SOFIA UNIVERSITY
ST. KLIMENT OHRIDSKI

Atomic Physics Department
University of Sofia "St. Kliment Ohridski"
Bulgaria

Machine Learning methods for precision measurements and detector automation at the CMS experiment

Author:
Elton Shumka

Scientific advisors:
Assoc. Prof. Borislav Pavlov
Assoc. Prof. Peicho Petkov

A thesis presented for the degree of
Doctor of Philosophy

Sofia, December 2024

Abstract

The work presented in this dissertation encompasses three main topics: physics analysis, detector monitoring with anomaly detection capabilities and detector automation. All three are characterized by the extensive use of Machine Learning (ML) algorithms and methods, in light of their boost in popularity and utility in the past decade. In physics analysis, such as in the B_s^0 precision measurement presented here, ML methods prove to be a valuable tool for increasing effective statistics by enhancing the tagging capability and for accounting for mismodeling effects in the Monte Carlo samples by facilitating a powerful per-event reweighting procedure. In detector monitoring, several complementary ML approaches are used for anomaly detection, crucial for anticipating detector failures in the near future. Lastly, the detector automation methodology relies on ML in order to implement a self-correctional feedback mechanism, which strives to automate detector operation.

Contents

Introduction

Part I of this dissertation gives a comprehensive overview of the detector setup and the theoretical background for the performed studies.

Chapter 1 introduces the Compact Muon Solenoid (CMS) experiment, in the context of the LHC accelerator complex and provides a detailed description of its subsystems: Tracker (TRK), Electromagnetic Calorimeter (ECAL), Hadronic Calorimeter (HCAL), and the Muon System. Furthermore, a deeper treatment of the Resistive Plate Chambers (RPC) subsystem, central to the two of the three works comprising the dissertation, is given, including their construction, properties and operation. A motivation for the necessity of a more sophisticated approach to operation strategies is also outlined.

Chapter 2 gives an overview of the theoretical foundation for the performed studies, starting with a general introduction to the Standard Model (SM) of particle physics. The subsequent sections focus on the topic of Flavor Physics, serving as a framework for the description of CP symmetry-violating (CPV) effects. B mesons are treated in detail, with special emphasis on the different categories of CPV scenarios that can arise in their decays. Lastly, the $B_s^0 \to J/\psi\phi(1020)$ decay is considered and the theoretical tools for its study are introduced. The main physics parameters relevant to this decay are presented, along with the state-of-the-art theoretical estimations and experimental measurements of their values.

Part II provides an exposition of the performed physics analysis work and the tools developed for detector monitoring and automation, and their corresponding results. Chapter 3 concerns the precision measurement of CPV effects in the $B_s^0 \to J/\psi\phi(1020)$ decay. A description of the used data sets, as well as the reconstruction and selection algorithms is provided as the foundational step of the analysis. The Deep Neural Network (DNN) based flavor-tagging strategy is considered in more detail, with a discussion of the steps taken to refine its performance and a comparison with earlier studies. The evaluation of the angular efficiency, crucial for the angular analysis performed on the decay's final state, is described in a dedicated section. Details are given about the Kernel Density Estimate (KDE) method paired with Fast-Fourier Transform (FFT), employed in order to tackle the computational complexity of the estimation of the multivariate angular efficiency function. The subsequent section showcases a novel per-event reweighting procedure, developed to deal with the intrinsic and unavoidable mismodeling in the produced Monte Carlo samples. The fit model,

incorporating all the components discussed up to that point, is presented. A subsequent short discussion of the fitting procedures serves as a lead up to the final results. The impact of the results is considered, along with a comparison to the previous measurements and estimations, introduced in Chapter 2.

Chapter 4 covers the second contribution constituting this dissertation, namely the ML-based tool for CMS RPC currents monitoring. It contains a description of the two basic approaches for anomaly detection that are used in this work: Generalized Linear Models (GLM) and Autoencoders. The benefits of both and their software implementation are discussed. The harnessing of the models' predictive power and the integration of their output in a decision-making logic is explained. The chapter ends with a section on the tool's deployment and the interface provided to the user.

Chapter 5 is dedicated to the RPC automation framework, intended for processing and analysis of non-event physics data from the RPC subsystem of CMS. The goal behind the development of this tool, and the problematic it intends to solve, serve as an entry point to an overview of the novel data-handling methods designed to address the numerous shortcomings of the existent infrastructure. The automation is segmented in fundamental task-processing units called automata, which are described individually, covering the methods used in each and the analysis for which they contribute. The chapter is finalized with a comprehensive vision of the automation framework as an integral part of a self-correctional feedback mechanism.

Part I

Motivation

Chapter 1

Experimental setup

1.1 The Large Hadron Collider

The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) is the most powerful particle accelerator build to-date. It is located in a 26.7 km tunnel at the French-Swiss border near Geneva, at an average depth under ground of $\sim$ 100 m. The tunnel was originally built for the Large Electron Positron accelerator, operated at CERN from 1989 to 2000s. LHC is part of the CERN accelerator complex (Fig. 1.1) and it accelerates protons up to energies of 13.6 TeV and heavy ions to 2.76 TeV per nucleon.

The acceleration of protons, before they reach the LHC, proceeds in four steps, each being performed by one of four accelerators in the chain, respectively. The first link in the chain is the linear accelerator LINAC 4 [1]. It receives hydrogen nuclei with an energy of 95 keV from a H^- source, forms a beam and accelerates it to 160 MeV.

In the next step, the proton beam from the linear accelerator is injected into the Proton Synchrotron Booster (PSB) [2]. It is a synchrotron with a circumference of 157 m. PSB accelerates the proton beam from 160 MeV to 2 GeV.

