly dilute the observed CP asymmetry by a factor

$$\mathcal{D}_{\text{tag}} = (1 - 2\omega_{\text{tag}})\,.$$

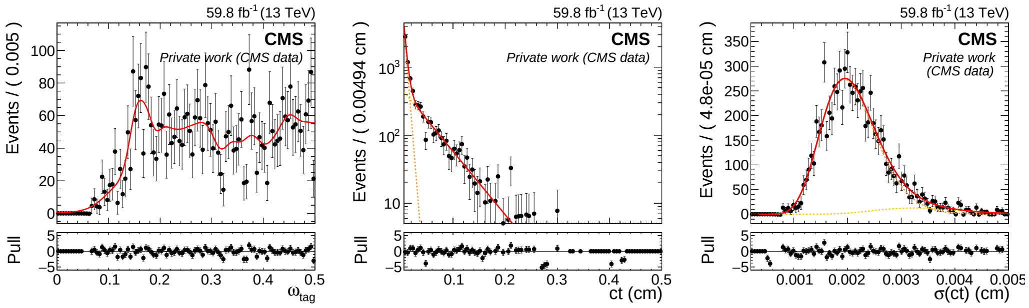
The origin and estimation of the tagging dilution are extensively described in Part III. This effect is incorporated into the fit model by modifying the time-



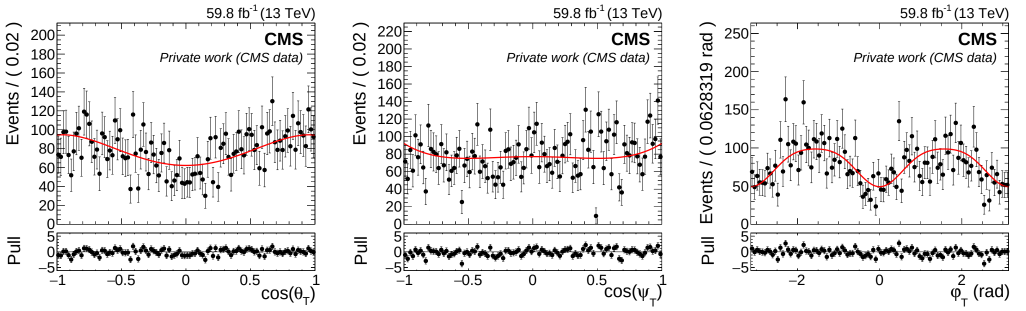
(a) ω_{tag} (2017). (b) Proper decay length (2017). (c) PDL uncertainty (2017).



(d) $\cos\theta_T$ (2017). (e) $\cos\psi_T$ (2017). (f) φ_T (2017).

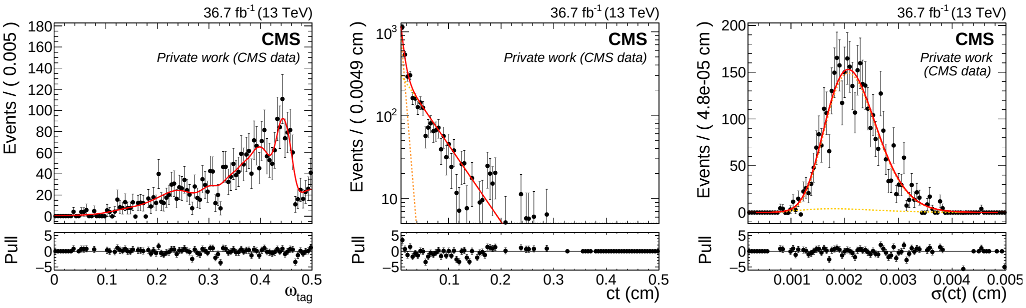


(g) ω_{tag} (2018). (h) Proper decay length (2018). (i) PDL uncertainty (2018).

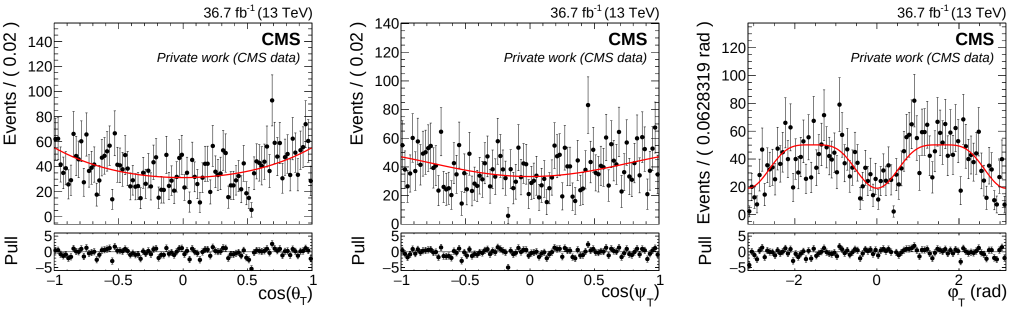


(j) $\cos\theta_T$ (2018). (k) $\cos\psi_T$ (2018). (l) φ_T (2018).

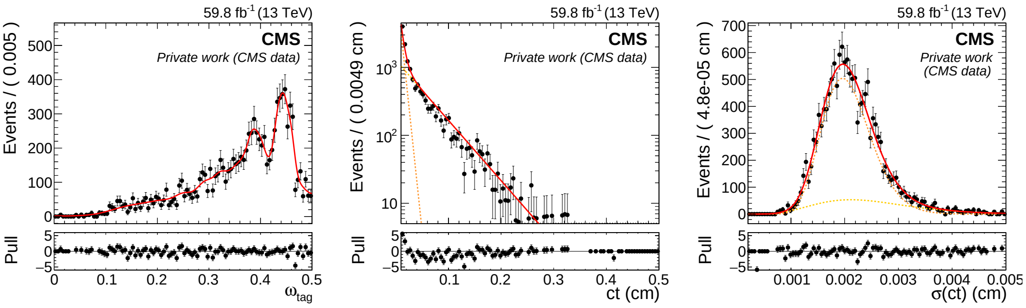
Figure 13.9: Distribution of the input observables in *dataset_JpsiMuon* after the subtraction of the signal and B^0 background. The solid line represents the 1D projection of the P_{bkg} fit, which will be used as pre-fit in the final fit.



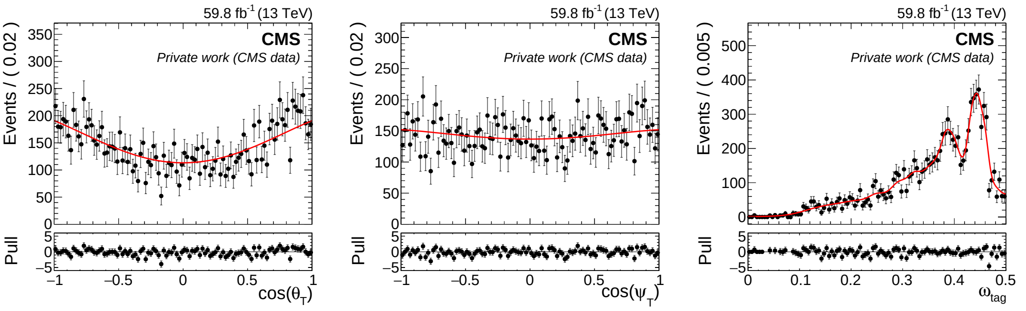
(a) ω_{tag} (2017). (b) Proper decay length (2017). (c) PDL uncertainty (2017).



(d) $\cos\theta_T$ (2017). (e) $\cos\psi_T$ (2017). (f) φ_T (2017).



(g) ω_{tag} (2018). (h) Proper decay length (2018). (i) PDL uncertainty (2018).



(j) $\cos\theta_T$ (2018). (k) $\cos\psi_T$ (2018). (l) φ_T (2018).

Figure 13.10: Distribution of the input observables in *dataset_JpsiTrkTrk* after the subtraction of the signal and B^0 background. The solid line represents the 1D projection of the P_{bkg} fit, which will be used as pre-fit in the final fit.

dependent functions $\mathcal{O}_i$ of the differential decay rate model[XII] to include the flavor tagging information, as:

$$\begin{aligned} O_i(\alpha, t) = N_i e^{-\Gamma_s t} \Bigg[& a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ & + c_i \xi^i_{\text{tag}} \left(1 - 2\omega^i_{\text{tag}}\right) \cos(\Delta m_s t) \\ & + d_i \xi^i_{\text{tag}} \left(1 - 2\omega^i_{\text{tag}}\right) \sin(\Delta m_s t) \Bigg], \end{aligned} \tag{13.37}$$

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[XIII].

13.2.5 *Decay length resolution*

In order to properly resolve the flavor oscillation of the B^0_s–$\overline{B}^0_s$ system, the proper decay time resolution must be significantly smaller than the oscillation period $T = 2\pi/\Delta m_s \sim 350$ fs. Since the proper decay length (PDL) ct and the proper decay time t differ only by a factor c, sometimes in this thesis the two terms are interchanged depending on which unit of measure is more suitable for the discussion.

Being able to properly observe the B^0_s 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^0_s and $\overline{B}^0_s$ decays, which is roughly given by:

$$a_{\text{CP}}(t) \propto \sin\phi_s \sin(\Delta m_s t)\,. \tag{13.38}$$

The impact of the decay time resolution (hereafter indicated with δ_t) can be estimated assuming a Gaussian resolution function, as [168]:

$$a^{\text{meas}}_{\text{CP}}(t) = \int_{-\infty}^{+\infty} a_{\text{CP}}(t') \frac{1}{\sqrt{2\pi}\delta_t} e^{-\frac{(t-t')}{2\delta_t^2}}\, dt' = e^{-\frac{\delta_t^2 \Delta m_s^2}{2}} a_{\text{CP}}(t)\,, \tag{13.39}$$

From Eq. 13.39 follows that the decay time resolution effectively damps the flavor oscillations amplitude, and, consequently, the observed CP asymmetry, by a dilution factor equal to

$$\mathcal{D}_{\text{time}} = e^{-\delta_t^2 \Delta m_s^2/2}\,, \tag{13.40}$$

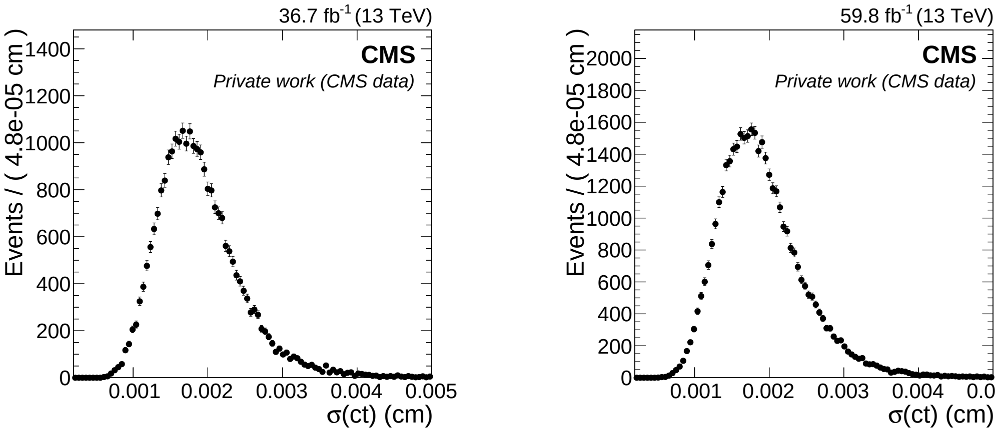
therefore the sensitivity on ϕ_s is directly proportional to $\mathcal{D}_{\text{time}}$[XIV]. The typical δ_t values in this analysis are of the order of $50 \sim 100$ fs, which correspond to dilution

---

[XII] The $\mathcal{O}_i$ functions have been first defined in Eq. 12.21.

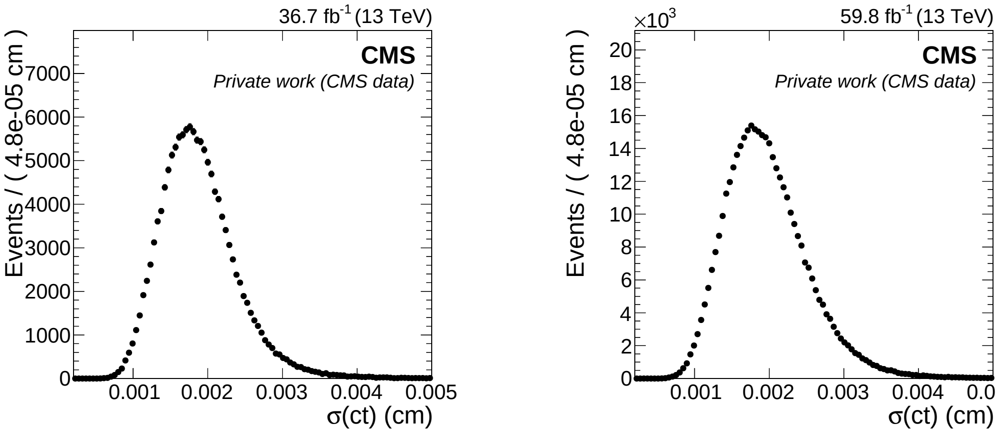
[XIII] The c_i and d_i terms change sign between $B^0_s \to J/\psi\, \phi(1020)$ and $\overline{B}^0_s \to J/\psi\, \phi(1020)$ decays, as can be seen from Eq. 12.15.