The beam is then injected into the Proton Synchrotron (PS) [3]. PS is the oldest operating synchrotron at CERN and has a circumference of 628 m. At the end of the acceleration in PS, the proton beam has reached an energy of up to 26 GeV. Apart from accelerating the beam, PS groups the beam into "bunches", where each bunch contains $\approx 1.15 \times 10^{11}$ protons.

In the final step, the proton beam is injected in the Super Proton Synchrotron (SPS) [4]. SPS has a circumference of 6.9 km and is the first accelerator in the chain to be located underground, at $\sim$ 30 m. Here, the beam obtains an energy of 450 GeV. The beam from SPS serves both the LHC as well as experiments situated in the CERN Prevessin area.

LHC contains 1232 main dipole magnets [7] (Fig. 1.2) of 15 m length each which steer the beam around the circular beam pipes. The dipoles are composed

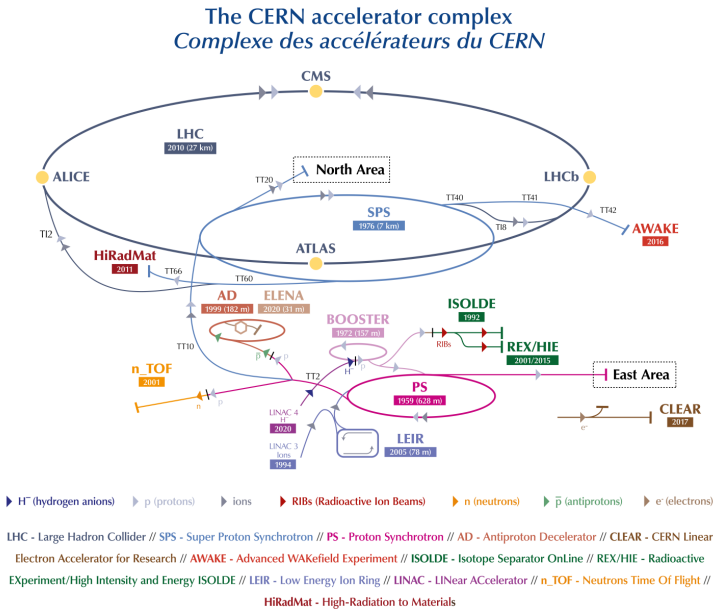


Figure 1.1: The accelerator complex at CERN. Image from [5].

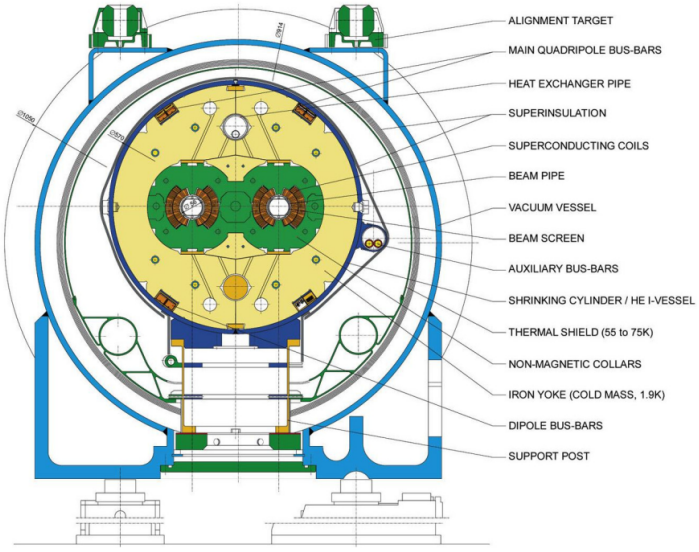


Figure 1.2: A cross section of an LHC dipole magnet. Image from [6].

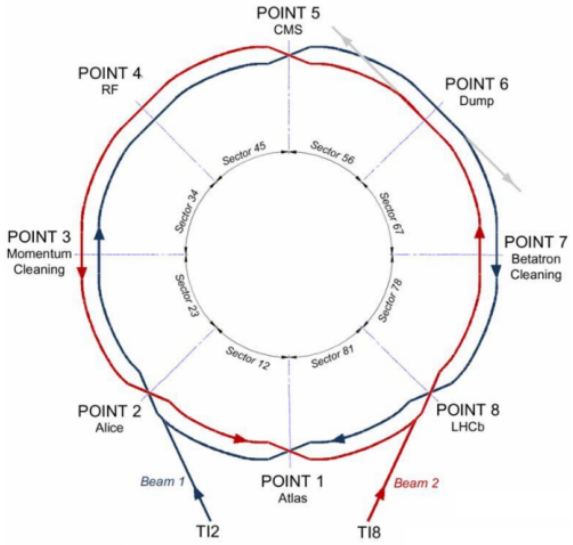


Figure 1.3: An overview of the LHC and the intersecting points on it, where the four major detectos are placed

of superconducting Cu-clad Nb-Ti and operate at the temperature of superfluid helium (1.9 K). A current of 12000 A flows through the dipoles' superconductor, producing a nominal magnetic field strength of 8.33 T.

Another set of magnets in the LHC are the 392 quadrupole magnets. The quadrupole magnets perform the function of focusing the proton beam. Each consecutive magnet focuses the beam in a direction perpendicular to the previous one. An illustration of this principle is given in Fig. 1.4.

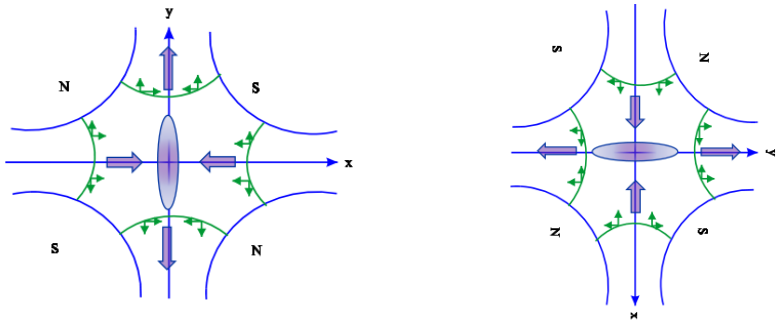


Figure 1.4: Illustration of the operation principle of quadrupole magnets.

The acceleration of the beam in the LHC occurs at eight locations along the accelerator ring, where Radio Frequency (RF) cavities are placed (Fig. 1.5).

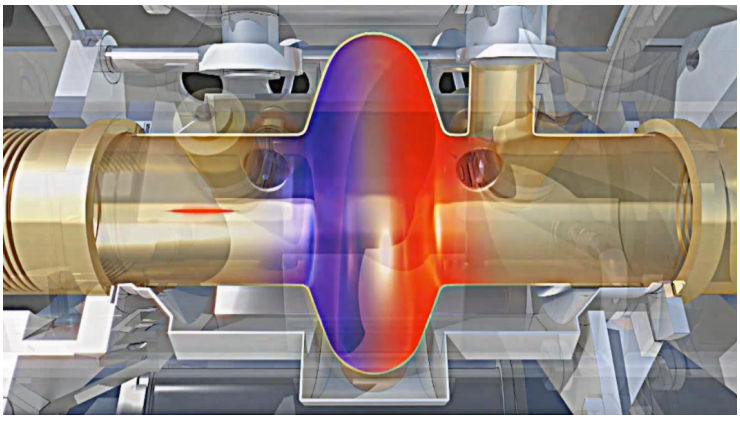


Figure 1.5: A graphical representation of an RF cavity, where the electric field is visualized in red and blue. Image from [8].

The electric field inside the RF cavities alternates with a harmonic of the revolution frequency of the beam, gradually accelerating each beam to the nominal energy of 6.5 TeV. With each turn, the beam gains 16 MeV of energy.

The full process takes 20 to 25 minutes, and once the beam reaches the nominal energy, it can be collided with the other beam circulating in a direction opposite to it. The center of mass energy during the collision is given by:

$$\sqrt{s} = 2E_{beam} \tag{1.1}$$

There are four locations along the accelerator ring where the beam pipes intersect and the collisions occur, as shown in Fig. 1.3. These are the locations where the four major LHC experiments are placed, namely: ATLAS [9], ALICE [10], CMS [10] and LHCb [11].

ALICE is an experiment (Fig. 1.6) devoted to the study of heavy-ion collisions and the associated processes, including the state of matter known as Quark-Gluon Plasma (QGP) [12].

LHCb, as the name of the experiment suggests, focuses on the study of b-quarks and the processes involving hadrons containing the b-quark and its antiparticle $\bar{b}$, a field of study known as B-physics. LHCb has gathered significant statistics for B-physics processes, giving contribution to questions related to matter-antimatter asymmetry, Beyond the Standard Model physics (BSM), CP violation[1] etc. As can be seen in Fig. 1.7, the LHCb detector does not enclose the interaction point (IP) as do the other three detectors described here, but rather sits on one side of the IP. It studies the products of the collision which fly off in the forward region.

ATLAS (Fig. 1.8) and CMS are general purpose detectors. Their scientific programmes are similar and cover a wide range of topics, from Higgs physics,

[1]This topic, being the focus of the analysis presented in this dissertation, is explored in detail in Chapter 2

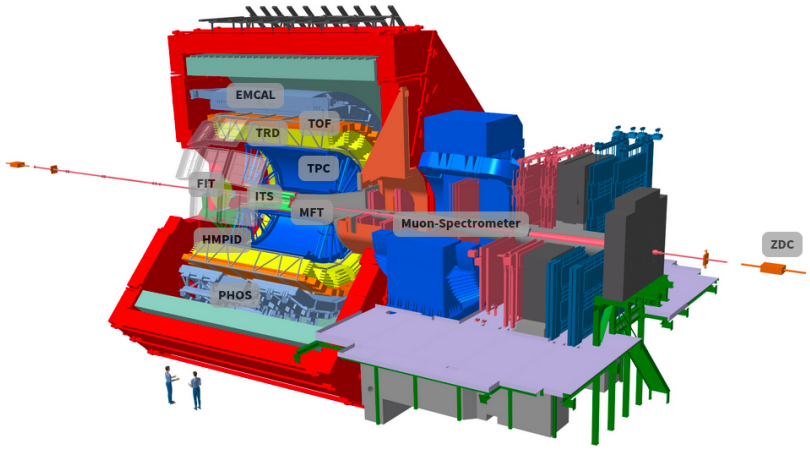


Figure 1.6: Schematic cutaway of the ALICE experiment at the CERN LHC. Image from [13].

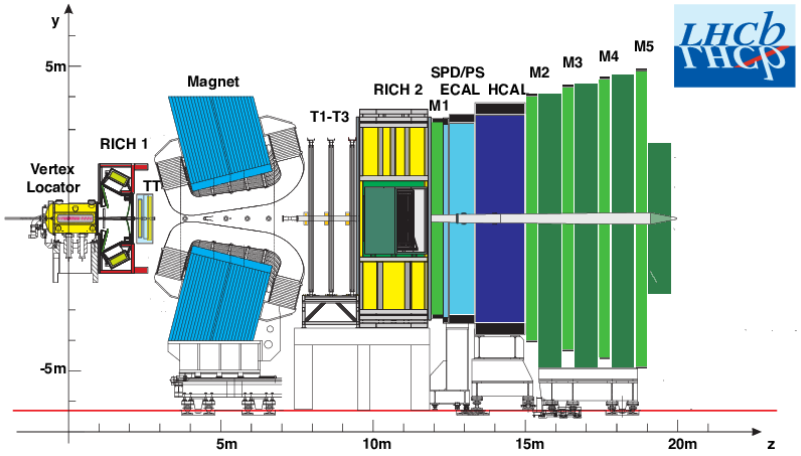


Figure 1.7: A longitudinal cross section of the LHCb detector, exhibiting all of its subdetectors. Image from [14].