[XIV] This is only true as a first approximation, as some terms of the decay rate are not directly related to the B^0_s mixing. However, these terms are less sensitive to ϕ_s and can be neglected in this discussion.



(a) σ_{ct} distribution in *dataset_JpsiMuon* (2017).

(b) σ_{ct} distribution in *dataset_JpsiMuon* (2018).



(c) σ_{ct} distribution in *dataset_JpsiTrkTrk* (2017).

(d) σ_{ct} distribution in *dataset_JpsiTrkTrk* (2018).

Figure 13.11: Proper decay length uncertainty in the various $B_s^0 \to J/\psi\,\phi(1020)$ data sets.

factors of approximately $0.2 \sim 0.7$. 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 Chapter 6.

The measured σ_{ct} distribution in the selected $B_s^0 \to J/\psi\,\phi(1020)$ data sample is shown in Fig. 13.11, with average values of the order of 15–20 μm. It is worth noting that the ct uncertainty σ_{ct} and the ct resolution δ_{ct} 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. 13.39) that *damps* the flavor oscillation and needs to be convolved with the original ct distribution to obtain the measured one. To use σ_{ct} in the fit model, it must be calibrated.

TIME RESOLUTION CALIBRATION The ct uncertainty σ_{ct} is calibrated to better describe the resolution δ_{ct} to account for any underestimate or overestimate of the ct uncertainties. This calibration can be performed by comparing event-by-event the measured σ_{ct} with the difference between the measured and *true* ct. This can be achieved either in simulations, where the true ct is known from the MC truth, or in data, exploiting prompt events that have a lifetime comparable with zero.

In this work, σ_{ct} uses the second method, prompt events in a $B_s^0 \to J/\psi\,\phi(1020)$ data sample. This sample is created by removing the selection requirement $ct > 60\ \mu\mathrm{cm}$ from the *dataset_JpsiMuon* data sets. *dataset_JpsiTrkTrk* is not used due to the displacement requirement in the HLT. Calibration is performed separately for 2017 and 2018, and assumed to be the same in both triggers.

The ct distribution is then fitted with the following model:

$$P(ct) \propto \left[f_{\mathrm{prompt}} P_{\mathrm{prompt}}(ct) + (1 - f_{\mathrm{prompt}}) P_{\mathrm{LL}}(ct)\right] \otimes G(ct)\,, \tag{13.41}$$

where

- f_{prompt} is the fraction of prompt events.
- $P_{\mathrm{prompt}}(ct) = \delta(0)$ is the PDF that models the prompt events. It consists of a Dirac's delta distribution centered in $ct = 0$ cm.
- $P_{\mathrm{LL}}(ct) = f_{\mathrm{decay}} e^{-ct/\tau_1} + (1 - f_{\mathrm{decay}}) e^{-ct/\tau_2}$ is the PDF modeling the *long lived* events. It is composed by two exponential distribution.
- $G(ct) = f_1 G_1(ct, 0, \sigma_2) + (1 - f_1) G_2(ct, 0, \sigma_2)$ is the resolution function (to be convolved). It is composed by two gaussian resolution functions centered in zero.

The fit is performed in the following bins of σ_{ct}:

0. [2, 15] μm;
1. [15, 20] μm;
2. [20, 25] μm;
3. [25, 30] μm;
4. [30, 50] μm.

For stability, the parameters f_{decay1}, τ_1, and τ_2 are fixed to the value obtained by the pre-fit over the whole σ_{ct} range. The fit results are shown in Fig. 13.12 and Fig. 13.13. It is noted that for the first bin, the long-lived component is not well described. This is probably due to the fact that the parameters of $P_{\mathrm{LL}}(ct)$ are fixed to the pre-fit. This is not considered a problem for the measurement of the resolution, since it only affects the tails and not the gaussian resolution.

Table 13.5: PDL uncertainty calibration results. The fitted model is $f_{\delta_{\text{eff}}}(\sigma_{ct}) = a + b \cdot \sigma_{ct}$.

Data set	a [cm]	b
2017	0.00016 ± 0.00006	1.01 ± 0.03
2018	0.00014 ± 0.00007	1.01 ± 0.03

Using the fit results, the effective time dilution can be calculated in each σ_{ct} bin as

$$\mathcal{D} = f_1 \exp\left\{\left(-\frac{\sigma_1 \Delta m_s^2}{2}\right)\right\} + (1 - f_1) \exp\left(-\frac{\sigma_2 \Delta m_s^2}{2}\right), \tag{13.42}$$

which can be used to estimate the effective PDL resolution for each σ_{ct} bin as:

$$\delta_{\text{eff}} = \sqrt{\frac{-2 \ln \mathcal{D}}{\Delta m_s^2}}. \tag{13.43}$$

Using the estimated δ_{eff}, the ct uncertainty is finally calibrated with a linear fit to the pairs $(\sigma_{ct}, \delta_{\text{eff}})_{i=0,\dots,4}$. The calibration results are shown in Fig. 13.14 and reported in Tab. 13.5. The results show a linear relationship between δ_{eff} and σ_{ct}, with a slope compatible with 1 and intercept $\approx 1.5\ \mu$m. This proves that the PDL uncertainty σ_{ct} is a very good approximation of the effective resolution, no scale effects are observed and only a flat correction of about $+1.5\ \mu$m is needed, which corresponds to an increase in uncertainty of approximately 5–10% depending on the σ_{ct} value.

The dependence of the δ_{ct}/σ_{ct} ratio as a function of ct has been checked in MC. No relevant dependence is found.

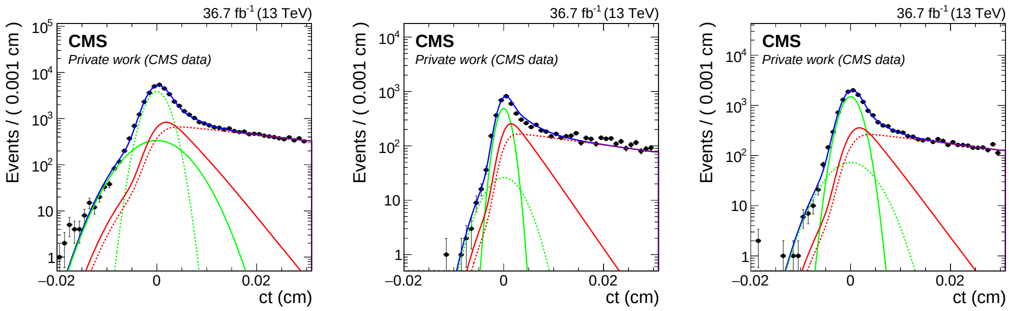
TIME RESOLUTION FIT MODEL The proper decay length resolution is included in the signal fit model from Section 13.2.1 through a Gaussian resolution function that makes use of the scaled per-event σ_{ct} as:

$$G(ct, \sigma_{ct}) = \exp\left[-\frac{1}{2}\left(\frac{ct}{\delta_{ct}}\right)^2\right] \approx \exp\left[-\frac{1}{2}\left(\frac{ct}{f_{\delta_{\text{eff}}}(\sigma_{ct})}\right)^2\right], \tag{13.44}$$

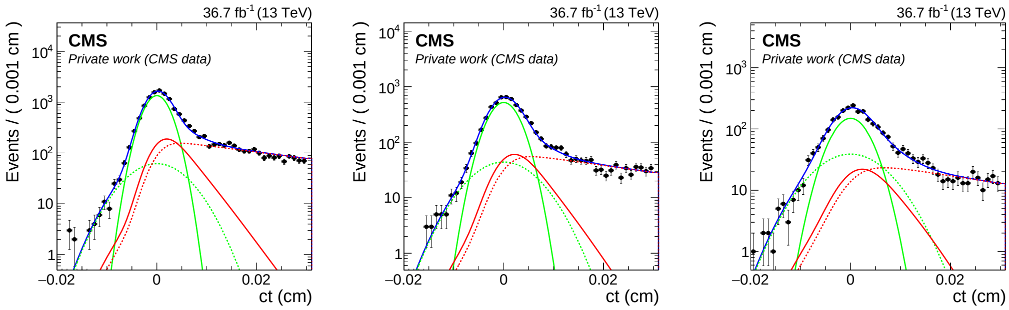
where $f_{\delta_{\text{eff}}}(\sigma_{ct})$ is the resolution calibration function described in the previous section. In the final fit, $G(ct, \sigma_{ct})$ is convolved with the $B_s^0 \to J/\psi\,\phi(1020)$ decay rate, effectively smearing the distribution as a function of time.

13.2.6 *Time efficiency*

The efficiency in reconstructing $B_s^0 \to J/\psi\,\phi(1020)$ candidates depends on the decay length of the B_s^0 meson. In particular, it is expected to decrease at large ct values, due to the decreasing tracking performance (as seen in Fig. 4.3). The proper decay

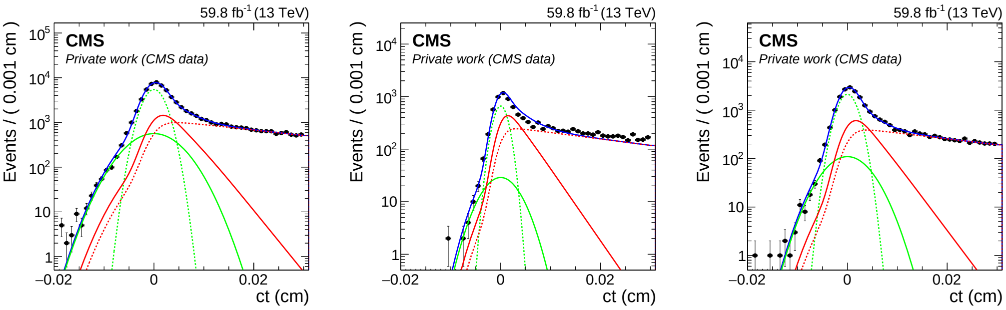


(a) Inclusive fit. (b) $\sigma_{ct} \in [2, 15]$ μm. (c) $\sigma_{ct} \in [15, 20]$ μm.

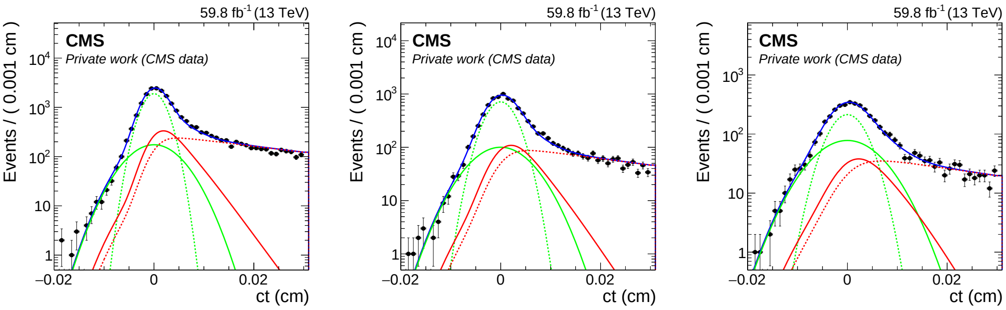


(d) $\sigma_{ct} \in [20, 25]$ μm. (e) $\sigma_{ct} \in [25, 30]$ μm. (f) $\sigma_{ct} \in [30, 50]$ μm.

Figure 13.12: Fit to the ct distribution in 2017 $B_s^0 \to J/\psi\,\phi(1020)$ data to measure the ct resolution. The solid blue line shows the fit to data (solid markers). The components of $P_{\text{LL}}(ct)$ are shown in red, while the components of the resolution function $G(ct)$ are shown in green.