BSM, B-physics, dark matter physics etc. The two experiments are famous for the joint discovery of the Higgs boson, a result published in 2012 [15] [16]. Section 1.2 offers an in-depth description of the CMS experiment.

Particle production in collisions

A bunch of protons passes through each point of the collider every 25 ns. During the collision of two oppositely moving bunches, only a very small fraction of the

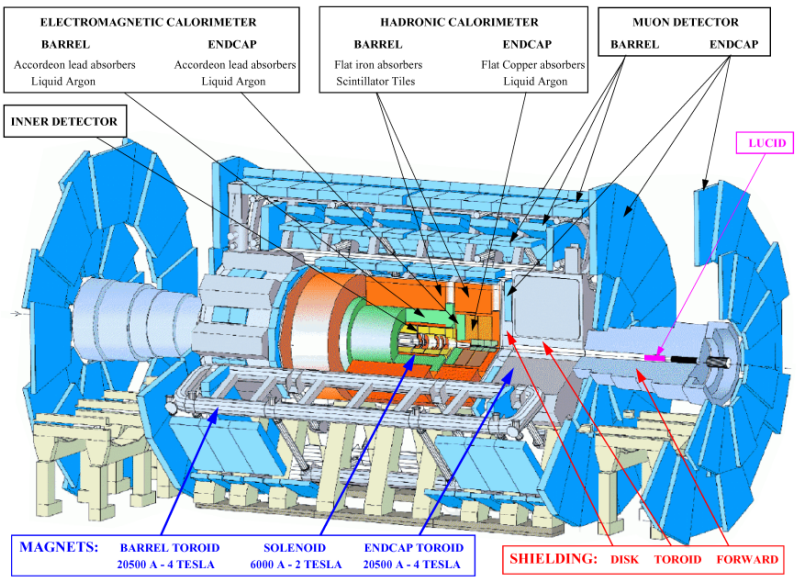


Figure 1.8: Schematic overview of the ATLAS experiment. Image from [17].

protons in the bunch will interact. In order to characterize the intensity and the number of interactions, two quantities are introduced: the LHC instantaneous luminosity (L) and the cross section (σ). They are related by the following expression:

$$L = \frac{1}{\sigma}\frac{dN}{dt} \tag{1.2}$$

where dN/dt is the number of events per unit time. Cross section is measured in units of surface (cm^2), or more commonly barn (b) ($1\ barn = 10^{-28}\ cm^2$). This implies units of number of events per $cm^{-2}s^{-1}$ for the instantaneous luminosity. Integrating the above equation gives the total number of events:

$$N = \int \frac{dN}{dt}dt = \sigma \int L dt = \sigma \mathcal{L}_{int} \tag{1.3}$$

where $\mathcal{L}_{int}$ is called the integrated luminosity.

For pp collision with $\sqrt{s} = 13$ TeV, the estimated total inelastic cross section is $\approx$ 70 mb. This leads to a number of $\approx$ 30 collisions per bunch crossing. Considering the fact that a bunch crossing occurs every 25 ns, the number of collision events at the CMS interaction point is $\approx 6 \times 10^8$ per second.

The analysis presented in this work is based on CMS physics data from the 2017 and 2018 data-taking periods. The measured luminosity provided by the LHC to CMS during these periods, as well as the integrated luminosity for the same periods, is shown in Figs. 1.9 and 1.10, respectively. Fig. 1.11 shows the mean number of interactions per bunch crossing during these same periods.

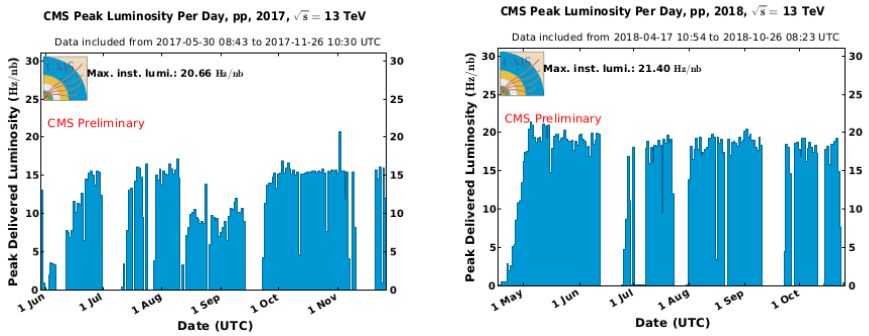


Figure 1.9: Instantaneous luminosity delivered from the LHC to CMS during the data-taking period in 2017 (left) and 2018 (right). Plots from [18].

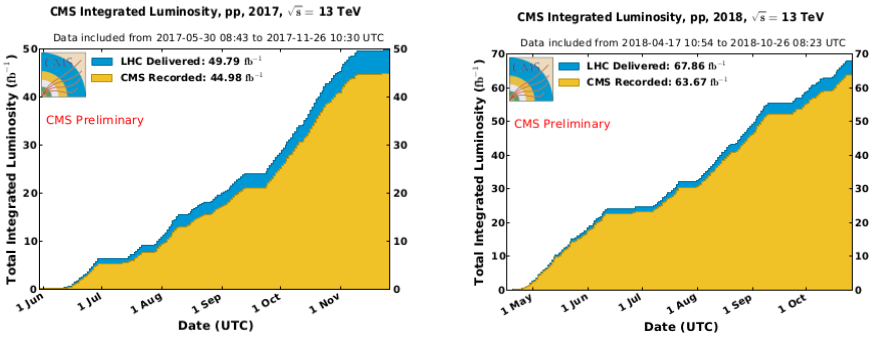


Figure 1.10: The evolution of the integrated luminosity delivered from the LHC to CMS during the data-taking period in 2017 (left) and 2018 (right). Plots from [18].