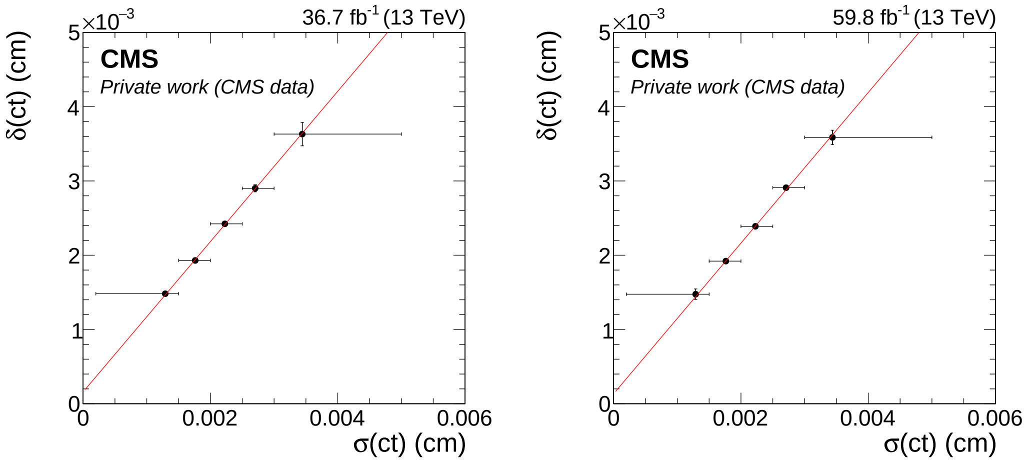


(a) Inclusive fit. (b) $\sigma_{ct} \in [2, 15]$ μm. (c) $\sigma_{ct} \in [15, 20]$ μm.



(d) $\sigma_{ct} \in [20, 25]$ μm. (e) $\sigma_{ct} \in [25, 30]$ μm. (f) $\sigma_{ct} \in [30, 50]$ μm.

Figure 13.13: Fit to the ct distribution in 2018 $B_s^0 \to J/\psi\,\phi(1020)$ data to measure the ct resolution. The solid blue line shows the fit to data (solid markers). The components of $P_{\mathrm{LL}}(ct)$ are shown in red, while the components of the resolution function $G(ct)$ are shown in green.



(a) 2017 data.

(b) 2018 data.

Figure 13.14: PDL uncertainty calibration fits. The measured effective PDL resolution are reported in the y-axis, while the average value of σ_{ct} in the various bins is reported in the x-axis. The fit results are shown with a solid red line. The error bar in the x-axis represents the bins' edges, and are not used in the fit, the uncertainty on the average values are used instead.

length efficiency (hereafter "ct efficiency") is estimated using $B^0 \to J/\psi K^*(892)^0$ decays in data. The $B^0 \to J/\psi K^*(892)^0$ decay has a topology extremely similar to $B^0_s \to J/\psi \phi(1020)$, but has the advantage that the two mass eigenstates of the B^0 have the same average decay time and can then be described by a single exponential distribution. The efficiency can be written as

$$\varepsilon(ct) \propto \frac{N(ct)}{N(ct_{\text{gen}}) \otimes G_{\text{MC}}(ct)}, \tag{13.45}$$

where

- $N(ct)$ is the ct distribution as *measured* in the $B^0 \to J/\psi K^*(892)^0$ sample, after removing the background through sideband subtraction[XV];
- $N(ct_{\text{gen}})$ is the ct distribution as *generated* with an exponential decay, where the decay time $c\tau^0_B$ was set equal to the value from [24];
- $G_{\text{MC}} = \text{PDF}(ct_{\text{TRUE}} - ct)$ is the resolution distribution; the σ_{ct} distribution is fitted to the data sample and calibrated using the result from paragraph 13.2.5.

The numerator and denominator of Eq. 13.45 are estimated, separately for each data set, as histograms with the following bins (the bin edges are given):

[XV] The trigger selection and offline reconstruction procedure is applied. The trigger `HLT_JpsiTrkTrk` is used here despite not being optimized for $B^0 \to J/\psi K^*(892)^0$ decays, because the misreconstruction at trigger level of the B^0 as B^0_s provides enough statistics.

- *dataset_JpsiMuon*: (0.006, 0.008, 0.010, 0.012, 0.014, 0.016, 0.018, 0.020, 0.022, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050, 0.055, 0.060, 0.065, 0.070, 0.075, 0.080, 0.085, 0.090, 0.095, 0.100, 0.125, 0.150, 0.175, 0.200, 0.225, 0.250, 0.275, 0.300, 0.350, 0.400, 0.500) cm;
- *dataset_JpsiTrkTrk*: same binning but starting from 0.01 cm instead[XVI].

The binned efficiency is obtained from the ratio of the two histograms. Given that the efficiency is computed in B^0 decays, the efficiency histogram is corrected according to the ratio of the efficiency histograms obtained in B_s^0 and B^0 simulated samples.

The efficiency histograms in *dataset_JpsiMuon* are fitted with a combination of Bernstein's polynomials as:

$$\varepsilon(ct) = \sum_{i=0}^{n-1} p_i \binom{n}{i} z^i (1-z)^{n-i}, \tag{13.46}$$

where

$$z = \frac{ct - ct_{min}}{ct_{max} - ct_{min}}, \tag{13.47}$$

ct_{min} is either 0.006 cm or 0.01 cm depending on the trigger, ct_{max} is 0.6 and $n = 6$.

In the canonical definition of Bernstein's polynomials the sum is from 0 to n, but here instead a sum from 0 to $n-1$ is used, to constrain the function to decrease monotonically to 0 when $z \to 1$, which improves the stability of the fit. Since the efficiency is not 0 at the end of the range, ct_{max} is not defined as the limit of the ct range, but is one bin further (0.6 cm).

In *dataset_JpsiTrkTrk* an additional term is added, to model the turn-on caused by the displacement requirement of the trigger, so the efficiency is fitted with

$$\varepsilon(ct) = \sigma\left(\frac{ct - ct_{\text{turn-on}}}{w_{\text{turn-on}}}\right) \sum_{i=0}^{n-1} p_i \binom{n}{i} z^i (1-z)^{n-i}, \tag{13.48}$$

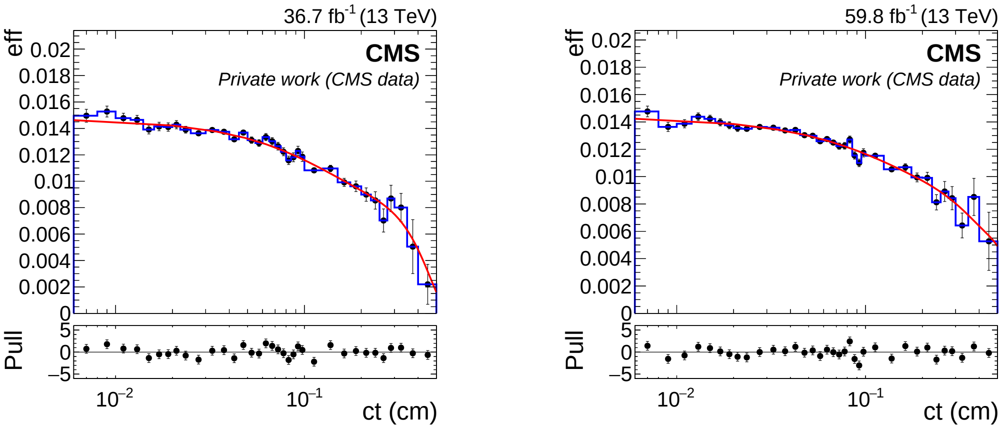
where σ is a sigmoid function and $ct_{\text{turn-on}}$ and $w_{\text{turn-on}}$ are parameters governing, respectively, its position and width.

The efficiency histograms (and the obtained efficiency functions) are shown in Fig. 13.15. As expected, the proper decay length efficiency decreases with increasing values of ct.

13.2.6.1 *PDL efficiency fit implementation*

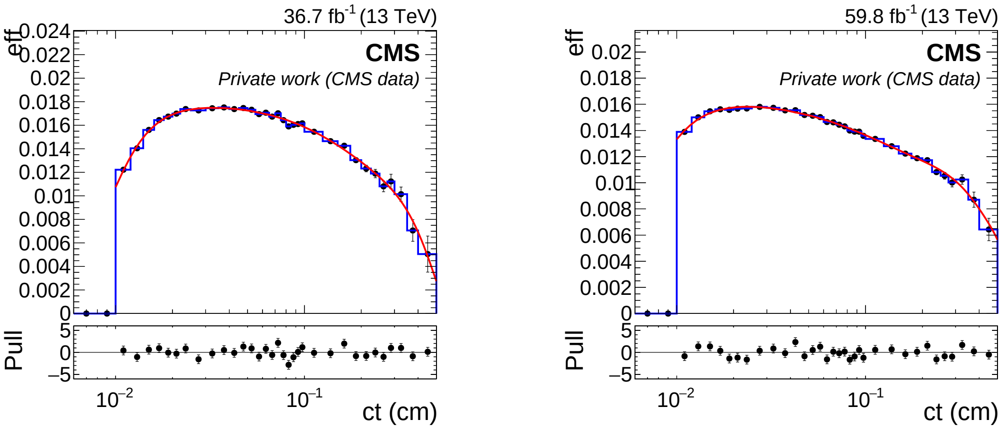
In the final fit model, the fitted PDL efficiency function should be included as a simple multiplicative term in the PDF describing the B_s^0 signal, as is the case for the angular distribution. However, this causes problems in the normalization of the fit model.

[XVI] That is, discarding the first two bins.



(a) PDL efficiency in *dataset_JpsiMuon* (2017).

(b) PDL efficiency in *dataset_JpsiMuon* (2018).



(c) PDL efficiency in *dataset_JpsiTrkTrk* (2017).

(d) PDL efficiency in *dataset_JpsiTrkTrk* (2018).

Figure 13.15: PDL Efficiency in the various $B_s^0 \to J/\psi\,\phi(1020)$ data sets.

For the maximum likelihood fit to work, the PDFs need to behave like real PDF, which includes the normalization to unity in the fit range. The RooFit framework used for the fit handles this transparently, adding normalization integrals inside the PDF definition. For many PDFs, the integral can be computed analytically as a function of the parameters, which is very fast. The addition of an efficiency, however, effectively breaks this, as it is a non-factorizable product, and requires thus a numerical integral. By itself this is not an issue (the angular efficiency, described in Section 13.2.7 has the same effect, yet it is added in the model), but it becomes a bottleneck when used in conjunction with a per-event PDL resolution. Since the decay rate $\tilde{f}(\theta, ct \mid \alpha, \xi_{\mathrm{tag}}, \omega_{\mathrm{tag}})$ is conditional to σ_{ct}, and σ_{ct} is computed in each event, each event gets a new normalization integral. The computation of $\mathcal{O}(10^5)$ numerical integrals for each of the $\mathcal{O}(100)$ fit iterations becomes unbearably slow and is thus avoided.

To significantly speed up the fit, the PDL efficiency is implemented as *weight*, which means that each event is given a weight of

$$w(ct) = \frac{1}{\varepsilon(ct)} \,. \tag{13.49}$$

13.2.7 *Angular efficiency*

The reconstruction and selection procedures distort the shape of the angular distributions of the final-state particles. Therefore, in order to obtain unbiased results, the angular efficiency needs to be taken into account in the final fit model. The angular efficiency is estimated using Kernel Density Estimation using the simulated $B_s^0 \to J/\psi\,\phi(1020)$ $(\Delta\Gamma_s = 0)$ samples.

13.2.7.1 *Kernel Density Estimation*

Kernel Density Estimation (KDE) is a non-parametric method to estimate the PDF of a random variable based on kernels used as weights. The method is described in detail in Ref. [169].

If $X_1, \ldots, X_n$ are real observations from a PDF f, the kernel estimate f_n of f is defined by:

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \tag{13.50}$$

where K is the kernel function, n is the number of data points, and h is the smoothing parameter or window width, which governs how much the data are smoothed to obtain the estimate. It turns out that the quality of the estimate is not strongly correlated with the choice of the kernel function, the latter being usually selected in such a way as to reduce the complexity of the calculation. Many functions are used, with the most widely used being the Gaussian distribution. The choice of the smoothing parameter is to some degree arbitrary, although several methods exist that help in selecting an optimal value for it (e. g., the Silverman rule and the Sheather-Jones algorithm). For example, if the data are sampled from a normal distribution with standard deviation σ, the integrated mean squared error can be minimized by setting:

$$h = 1.06\,\sigma n^{-1/5} \tag{13.51}$$

In this case, σ is evaluated from the data sample.

An additional option for the choice of h is the adaptive method. In an adaptive KDE, the constant h is replaced by a function $h(X_i) \propto 1/\sqrt{\delta(X_i)}$, where $\delta(X_i)$ is the point density around X_i

Although both methods guarantee a very good estimate of the required PDF, their main downside is the computational complexity in the multivariate case [170].

In the RooFit framework, multidimensional KDE PDFs are implemented as the `RooNDKeysPdf` class, but do not scale well with the number of data points. For this reason, an in-house class is implemented for optimal performance in the multivariate case, by using the Fast Fourier Transform (FFT).

13.2.7.2 *KDE with FFT*

A *non-adaptative* KDE can be interpreted as a convolution of the kernel function with the empirical PDF of the data (a sequence of Dirac deltas centered around each point). Both the empirical PDF and the kernel function can be approximated with a discretized grid, simplifying the computation; for the data, this is equivalent to creating an histogram. The convolution interpretation can be used to transform the KDE computation into a product of the Fourier transforms of the two functions, which, in the discrete grid, can be implemented as a Fast Fourier Transform.

As described in Section 12.3.1, the angular distributions of the decay products are described in terms of the three decay angles $\Theta = (\theta_T, \psi_T, \phi_T)$, defined in the transversity basis, as shown in Fig. 12.4.

In order to perform the discrete Fourier Transform, the data is distributed on a 3D grid, where each dimension is related to one of the angular variables $\cos\theta_T$, $\cos\psi_T$, and φ_T. The estimated function is then obtained by means of interpolation. The steps of the algorithm are as follows:

1. Construct a three-dimensional histogram of the data.
2. Determine the bandwidth h by using either the Silverman rule (also known as the rule-of-thumb method [171]) or the Sheather-Jones algorithm.
3. Apply mirrored boundary conditions.
4. Calculate the Fourier transforms of the data and of the Gaussian kernel centered in the middle of the histogram range.

5. Calculate the product of the two Fourier images.
6. Calculate the Inverse Fourier transform of the product.
7. Shift back the resulting histogram to account for the shift introduced by the centered Gaussian kernel.
8. Return the resulting PDF on request.

Unlike the decay length efficiency, here the resolution effects are small enough to be neglected, so the efficiency is just the ratio of the reconstructed and generated angular distributions. The above algorithm is applied for both the reconstructed angular distribution obtained from the simulated $B_s^0 \to J/\psi\,\phi(1020)$ $(\Delta\Gamma_s = 0)$ samples and the generated angular distribution from the simulated $B_s^0 \to J/\psi\,\phi(1020)$ $(\Delta\Gamma_s = 0$, no filters) sample, leading to two KDE approximated distributions:

$$\tilde{\mathrm{N}}_{\mathrm{reco}}(\cos\theta_T, \cos\psi_T, \varphi_T)\,, \tag{13.52}$$

$$\tilde{\mathrm{N}}_{\mathrm{gen}}(\cos\theta_T, \cos\psi_T, \varphi_T) \tag{13.53}$$

The efficiency is obtained from the ratio of the two distributions:

$$\tilde{\varepsilon}(\cos\theta_T, \cos\psi_T, \varphi_T) = \frac{\tilde{\mathrm{N}}_{\mathrm{reco}}(\cos\theta_T, \cos\psi_T, \varphi_T)}{\tilde{\mathrm{N}}_{\mathrm{gen}}(\cos\theta_T, \cos\psi_T, \varphi_T)}\,. \tag{13.54}$$

The 1D projections of the angular efficiency functions are shown in Fig. 13.16, Fig. 13.17, Fig. 13.18, and Fig. 13.19.

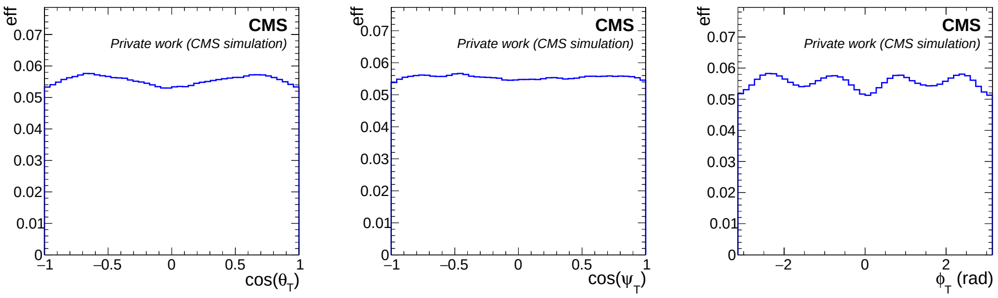
13.3 FIT TO DATA

13.3.1 *Fit procedure*

The physics parameters are extracted with an unbinned extended maximum-likelihood fit performed simultaneously on all the four (two triggers and two years) samples. The PDFs for the 4 samples share only the decay rate, and thus the physics parameters, while every other model and parameter is separate.

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. Some PDFs, indicated as "*pre-fitted*" in the following section, use the pre-fit only as starting value, and are otherwise free to float in the final fit. Others istead, referred to as "*fixed*", have their parameters fixed to the pre-fit value when used in the final fit.

The first step in the procedure is the fit of the B^0 background templates. This fit is performed in the $B^0 \to J/\psi\,K^*(892)^0$ simulated samples, independently for each

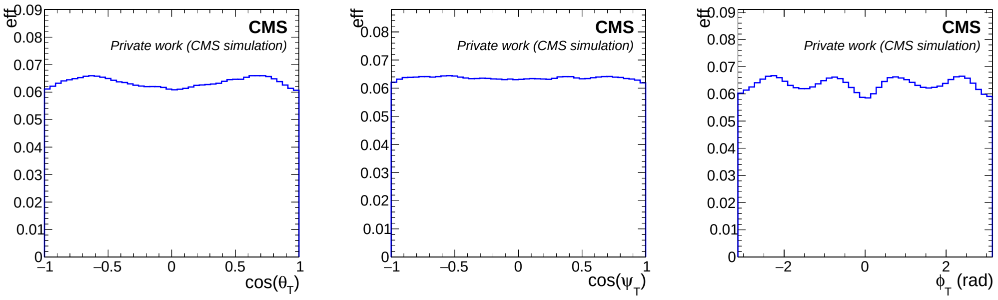


(a) Projection on cos θ_T.

(b) Projection on cos ψ_T.

(c) Projection on cos φ_T.

Figure 13.16: Angular efficiency one-dimensional projections for *dataset_JpsiMuon* (2017).

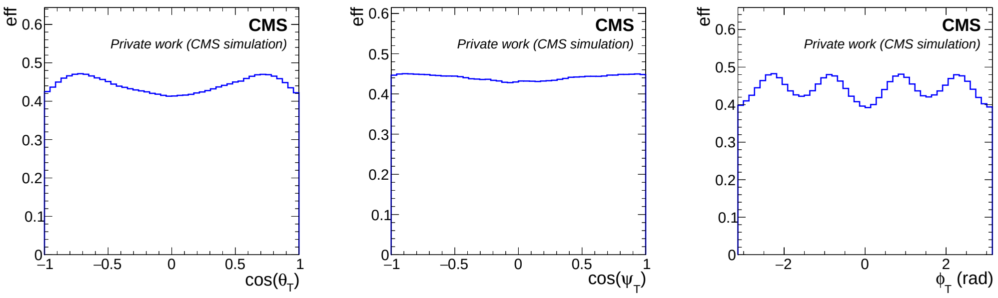


(a) Projection on cos θ_T.

(b) Projection on cos ψ_T.

(c) Projection on cos φ_T.

Figure 13.17: Angular efficiency one-dimensional projections for *dataset_JpsiMuon* (2018).



(a) Projection on cos θ_T.

(b) Projection on cos ψ_T.

(c) Projection on cos φ_T.

Figure 13.18: Angular efficiency one-dimensional projections for *dataset_JpsiTrkTrk* (2017).

year and trigger. In the fit, the B^0 background distribution components are considered independent from each other, and thus fitted separately in each observable. The only distribution of B^0 background that is not estimated in this way is the σ_{ct} distribution, which was observed to be significantly lower in simulated samples. Due to the similar topology of the decay, it is instead considered equal to the signal σ_{ct} distribution. All other distributions are fixed to their estimation in the MC sample. The yields of the B^0 background are also fixed, using the procedure described in paragraph 13.2.2.

Once the B^0 background yields and mass distribution are fixed, a fit is performed in the invariant mass observable, to obtain the yields of the signal and combinatorial background components. This is both used as a pre-fit and needed for the sideband subtraction in the subsequent steps. The yields for each value of ξ_{tag} are separate in the final fit, but estimated with a single pre-fit, with a correction based on the B_s^0, $\overline{B}_s^0$ and untagged fractions of the total sample. This is not a problem, as they are left free to float in the final fit. On the other hand, this means that all distributions pre-fitted or fixed in the next paragraphs (save, of course, for the ω_{tag} distributions) are shared between the different ξ_{tag} values.

All combinatorial components are pre-fitted to the data after executing signal and B^0 subtraction from the mass sidebands. Also for combinatorial background, the distributions are separate for each of the four subsamples. The parameters of the σ_{ct} distributions are fixed, while the ω_{tag} distribution is non-parametric and thus trivially considered fixed. All other parameters are left free to float.

Finally, pre-fits are executed on the signal distributions for σ_{ct} and ω_{tag} using background subtracted data, separately for each subsample. The signal distribution for σ_{ct} is considered shared with B^0 background and thus fitted with the subtraction of only the combinatorial background. Also for the signal, the σ_{ct} and ω_{tag} distributions are fixed to their pre-fitted shapes.

The pre-fitted and fixed PDFs are summarized in Tab. 13.6.

After all the cycles of pre-fits, the UML fit to the combined samples is finally performed.

13.3.2 *Fit results*

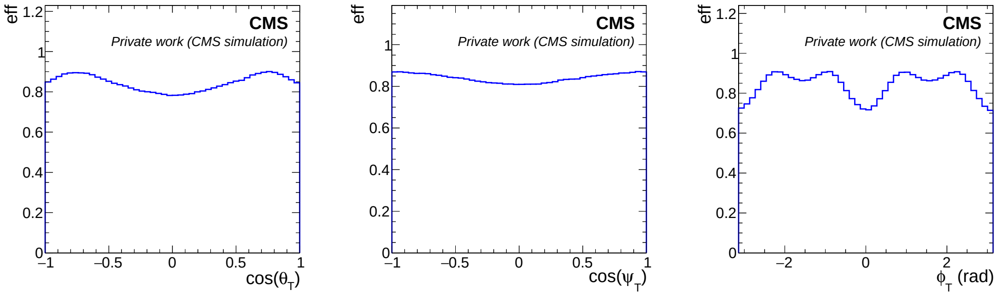
The results of the fit with their statistical uncertainties are given in Tab. 13.7. Statistical uncertainties are obtained with the bootstrap procedure described in paragraph 13.1.2. The statistical correlation matrix is similarly obtained from the bootstrap procedure and is reported in Tab. 13.8. The distributions of the input observables and the corresponding fit projections are shown in Fig. 13.20, 13.21, 13.22 and 13.23. The fit results are discussed more in detail in Chapter 15, after the evaluation of the systematic uncertainties.

Table 13.6: List of pre-fitted and fixed 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. The "bkg" and "bkgB^0" subscripts indicate quantities related to the combinatorial and B^0 components of the background, respectively. The combinatorial background PDFs are pre-fitted in the data mass sidebands. The B^0 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. $P_{\mathrm{bkg}B^0}(\sigma_{ct})$ is set to be equal to $P_{\mathrm{sig}}(\sigma_{ct})$).

Object	Pre-fitted	Fixed
$\tilde{f}(\theta, ct \mid \alpha, \xi_{\mathrm{tag}}, \omega_{\mathrm{tag}})$		
$\varepsilon(ct)\varepsilon(\Theta)$	$-$	✓
$f_{\delta_{\mathrm{eff}}}(\sigma_{ct})$	✓	✓
N_{sig}	✓	
$P_{\mathrm{sig}}(m_{B_s^0})$	✓	
$P_{\mathrm{sig}}(\sigma_{ct})$	✓	✓
$P_{\mathrm{sig}}(\omega_{\mathrm{tag}})$	✓	✓
N_{bkg}	✓	
$P_{\mathrm{bkg}}(m_{B_s^0})$	✓	
$P_{\mathrm{bkg}}(ct)$	✓	
$P_{\mathrm{bkg}}(\sigma_{ct})$	✓	✓
$P_{\mathrm{bkg}}(\Theta)$	✓	
$P_{\mathrm{bkg}}(\omega_{\mathrm{tag}})$	✓	✓
$N_{\mathrm{bkg}B^0}$	✓	✓
$P_{\mathrm{bkg}B^0}(m_{B_s^0})$	✓	✓
$P_{\mathrm{bkg}B^0}(ct)$	✓	✓
$P_{\mathrm{bkg}B^0}(\sigma_{ct})$	$-$	$-$
$P_{\mathrm{bkg}B^0}(\Theta)$	✓	✓
$P_{\mathrm{bkg}B^0}(\omega_{\mathrm{tag}})$	✓	✓

Table 13.7: Results of the fit to data for the physics parameters of most interest. The quoted statistical uncertainties are obtained with the procedure described in paragraph 13.1.2. The results are converted to the units of measure used by the PDG.