1.2 The CMS detector

The Compact Muon Solenoid (CMS) detector [19], shown in Fig. 1.12, is located at entry point 5 of the LHC accelerator tunnel, at a depth of about 100 meters below the surface, at the tunnel's side opposite the CERN main campus. It is a general-purpose detector. As such, it has a very diverse physics programme, as mentioned above, with studies being performed in areas ranging from Higgs boson physics, flavor physics, to more exotic topics like Beyond the Standard Model (BSM) physics and the search for dark matter.

The CMS detector has a cylindrical geometry, with a series of concentric layers of detectors, varying in function and purpose, stacked on top of each other. The detector is 21 m in length and 15 m in diameter. It consists of a central, so-called barrel part, where the detectors are placed parallel to the beam pipe and two enclosing parts, called endcaps where the detectors are

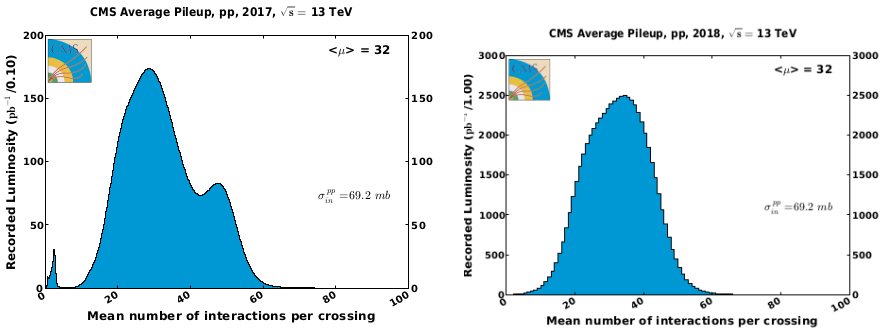


Figure 1.11: Mean number of interactions per bunch crossing for 2017 (left) and 2018 (right). Plots from [18].

placed perpendicular to the beam pipe. It weighs 14000 tons, making it the heaviest of the LHC experiments.

The CMS Collaboration includes more than 5000 scientists, engineers and students.

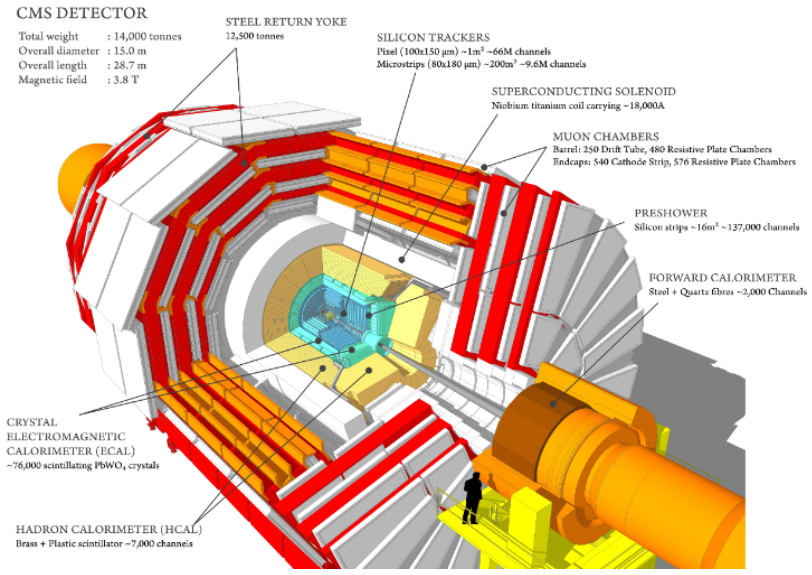


Figure 1.12: Schematic view of the CMS detector and its subsystems. Image from [16].

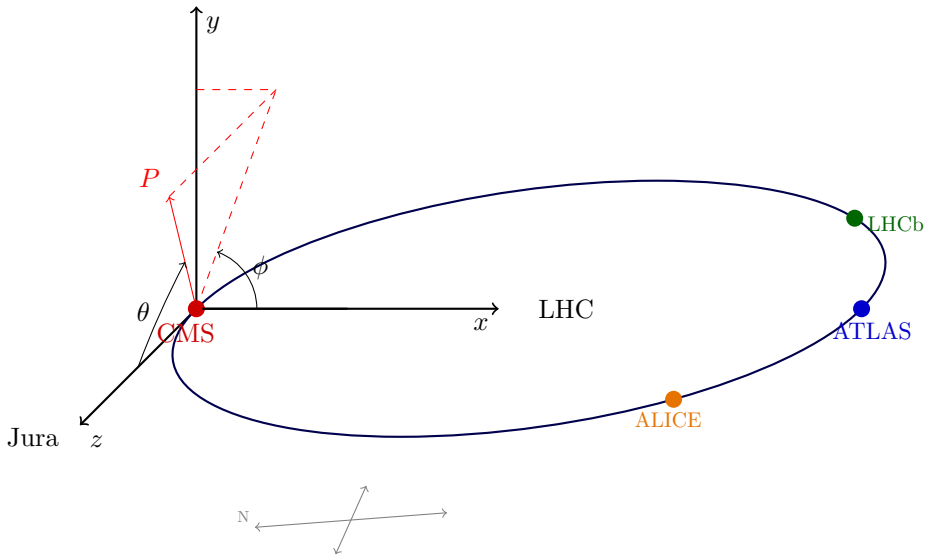


Figure 1.13: The CMS coordinate system. Figure from [20].