Parameter	Fit value
ϕ_s [mrad]	-21 ± 42
$\Delta\Gamma_s$ [ps^{-1}]	0.078 ± 0.010
Γ_s [ps^{-1}]	0.6637 ± 0.0029
Δm_s [$\hbar$ps^{-1}]	$17.737^{+0.067}_{-0.061}$
$\lvert\lambda\rvert$	0.978 ± 0.027
$\lvert A_0\rvert^2$	$0.5235^{+0.0029}_{-0.0032}$
$\lvert A_\perp\rvert^2$	0.2476 ± 0.0045
$\lvert A_S\rvert^2$	$0.0060^{+0.0062}_{-0.0049}$
$\delta_\parallel$ [rad]	3.146 ± 0.023
$\delta_\perp$ [rad]	2.95 ± 0.14
$\delta_{S\perp}$ [rad]	0.635 ± 0.37

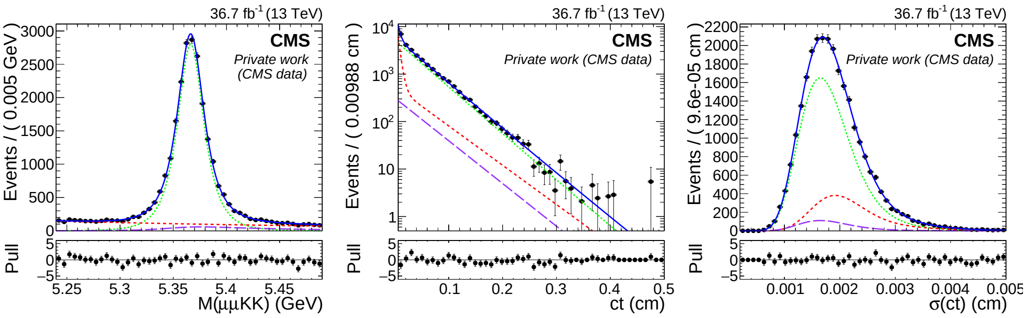


(a) Projection on $\cos\theta_T$.

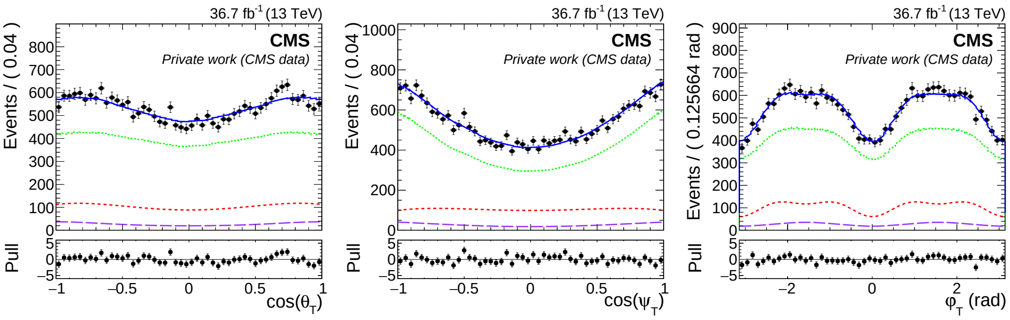
(b) Projection on $\cos\psi_T$.

(c) Projection on $\cos\varphi_T$.

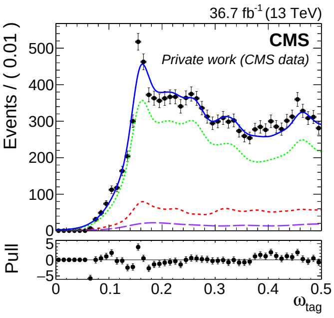
Figure 13.19: Angular efficiency one-dimensional projections for *dataset_JpsiTrkTrk* (2018).



(a) Invariant mass distribution. (b) Proper decay length distribution. (c) PDL uncertainty distribution.

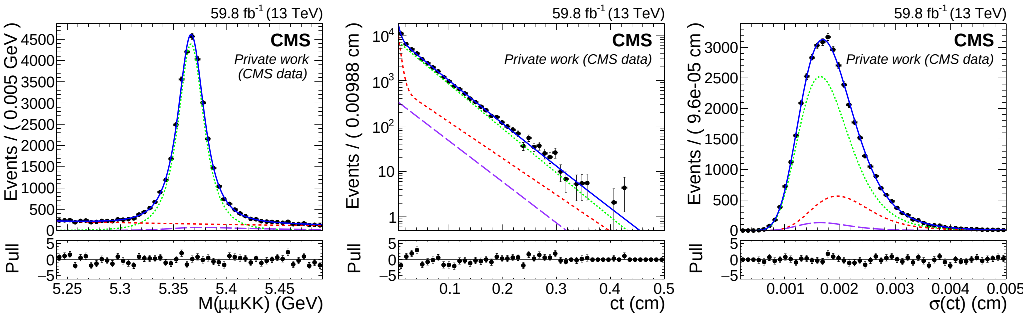


(d) cos θ_T distribution. (e) cos ψ_T distribution. (f) φ_T distribution.

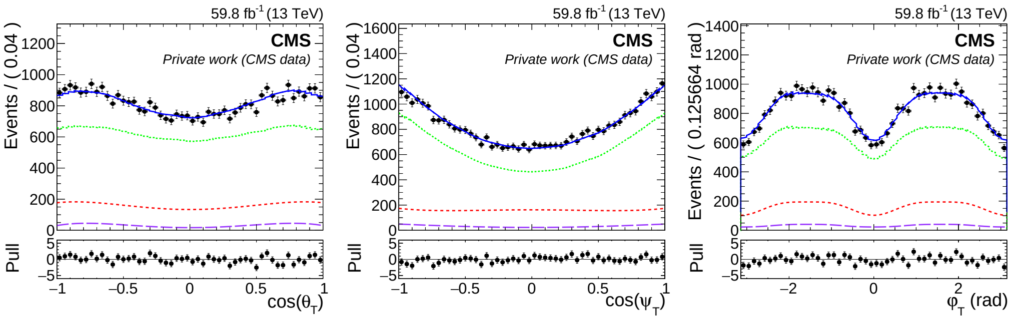


(g) ω_{tag} distribution.

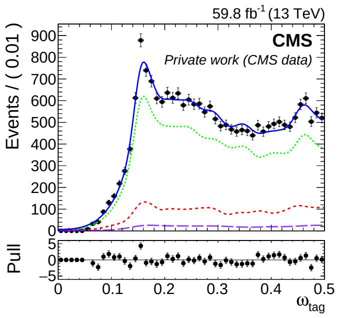
Figure 13.20: Distribution of the input observables and the corresponding fit projections in *dataset_JpsiMuon* (2017).



(a) Invariant mass distribution. (b) Proper decay length distribution. (c) PDL uncertainty distribution.

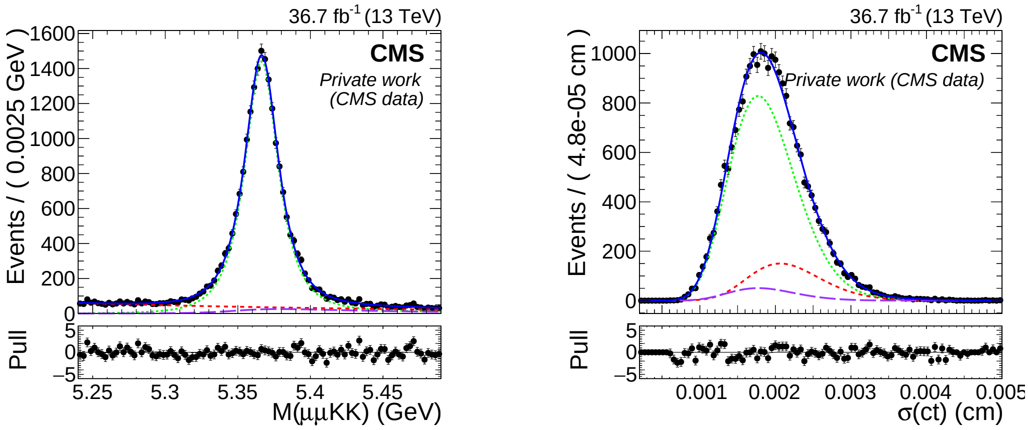


(d) $\cos\theta_T$ distribution. (e) $\cos\psi_T$ distribution. (f) φ_T distribution.



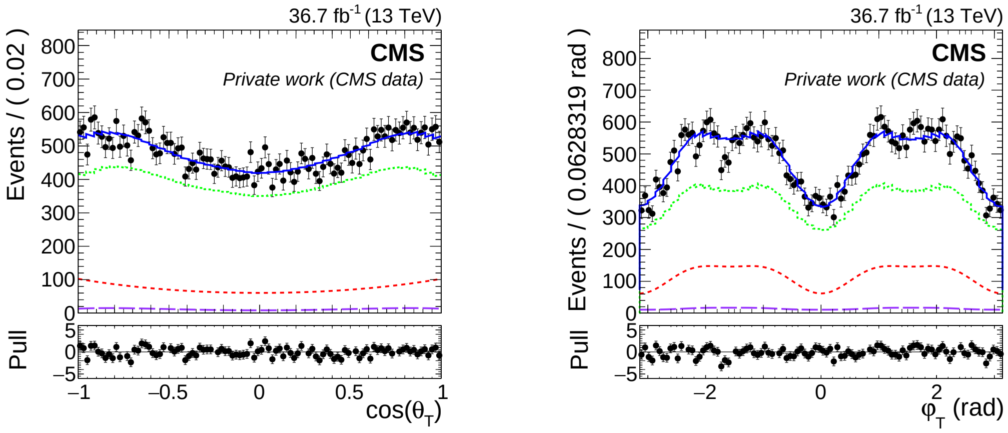
(g) ω_{tag} distribution.

Figure 13.21: Distribution of the input observables and the corresponding fit projections in *dataset_JpsiMuon* (2018).



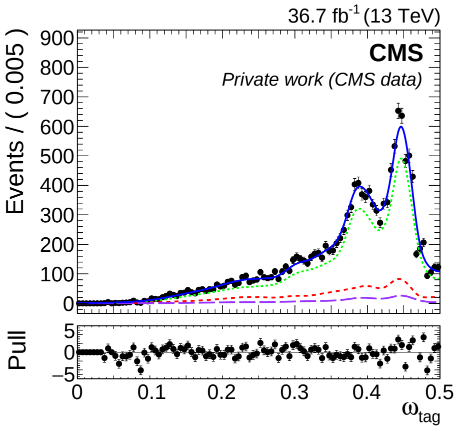
(a) Invariant mass distribution.

(b) PDL uncertainty distribution.



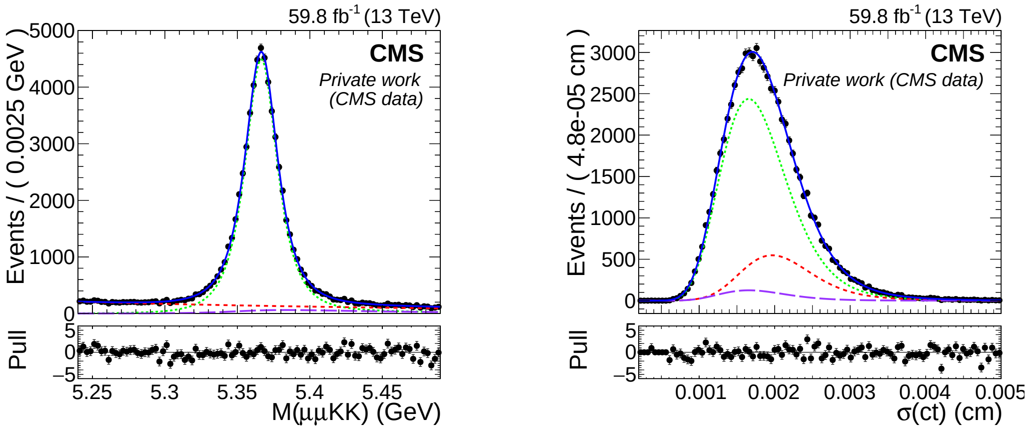
(c) $\cos\theta_T$ distribution.

(d) ω_{tag} distribution.