Coordinate system

The physics events registered with the CMS detector are described using a coordinate system centered at the nominal LHC collision point. The x-axis points inward toward the beam pipe, the y-axis points upward while the z-axis is tangential to the beam pipe and completes a right-handed coordinate system. The coordinate system is illustrated in Fig. 1.13.

Apart from Cartezian coordinates, cylindrical or spherical coordinates are often used as well, as this is also favored by the geometry of the detector itself. The x-y plane is referred to as the "transverse plane". The azimuthal angle ϕ is defined as the angle between the position vector's projection onto the transverse plane and the x-axis. The polar angle θ is defined as the angle between the position vector and the z-axis. In relativistic kinematics, the pseudorapidity is used interchangeably with the polar angle, with the two being connected by:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \tag{1.4}$$

The use of the pseudorapidity is favored by the fact that it is a Lorentz invariant quantity. The relationship between the two is shown in Fig. 1.14.

The solenoid magnet

A crucial component of the CMS experiment is the solenoid magnet [21]. It is the most powerful solenoid magnet ever made, constructed of a cylindrical coil of superconducting fibers that generates a magnetic field of ~ 4T, as shown in

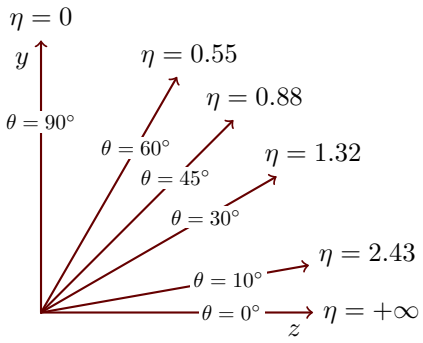


Figure 1.14: Polar angle and pseudorapidity

Fig. 1.15. In order to generate this field, a current of 18.5 kA is circulated in the coils. The magnet and the iron return yoke that confines the field lines to the volume of the detector, constitute the bulk of the detector's mass, weighing in at 12500 tons.

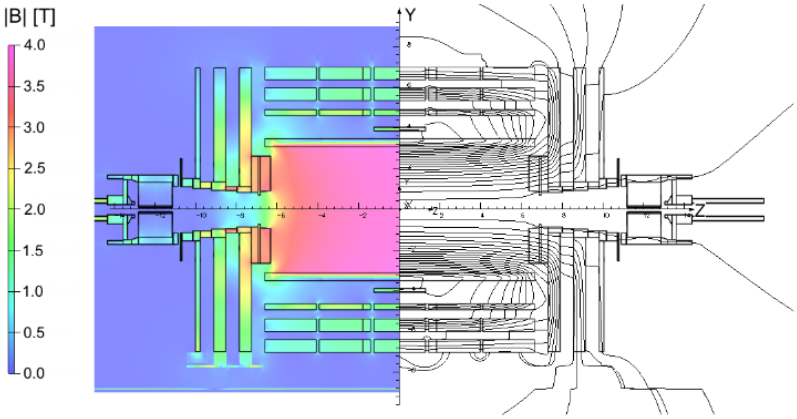


Figure 1.15: Field lines map of the CMS solenoid magnet. Image from [22].

The purpose of the solenoid magnet is to bend the trajectories of the charged particles produced at the interaction point and at secondary vertices inside the detector. The inside of the solenoid houses the silicon trackers and the calorimeters. On the outside of the magnet, the muon system is located.
Each subsystem is discussed separately in the following subsections.

1.2.1 Tracker

The CMS Tracker [23] is designed to precisely measure the trajectories of the particles coming from the primary vertex of the interaction point as well as to

make possible the reconstruction of secondary vertices. Measuring a particle's trajectory requires for the detector to interact with the particle at multiple points. A requirement for the tracker is to minimize the impact during each interaction, in order to avoid the modification of the particle's trajectory.

The overall dimensions of the tracker system are 5.8m in length and 2.5m in diameter. As the innermost layer in the detector, the tracker surrounds the interaction point. It is faced with the highest flux of particles coming from the interaction point and as a result it is faced with the highest amount of radiation damage in comparison to all other CMS subsystems. Therefore, the materials used in its construction have to be such that they are able to endure the hostile conditions and at the same time provide the required performance, in terms of granularity, response time etc. These constraints led to the decision to construct the entire tracker out of silicon.

The CMS tracker consists of two types of silicon-based technologies: pixel detectors and silicon microstrip detectors. A cross-section of the layout of these detectors, in both the barrel and endcap regions, is shown in Fig. 1.16.

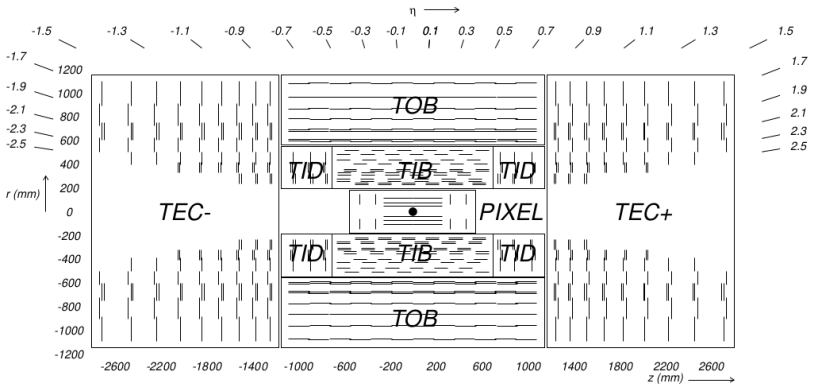


Figure 1.16: Schematic cross-section of the tracker layout. Image from [24].

The pixel detector [2] is the first layer of the tracker. It extends radially from 4.4 cm to 10.2 cm with respect to the beam pipe, with two additional disks of pixel modules on each side (Fig. 1.17). A total of 66 million pixels provide the fine granularity required to register the high flux of charged particles passing through this subsystem. Three space points are registered with high precision for each passing charged particle, making possible the reconstruction of their trajectories.

The silicon microstrip detector occupies the region after the pixel tracker, as seen from the interaction point. Radially, it extends from 20 cm to 116 cm and is composed of three different subsystems.

First, the Tracker Inner Barrel and Disks (TIB/TID) consists of 4 barrel

[2] During Long Shutdown 2 (LS2) of the LHC, the pixel detector was extracted and refurbished in preparation for Run-III [25].

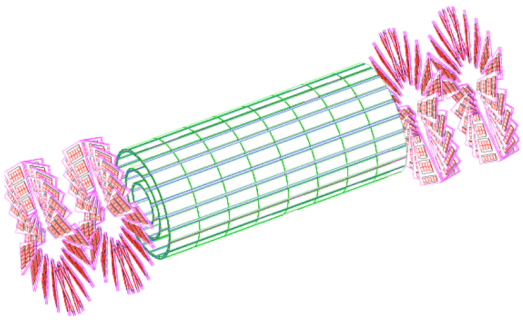


Figure 1.17: 3D layout of the silicon pixel tracker

layers and 3 disks at each end of 320 μm thick silicon micro-strip sensors, placed parallel to the beam pipe in the barrel and radially in the disks. It provides 4 position measurements in $r - \phi$, with a resolution of ~ 10 μm.

The Tracker Outer Barrel (TOB) surrounds the TIB/TID. It has 6 layers of $500\mu m$ thick micro-strip sensors. TOB provides an additional 6 $r - \phi$ measurements with resolutions of $53\mu m$ and $35\mu m$, for the two different strips pitches, used in different layers.

Beyond the distance of ± 118 cm in z, to where TOB extends, the Tracker EndCaps (TEC+ and TEC-) are positioned. Each TEC consists of 9 disks containing up to 7 rings of micro-strip detectors. Thus, this subsystem guarantees an additional 9 position measurements in ϕ, per trajectory.

1.2.2 Electromagnetic Calorimeter

The CMS Electromagnetic Calorimeter (ECAL) is located between the tracker and the hardon calorimeter (HCAL). Its purpose is to contain all the electrons, positrons, and photons passing through it, and measure their energies. Its layout is illustrated in Fig. 1.18.

ECAL consists of $80,000$ lead tungstate crystals ($PbWO_4$) [27] that act as scintillators. The lead tungstate crystals have a high density ($8.28 g/cm^3$) and a short radiation length (0.89 cm) making them appropriate for work in the hostile, high-radiation environment of the LHC while also providing the needed properties of a calorimeter with fine granularity and a compact size. An additional positive property of the $PbWO_4$ crystals is their scintillation decay time, which is of the same order of magnitude as the LHC bunch crossing time of 25 ns. Approximately 80% of the light produced in the ECAL during one bunch crossing will be emitted before the next bunch crossing takes place. The produced light is detected by two types of photodetectors: avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps. An image of the crystal and the photodetectors is shown in Fig. 1.19. **The barrel part of the ECAL (EB)** covers the pseudorapidity range $|\eta| < 1.479$. EB contains a total of $61,200$ crystals, for a granularity of 360-fold in ϕ and (2×85)-fold in η. The

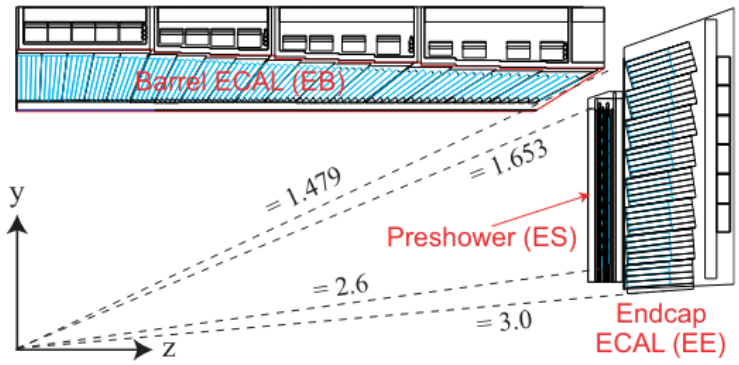


Figure 1.18: Layout of one of the quadrants of the CMS ECAL. Image from [26].

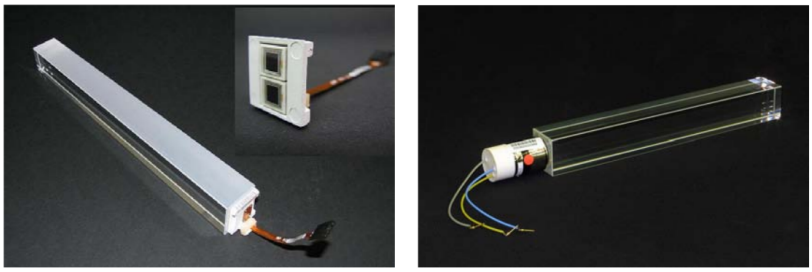


Figure 1.19: $PbWO_4$ crystals and the corresponding photodetectors: APD for the barrel (left) and VPT for the endcaps (right)

crystals in EB have a tapered shape, with the shape varying with position in η. They are combined in groups called submodules, which are further combined in modules (Fig. 1.20) and supermodules.

The ECAL endcaps (EE) cover the pseudorapidity range $1.479 < |\eta| < 3.0$. The crystals in EE, in contrast to the ones on EB, are identical in shape and are combined in units of 5×5 crystals, called supercrystals.