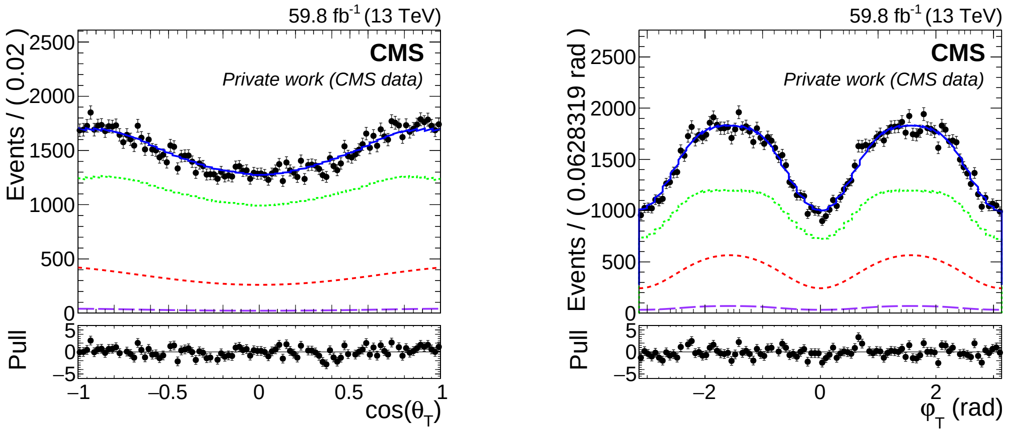
(e) ω_{tag} distribution.

Figure 13.22: Distribution of the input observables and the corresponding fit projections in *dataset_JpsiTrkTrk* (2017).



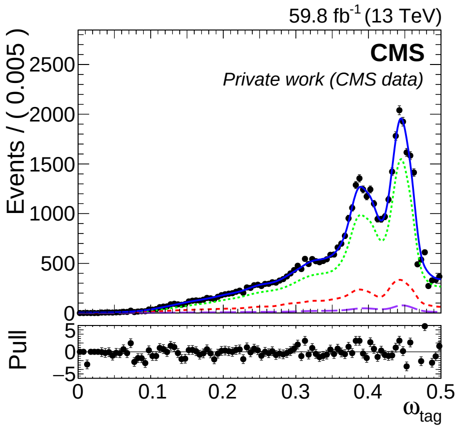
(a) Invariant mass distribution.

(b) PDL uncertainty distribution.



(c) cos θ_T distribution.

(d) ω_{tag} distribution.



(e) ω_{tag} distribution.

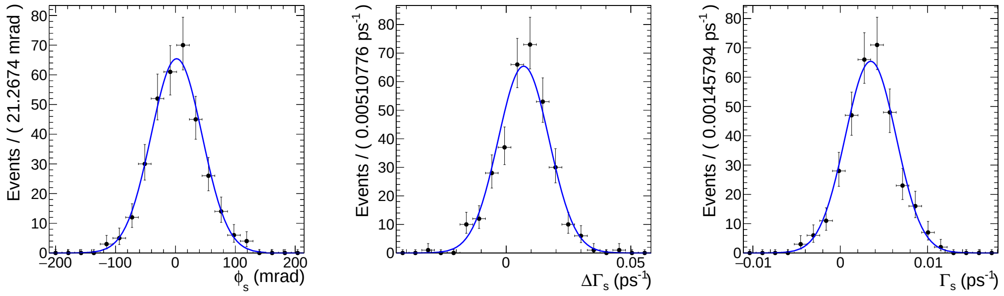
Figure 13.23: Distribution of the input observables and the corresponding fit projections in *dataset_JpsiTrkTrk* (2018).

Table 13.8: Statistical correlations between the main physics parameters of interest as evaluated with the same procedure as the statistical uncertainties. The lower half of the matrix is not reported as it is symmetrical. Correlations larger than 0.5 are underlined.

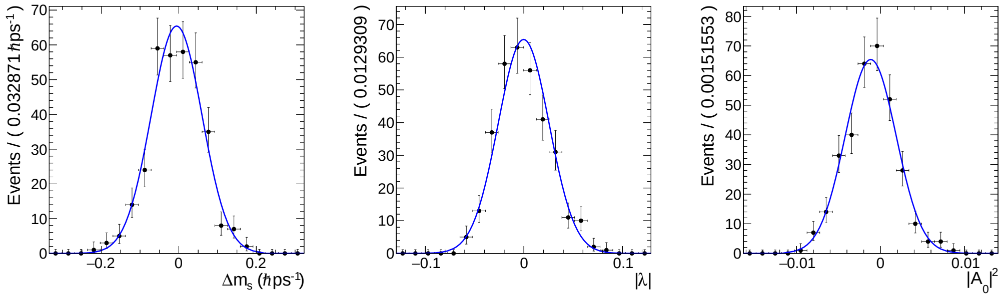
	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$	$\delta_\perp$	$\delta_{S\perp}$	Γ_s	$\Delta\Gamma_s$	Δm_s	$\lvert\lambda\rvert$	ϕ_s
$\lvert A_0\rvert^2$	+1.00	$\underline{-0.70}$	−0.067	−0.036	−0.0029	0.019	−0.39	$\underline{0.66}$	−0.0018	0.11	0.0030
$\lvert A_\perp\rvert^2$		+1.00	0.078	0.028	0.016	−0.037	0.48	$\underline{-0.74}$	0.015	−0.086	−0.027
$\lvert A_S\rvert^2$			+1.00	0.0055	0.047	$\underline{-0.57}$	0.017	−0.035	−0.083	−0.010	0.043
$\delta_\parallel$				+1.00	0.10	0.054	0.0083	−0.018	−0.0069	−0.035	−0.028
$\delta_\perp$					+1.00	−0.092	−0.029	−0.031	$\underline{0.80}$	0.00035	0.13
$\delta_{S\perp}$						+1.00	−0.019	0.030	0.013	−0.056	−0.035
Γ_s							+1.00	$\underline{-0.60}$	−0.0087	−0.042	−0.067
$\Delta\Gamma_s$								+1.00	−0.026	0.062	−0.011
Δm_s									+1.00	0.0041	0.10
$\lvert\lambda\rvert$										+1.00	0.14
ϕ_s											+1.00

Table 13.9: Central values and standard deviation of the distributions obtained from the pseudo-experiments generated to validate the fit model. The values are estimated as the arithmetic mean μ and the unbiased sample variance s.

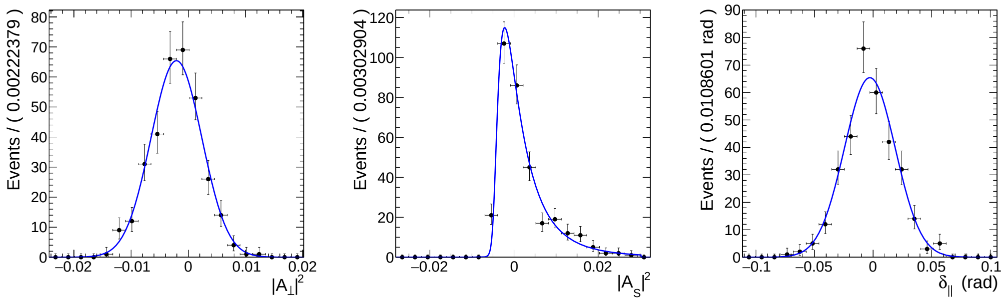
Parameter	Central value	Standard deviation
ϕ_s [mrad]	1.9	42
$\Delta\Gamma_s$ [ps^{-1}]	0.0070	0.010
Γ_s [ps^{-1}]	0.0035	0.0029
Δm_s [$\hbar$ps^{-1}]	−0.0057	0.066
$\lvert\lambda\rvert$	−0.0002	0.027
$\lvert A_0\rvert^2$	−0.0012	0.0030
$\lvert A_\perp\rvert^2$	−0.0021	0.0045
$\lvert A_S\rvert^2$	0.0022	0.0061
$\delta_\parallel$ [rad]	−0.0028	0.023
$\delta_\perp$ [rad]	−0.0151	0.14
$\delta_{S\perp}$ [rad]	0.123	0.37



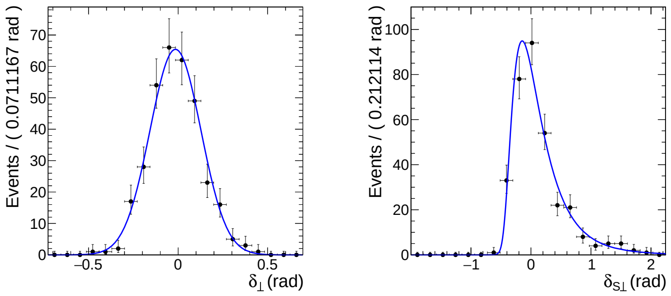
(a) ϕ_s distribution. (b) $\Delta\Gamma_s$ distribution. (c) Γ_s distribution.



(d) Δm_s distribution. (e) $|\lambda|$ distribution. (f) $|A_0|^2$ distribution.



(g) $|A_\perp|^2$ distribution. (h) $|A_S|^2$ distribution. (i) $\delta_\parallel$ distribution.



(j) $\delta_\perp$ distribution. (k) $\delta_{S\perp}$ distribution.

Figure 13.24: Distribution of the physics parameters as obtained from the 500 pseudo-experiments generated to validate the fit model. The solid line represents a gaussian distribution fit.

14

SYSTEMATIC UNCERTAINTIES

This chapter provides a description of possible systematic uncertainty sources in the ϕ_s measurement. Since the analysis is not the core of the thesis, only the systematic related to tagging are evaluated, in addition to the bias intrinsic to the model to provide a comparison of the uncertainty magnitude.

Possible systematic effects are divided into two categories:

- **Type-I: unaccounted uncertainties**. This category accounts for the finite statistics of the simulated samples and uncertainties in the parameters that are fixed in the final fit, such as those of the efficiencies and calibration functions. These uncertainties are always propagated to the final results and if found to be negligible the expression "$<10^{-x}$" will be reported. This class of uncertainties can be evaluated by re-performing the measurement several times ($\mathcal{O}(100)$), each time with a different *sampling* of the parameters whose uncertainty is to be propagated. The sampling follows the statistical uncertainties and correlations of the parameters in question. The RMS (or standard deviation) of the distribution of the results will be taken as the systematic uncertainty. Only one uncertainty of this type is provided here, the one related to the fit uncertainty in the flavor tagging calibration.

- **Type-II: methods and model assumptions**. This category accounts for the possible bias induced by the assumptions made in the fit model and the analysis methods. These assumptions are tested with a binary variation, where the alternative hypothesis is tested in the UML fit. If the difference Δ_{syst} between the new and reference results is deemed significant (a conservative choice was made to use $> 10\%$ of the statistical uncertainty as the limit), the systematic uncertainty can be evaluated as

  $$\sigma_{\text{syst}} = \frac{\Delta_{\text{syst}}}{2} \quad \text{(if relevant)}\,, \tag{14.1}$$

  assuming a uniform distribution for the bias. If the tests show no significant bias, the related systematic uncertainty will not be evaluated and "$-$" will be reported. The uncertainties of this type presented here are those related to the flavor tagging calibration method, i. e. assuming flavor equivalence between B^+ and B_s^0 mesons[I].

[I] Note that the OS and SS taggers here follow different choices: in OS taggers B^+ and B_s^0 mesons are assumed equivalent and the systematic uncertainty tests the case in which they are not; the SS tagger instead always corrects for the difference, and the systematic is meant to evaluate the effect of the correction.

The fit bias is the only systematic uncertainty that does not fall into either of the two categories.

14.1 FIT BIAS

Possible biases arising from the fitting procedure are evaluated from the reference bootstrap distributions described in paragraph 13.1.2, sharing the generated replicas used for the statistical error estimation (the variance of the parameters in the bootstrap sample is used as statistical error, while the average is used here for the fit bias).

For each parameter, the central value of the corresponding bootstrap distribution is taken as the corresponding systematic uncertainty, if different from zero by more than its error:

$$\sigma_{\text{bias}}(V) = \left|\widehat{t(V)}\right|, \quad \text{if} \left|\frac{\widehat{t(V)}}{\sigma_{\widehat{t(V)}}}\right| > 1\,, \tag{14.2}$$

where $\widehat{t(V)}$ is the arithmetic mean of the distribution and $\sigma_{\widehat{t(V)}} = \text{RMS}(t(V))/\sqrt{N_{\text{BS}}}$ is its uncertainty. The obtained systematic uncertainties are reported in Tab. 14.1.

Table 14.1: Summary of the systematic uncertainties related to the fit bias. The dashes (–) indicate that the uncertainty is not evaluated as no bias is observed.

Parameter	Central value
ϕ_s [mrad]	–
$\Delta\Gamma_s$ [ps^{-1}]	0.0070
Γ_s [ps^{-1}]	0.0035
Δm_s [$\hbar$ps^{-1}]	0.0057
$\lvert\lambda\rvert$	–
$\lvert A_0\rvert^2$	0.0012
$\lvert A_\perp\rvert^2$	0.0021
$\lvert A_S\rvert^2$	0.0022
$\delta_\parallel$ [rad]	0.0028
$\delta_\perp$ [rad]	0.0151
$\delta_{S\perp}$ [rad]	0.123

14.2 FLAVOR CALIBRATION UNCERTAINTY

Each of the calibration functions for the taggers described in Part III has an associated statistical uncertainty in the parameters, due to the limited statistic of the calibration sample. This statistical uncertainty is propagated to the final results by repeating the fit 200 times, each time with a random variation on *all* the calibration parameters and taking into account the correlations between the parameters in the same calibration fit (i. e. a and b). In case of the Same Side tagger, the variation is performed simultaneously on all three steps of the calibration (B^+ MC, B_s^0 MC and B^+ data). To be more conservative in the estimation, each uncertainty is multiplied by the normalized χ^2 of the respective fit (usually around $2 \sim 3$). The RMS of the distribution of each extracted physics parameter is taken as the systematic uncertainty. The systematic uncertainties obtained are reported in Tab. 14.2.

Table 14.2: Summary of the systematic uncertainties related to the flavor calibration uncertainty. Uncertainty negligible with respect to the statistic uncertainty are shown as $< 10^n$, where 10^n is the order of the last digit of the statistical uncertainty.

Parameter	RMS
ϕ_s [mrad]	1
$\Delta\Gamma_s$ [ps^{-1}]	$< 10^{-3}$
Γ_s [ps^{-1}]	$< 10^{-4}$
Δm_s [$\hbar$ps^{-1}]	$< 10^{-3}$
$\|\lambda\|$	$< 10^{-3}$
$\|A_0\|^2$	$< 10^{-4}$
$\|A_\perp\|^2$	$< 10^{-4}$
$\|A_S\|^2$	0.0001
$\delta_\parallel$ [rad]	0.001
$\delta_\perp$ [rad]	$< 10^{-3}$
$\delta_{S\perp}$ [rad]	$< 10^{-2}$

14.3 CALIBRATION METHOD IN OS TAGGING ALGORITHMS

In the calibration of OS taggers, an assumption is made that the algorithm's behavior when the signal meson is a B^+ is the same as when it is a B_s^0. This assumption is not unfounded, as OS taggers are explicitly constrained to avoid the region where the signal meson can interfere, with a $\Delta R > x$ cut around the signal. Nonetheless, a systematic uncertainty is associated with the hypothesis in which the response is different. As a reminder, the original calibration function is obtained from a fit in a *data* sample of $B^+ \to J/\psi K^+$ decays. A correction function is obtained from simulated $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\,\phi(1020)$ decays and combined with the original calibration. The new calibration follows the flavor correction procedure described in Eq. 10.9 for the SS tagger: the role of the correction is, after all, the same, the only difference being that in the SS tagger the B^+-B_s^0 difference is expected to be significant.

This is a Type-II systematic, and is thus implemented as a single fit with the new calibration, where the half-difference is taken as value for the uncertainty. The results are provided in Tab. 14.3.

Table 14.3: Summary of the systematic uncertainties related to the OS taggers calibration B^+-B_s^0 correction.

Parameter	Half-difference
ϕ_s [mrad]	–
$\Delta\Gamma_s$ [ps^{-1}]	–
Γ_s [ps^{-1}]	–
Δm_s [$\hbar$ps^{-1}]	0.029
$\lvert\lambda\rvert$	0.006
$\lvert A_0\rvert^2$	–
$\lvert A_\perp\rvert^2$	–
$\lvert A_S\rvert^2$	0.0015
$\delta_\parallel$ [rad]	0.029
$\delta_\perp$ [rad]	0.040
$\delta_{S\perp}$ [rad]	0.13

14.4 CALIBRATION METHOD IN SS TAGGING ALGORITHMS

The calibration of the SS-tagger relies on simulated samples to correct for differences between the hadronization of the B_s^0 and B^+ mesons. We test the impact of a possible mismodeling in MC and biases in this procedure by repeating the fit without applying this correction, that is, using directly the calibration obtained from B^+ data. This is possibly the most *extreme* stress test possible for the SS-tagger, as it changes the calibration parameters by almost 10%[II].

This too is a Type-II systematic, and the results are provided in Tab. 14.4.

14.5 SYSTEMATIC UNCERTAINTIES SUMMARY

The systematic uncertainties are summarized in Tab. 14.5. The final systematic uncertainty in the CP-violating phase ϕ_s is estimated to be 1 mrad, 2% of the statistical one.

[II] The b parameter is changed from 0.895 (MC corrected value) to 0.970 (measured in B^+ data).

Table 14.4: Summary of the systematic uncertainties related to the SS tagger calibration B^+-B_s^0 correction.

Parameter	Half-difference
ϕ_s [mrad]	–
$\Delta\Gamma_s$ [ps^{-1}]	–
Γ_s [ps^{-1}]	–
Δm_s [$\hbar$ps^{-1}]	–
$\lvert\lambda\rvert$	–
$\lvert A_0\rvert^2$	–
$\lvert A_\perp\rvert^2$	–
$\lvert A_S\rvert^2$	–
$\delta_\parallel$ [rad]	0.0024
$\delta_\perp$ [rad]	–
$\delta_{S\perp}$ [rad]	–

The fit bias is, as expected, the dominant systematic uncertainty source for several parameters, especially those not related to flavor tagging such as Γ_s, $\Delta\Gamma_s$, or the amplitudes.

Fluctuations of the calibration functions, as can be noticed, do not cause large differences in the physics parameters. Among the tag systematic uncertainties, however, this was the only one that generated a significant, albeit small, uncertainty in ϕ_s.

The SS B^+-B_s^0 uncertainty is only relevant in the $\delta_\parallel$ variable, and even there it is negligible compared both to the statistical uncertainty and to the OS B^+-B_s^0 calibration. This is likely caused by the fact that by construction all possible variations on the SS calibration are symmetrical and affect in the same way B_s^0 and $\overline{B}_s^0$ mesons. Additionally, the strength of the SS tagger lies in its high efficiency and not in the low dilution; this means that most variation caused by changes in the calibration affect events close to $\omega_{\text{tag}} = 0.5$, where the change is small. The result proves the stability of the oscillation parameters with respect to the Same Side calibration method.

Finally, the B^+-B_s^0 factorization in the OS is possibly the largest tag-related uncertainty. The difference between OS and SS is that in the OS case, specifically for lepton taggers, is the presence of the intercept a in the calibration. Shifts in a have two effects: they affect equally high and low mistag events, and can cause a change in ξ_{tag}. The different response to changes in the OS and SS calibration can be interpreted as the analysis being robust against factors of ω_{tag}, but less robust against shifts in ξ_{tag}.

Table 14.5: Summary of the systematic uncertainties for the physics parameters. The dashes (–) indicate 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.

	ϕ_s	$\Delta\Gamma_s$	Γ_s	Δm_s	$\lvert\lambda\rvert$	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$	$\delta_\perp$	$\delta_{S\perp}$
	[mrad]	[ps^{-1}]	[ps^{-1}]	[$\hbar$ps^{-1}]					[rad]	[rad]	[rad]
Stat. uncertainty	42	0.010	0.0029	0.066	0.027	0.0030	0.0045	0.0061	0.023	0.14	0.37
Fit bias	–	0.0070	0.0035	0.0057	–	0.0012	0.0021	0.0022	0.0028	0.0151	0.123
Tag calib. uncertainty	1	$< 10^{-3}$	$< 10^{-4}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	$< 10^{-4}$	0.0001	0.001	$< 10^{-3}$	$< 10^{-2}$
OS tagging B^+-B_s^0	–	–	–	0.029	0.006	–	–	0.0015	0.029	0.040	0.13
SS tagging B^+-B_s^0	–	–	–	–	–	–	–	–	0.0024	–	–
Total syst. uncertainty	1	0.0070	0.0035	0.030	0.006	0.0012	0.0021	0.0027	0.029	0.043	0.18

15

SUMMARY AND CONCLUSIONS

This thesis presents the development and calibration of four DNN based flavor tagging algorithms used for CP violation measurements in CMS.

The OS-lepton based taggers have relatively simple physics interpretation as they are based on semileptonic decays of opposite side b-quarks, and prove to have a great performance in CMS, having a tagging power P_{tag} equal to

$$P_{\text{tag}}^{\text{OS-}\mu} = 1.535 \pm 0.006\%$$

$$P_{\text{tag}}^{\text{OS-}e} = 0.4096 \pm 0.0003\% \, .$$

The OS-μ tagger in particular makes use of the great muon reconstruction capabilities of CMS, and can optionally be used at trigger level to select events with no bias on the proper decay length.

The jet based tagger uses b-tagged jets to identify the direction of the opposite side b-hadron, and looks inclusively at the tracks associated to the jet to infer its flavor. The algorithm uses a *DeepSets* based architecture to avoid depending on a choice of order in the input tracks. The OS-jet tagger was found to have a performance comparable to that of the electron based tagger

$$P_{\text{tag}}^{\text{OS-jet}} = 0.4571 \pm 0.0004\% \, ,$$

limited mostly by the poor jet clustering performances in the low-energy regimes used in B-Physics analyses

The Same Side tagger represents the most important innovation among the ones developed during this thesis. The tagger analyzes inclusively all the activity around the signal meson, and is the first same side tagger able to infer the flavor without access to particle identification, which is not available in CMS. With a P_{tag} of

$$P_{\text{tag}}^{\text{SS}} = 3.39 \pm 0.01\% \, ,$$

the tagger is a factor of 2 above the second best tagger of the suite (the OS-μ tagger). The great performance comes at the cost of ease of calibration, since this tagger cannot assume equivalence in the response to B^+ and B_s^0 mesons. The final selected method of difference factorization using simulated samples however proved to be stable and to cause minimal systematic uncertainties in a real world measurement of CP violation. The tagger can also trivially be augmented with PID information, if available. A simulation using the predicted performance of the future MTD timing detector in CMS showed a 14% improvement in tagging power in a worst case scenario and 24% in a standard scenario.

The total performance of the suite of taggers, also taking into account the statistical combination of probabilities from different independent sources, was measured to be

$$P_{\text{tag}}^{\text{tot}} = 5.59 \pm 0.02\% ,$$

one of the highest among hadronic collider experiments.

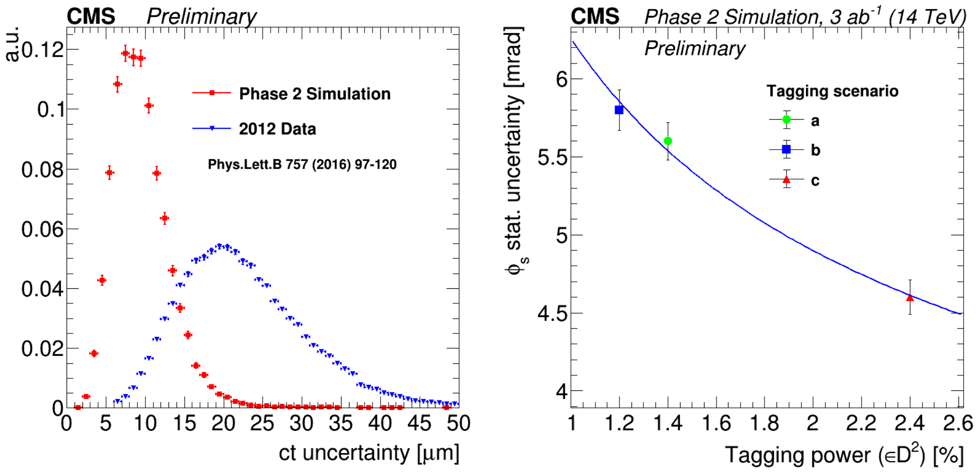
A measurement of the CP violating phase ϕ_s in the $B_s^0 \to J/\psi\,\phi(1020)$ channel was performed as a benchmark for the performance of the tagging algorithms and to evaluate the impact that the calibration methods have in a real analysis. The analysis was performed on $96.48\,\text{fb}^{-1}$ collected in 2017 and 2018, but was artificially reduced in statistics to avoid the unblinding of an official analysis ongoing in CMS. The analysis showed a good precision in the measurement of ϕ_s compared to previous CMS results [4], which this measurement subsumes. All measured parameters were found consistent with their values in the world average. The final result for ϕ_s was

$$\phi_s = -21 \pm 42(\text{stat.}) \pm 1(\text{model bias} + \text{tag})\ \text{mrad}$$

Most importantly, an evaluation of the tagging related systematic uncertainties showed them to be negligible compared to the statistical ones and at most comparable with the intrinsic bias of the UML fit procedure.

The official ϕ_s measurement in CMS is proceeding using the full statistics available. The addition of all the taggers developed for this thesis will greatly enhance the precision of the analysis, which will achieve a projected statistical uncertainty for ϕ_s of ~ 25 mrad.

Going further in the future, all future projections predict the ϕ_s measurement to be statistically limited. 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. [172, 173]. Figure 15.1 shows the simulated proper decay length uncertainty of the Phase-2 CMS detector and the ϕ_s sensitivity at the end of Phase-2. The projected sensitivity on ϕ_s is given in Ref. [172] as a function of P_{tag}, and reaches up to $P_{\text{tag}} = 2.6\%$ as a best case scenario, with a precision in the 5–6 mrad range. Extrapolating outside the axis range, the measured $P_{\text{tag}} \sim 5.6\%$ in this thesis leads to a predicted precision of $\sim$ 3–4 mrad.



(a) Proper decay length uncertainty distribution in Phase-0 and Phase-2 simulations.

(b) Expected ϕ_s sensitivity in Phase-2.

Figure 15.1: *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 3.2.2). *b*): ϕ_s sensitivity in Phase-2 as a function of the available tagging power, obtained from MC pseudo-experiments. Figure from Ref. [172].

BIBLIOGRAPHY

[1] 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.

[2] 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.

[3] 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.

[4] CMS Collaboration. "Measurement of the CP-violating phase ϕ_s in the $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^-\,K^+K^-$ channel in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *Physics Letters B* 816 (2021), p. 136188. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2021.136188.

[5] J. Goldstone, A. Salam, and S. Weinberg. "Broken symmetries." In: *Phys. Rev.* 127 (1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965.

[6] S. Glashow. "Partial symmetries of weak interactions." In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[7] 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.

[8] 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.

[9] S. Weinberg. "A model of leptons." In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[10] 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.

[11] 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.

[12] 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].

[13] 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].

[14] 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].

[15] 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/.

[16] M. Gell-Mann. "Symmetries of baryons and mesons." In: *Phys. Rev.* 125 (3 Feb. 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067.

[17] J Goldstone. "Field theories with "superconductor" solutions." In: *Nuovo Cimento* 19.CERN-TH-118 (Aug. 1960), pp. 154–164. DOI: 10.1007/BF02812722.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] N. Cabibbo. "Unitary symmetry and leptonic decays." In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[23] E. Noether. "Invariante variationsprobleme." (German). In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918), pp. 235–257.

[24] Particle Data Group. "Review of particle physics." In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[25] Y. Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021." In: *Phys. Rev. D* 107 (5 2023), p. 052008. DOI: 10.1103/PhysRevD.107.052008.

[26] 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.

[27] 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.

[28] R. Jost. “Eine bemerkung zum CPT-Theorem.” (German). In: *Hel. Phys. Acta* 30 (1957), pp. 409–416.

[29] 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`.

[30] 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`.

[31] 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]`.

[32] L. Wolfenstein. “Parametrization of the Kobayashi-Maskawa matrix.” In: *Phys. Rev. Lett.* 51 (1983), p. 1945. DOI: `10.1103/PhysRevLett.51.1945`.

[33] CKMfitter Group. *CKMfitter*. `http://ckmfitter.in2p3.fr`. Accessed: 2022-06-17. 2022.

[34] 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`.

[35] 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`.

[36] 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`.

[37] 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`.

[38] 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`.

[39] 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`.

[40] 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]`.

[41] BaBar Collaboration. "Measurement of time-dependent CP asymmetry in $B^0 \rightarrow c\bar{c}K^{(*)0}$ decays." In: *Phys. Rev. D* 79 (2009), p. 072009. DOI: `10.1103/PhysRevD.79.072009`. arXiv: `0902.1708 [hep-ex]`.

[42] 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`.

[43] 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`.

[44] 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`.

[45] 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`.

[46] 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`.

[47] 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`.

[48] 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`.

[49] 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]`.

[50] 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]`.

[51] 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.

[52] J. Wenninger. "Approaching the Nominal Performance at the LHC." In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14â19 May, 2017* (Copenhagen, Denmark). International Particle Accelerator Conference 8. https://doi.org/10.18429/JACoW-IPAC2017-MOYAA1. Geneva, Switzerland: JACoW, 2017, pp. 13–18. ISBN: 978-3-95450-182-3. DOI: https://doi.org/10.18429/JACoW-IPAC2017-MOYAA1.

[53] R Bruce. "REVIEW OF LHC RUN 2 MACHINE CONFIGURATIONS." In: Virtual, Feb. 2019. URL: https://indico.cern.ch/event/751857/timetable/#32-machine-configuration-20-10.

[54] E. Mobs. *The CERN accelerator complex - 2019*. General Photo. July 2019. URL: https://cds.cern.ch/record/2684277.

[55] A. Team. *Diagram of an LHC dipole magnet*. June 1999. URL: https://cds.cern.ch/record/40524.

[56] ALICE Collaboration. "The ALICE experiment at the CERN LHC." In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[57] ATLAS Collaboration. "The ATLAS experiment at the CERN Large Hadron Collider." In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[58] LHCb Collaboration. "The LHCb detector at the LHC." In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[59] 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).

[60] C. Pralavorio. *Record luminosity: well done LHC*. Nov. 2017. URL: https://cds.cern.ch/record/2295027.

[61] CMS Collaboration. *CMS luminosity public results*. 2019. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Online; accessed November 2020.

[62] G. Antchev et al. "First measurement of elastic, inelastic and total cross-section at $$\sqrt{s}=13$$ TeV by TOTEM and overview of cross-section data at LHC energies." In: *The European Physical Journal C* 79.2 (2019), p. 103. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6567-0.

[63] 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]. arXiv: 2010.11906 [hep-ph].

[64] W. Sirling. *Private communication*. URL: http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html. Online; accessed November 2020.

[65] CMS Collaboration. "The CMS experiment at the CERN LHC." In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[66] 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.

[67] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Tech. rep. CERN-LHCC-97-010. 1997. URL: https://cds.cern.ch/record/331056.

[68] 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].

[69] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Tech. rep. CERN-LHCC-98-006. 1997. URL: http://cds.cern.ch/record/368412.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-031. 1997. URL: http://cds.cern.ch/record/357153.

[75] CMS Collaboration. *The CMS muon project: Technical Design Report*. Tech. rep. CERN-LHCC-97-032. 1997. URL: http://cds.cern.ch/record/343814.

[76] 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].

[77] CMS Collaboration. *The Phase-2 upgrade of the CMS tracker*. Tech. rep. CMS-TDR-014. June 2017. URL: https://cds.cern.ch/record/2272264.

[78] CMS Collaboration. *CMS tracker detector performance results*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK. Online; accessed November 2020.

[79] 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.

[80] 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)].

[81] 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].

[82] CMS Collaboration. *CMS computing: Technical Design Report*. Tech. rep. CERN-LHCC-2005-023. 2005. URL: https://cds.cern.ch/record/838359.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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].

[89] 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].

[90] 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.

[91] 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.

[92] 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.

[93] T. C. Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker.” In: *Journal of Instrumentation* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009.

[94] “2017 tracking performance plots.” In: (2017).

[95] 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.

[96] 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.

[97] “Muon Reconstruction and Identification Performance with Run-2 data.” In: (2020).

[98] 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].

[99] A. M. Sirunyan et al. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC.” In: *JINST* 16.05 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[100] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-kt jet clustering algorithm.” In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063.

[101] 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.

[102] 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.

[103] 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.

[104] 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`.

[105] 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`.

[106] 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]`.

[107] 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]`.

[108] 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]`.

[109] 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`.

[110] 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`.

[111] 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`.

[112] CMS Collaboration. *CMS Offline Software*. URL: `http://cms-sw.github.io/`. Online; accessed November 2020.

[113] 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`.

[114] 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/`.

[115] W. Verkerke and D. Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: `physics/0306116 [physics.data-an]`.

[116] 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`.

[117] D. P. Kingma and J. Ba. “Adam: a method for stochastic optimization.” 2017. arXiv: `1412.6980 [cs.LG]`. arXiv: 1412.6980 [cs.LG].

[118] 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]`. arXiv: 1207.0580 [cs.NE].

[119] 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`.

[120] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. “On calibration of modern neural networks.” 2017. arXiv: `1706.04599 [cs.LG]`. arXiv: 1706.04599 [cs.LG].

[121] J. C. Platt. “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods.” In: *Advances in Large Margin Classifiers*. MIT Press, 1999, pp. 61–74.

[122] F. Chollet et al. *Keras*. 2015. URL: `https://keras.io`. Online; accessed November 2020.

[123] 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`.

[124] K. Albertsson, S. Gleyzer, A. Hoecker, L. Moneta, P. Speckmayer, J. Stelzer, J. T. E. von Toerne, H. Voss, and S. Wunsch. *TMVA User Guide*. 2020.

[125] 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.

[126] 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.

[127] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia. “Jet flavour classification using DeepJet.” In: *Journal of Instrumentation* 15.12 (2020), P12012. DOI: `10.1088/1748-0221/15/12/P12012`.

[128] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola. “Deep Sets.” 2017. arXiv: `1703.06114 [cs.LG]`. arXiv: 1703.06114 [cs.LG].

[129] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. 2016. arXiv: `1511.07289 [cs.LG]`.

[130] F. Sikler. *Spectra of Charged Pions, Kaons, and Protons Identified via Tracker Energy Loss from CMS*. Tech. rep. Geneva: CERN, 2012.

[131] 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)].

[132] LHCb Collaboration. "A new algorithm for identifying the flavour of B_s^0 mesons at LHCb. A new algorithm for identifying the flavour of B_s^0 mesons at LHCb." In: *JINST* 11 (2016), P05010. DOI: `10.1088/1748-0221/11/05/P05010`. arXiv: `1602.07252`.

[133] C. CMS. *A MIP Timing Detector for the CMS Phase-2 Upgrade*. Tech. rep. Geneva: CERN, 2019.

[134] "Update of the MTD physics case." In: (2022).

[135] ATLAS Collaboration. "Measurement of the *CP*-violating phase ϕ_s in $B_s^0 \to J/\psi\phi$ decays in ATLAS at 13 TeV." In: *Eur. Phys. J. C* 81.4 (2021), p. 342. DOI: `10.1140/epjc/s10052-021-09011-0`. arXiv: `2001.07115 [hep-ex]`.

[136] 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]`.

[137] 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`.

[138] 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`.

[139] 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]`.

[140] 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`.

[141] CMS Collaboration. “Measurement of the CP-violating phase ϕ_s in the $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ channel in proton-proton collisions at $\sqrt{s} = 13$ TeV.” July 2020. arXiv: 2007.02434 [hep-ex]. arXiv: 2007.02434 [hep-ex]. Submitted to Phys. Lett. B.

[142] 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].

[143] 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].

[144] 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].

[145] 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].

[146] 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].

[147] 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].

[148] 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].

[149] 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].

[150] 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].

[151] 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].

[152] 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].

[153] 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].

[154] 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].

[155] 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].

[156] 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].

[157] 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].

[158] R. Workman et al. "Review of Particle Physics." In: (). to be published (2022), accessed: 2022-06-17.

[159] 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.

[160] 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)].

[161] 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].

[162] R. A. Fisher. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd, 1925.

[163] 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.

[164] R. J. Barlow. "Extended maximum likelihood." In: *Nucl. Instrum. Meth. A* 297 (1990), pp. 496–506. DOI: 10.1016/0168-9002(90)91334-8.

[165] 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.

[166] C. Langenbruch. "Parameter uncertainties in weighted unbinned maximum likelihood fits." In: *The European Physical Journal C* 82.5 (2022), p. 393. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10254-8.

[167] B. Efron. "Bootstrap Methods: Another Look at the Jackknife." In: *The Annals of Statistics* 7.1 (1979), pp. 1 –26. DOI: 10.1214/aos/1176344552.

[168] 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.

[169] B. W. Silverman. "Kernel Density Estimation using the Fast Fourier Transform." In: *Journal of the Royal Statistical Society* Series C (Applied statistics) (1982).

[170] A. Gramacki and J. Gramacki. "FFT-based Fast Computation of Multivariate Kernel Density Estimators with Unconstrained Bandwidth Matrices." In: *arXiv* (2016).

[171] S. Weglarczyk. "Kernel density estimation and its application." In: *ITM Web of Conferences* (2018).

[172] CMS Collaboration. *CP-violation studies at the HL-LHC with CMS using* B^0_s *decays to* $\mathrm{J}/\psi\phi(1020)$. Tech. rep. CMS-PAS-FTR-18-041. 2018. URL: http://cds.cern.ch/record/2650772.

[173] 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].