In order for the ECAL to function properly and to exhibit the nominal energy resolution, it has to be maintained at a constant temperature. This is due to the fact that both the number of photons emmited by the crystals and the amplification produced by the photodetectors are sensitive to temperature variations. Measurements performed under test beam conditions show an overall variation of $(-3.8 \pm 0.4)\%\ ^\circ C^{-1}$. A water flow cooling system maintains the temperature at the ECAL nominal value of $18 \pm 0.05^\circ C$.

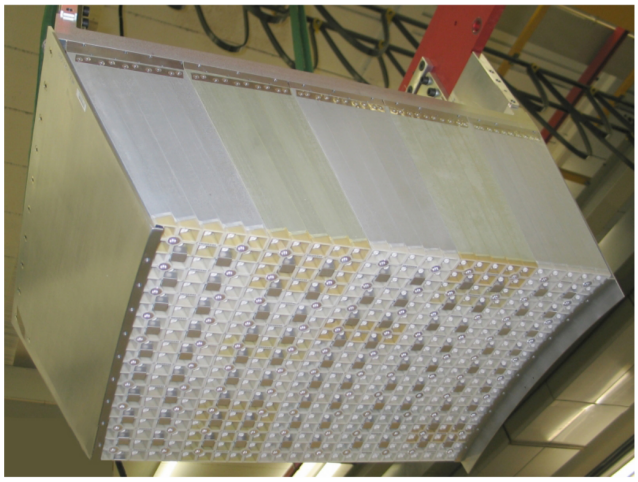


Figure 1.20: Image of an ECAL barrel module

1.2.3 Hadron Calorimeter

The CMS Hadron Calorimeter (HCAL) [28] surrounds the ECAL system (Fig. 1.21) and acts as an absorber for all the hadrons produced in the primary or secondary vertices in the volume inside of it. HCAL is composed of a barrel and two end-

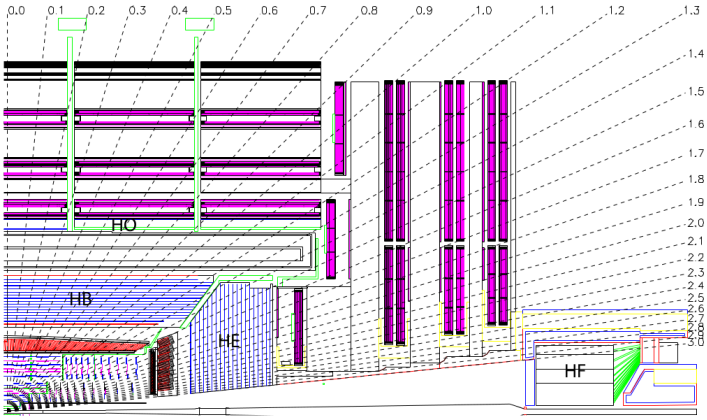


Figure 1.21: The HCAL location in the CMS solenoid cavity

cap parts on both sides of it, plus a forward calorimeter and an outer calorimeter to complete the setup.

The barrel part of the HCAL (HB) covers the pseudorapidity region $|\eta| < 1.3$ and is composed of two sections or half-barrels. Radially, it extends from the outer layer of the ECAL at $R = 1.77m$ to the inner layer of the mag-

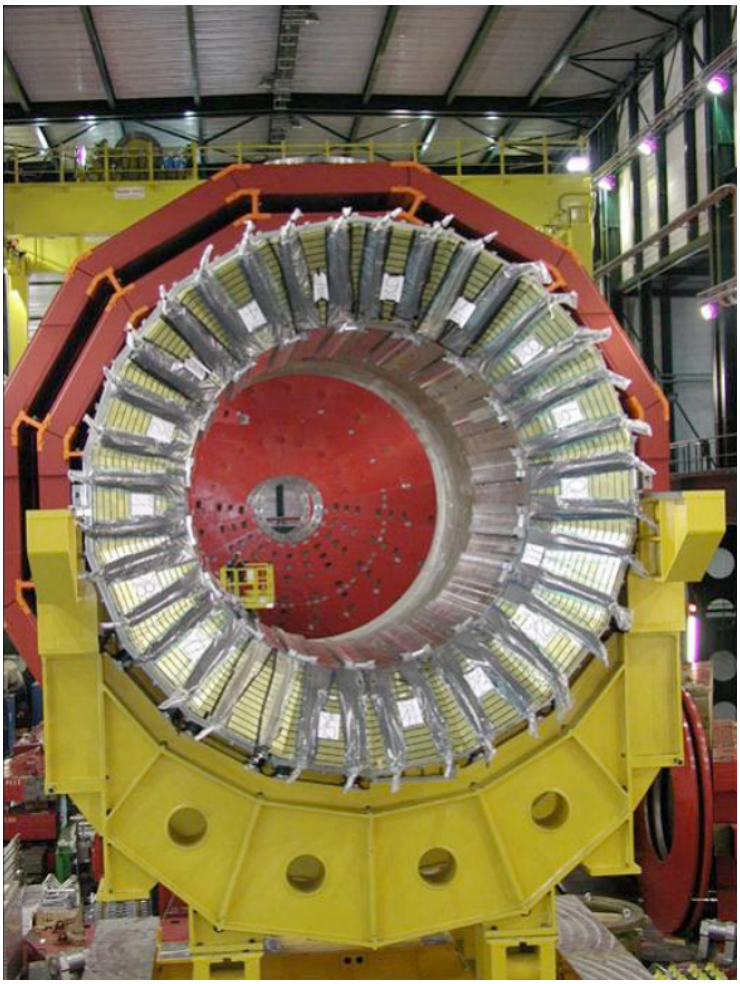


Figure 1.22: Image of one of the HCAL half-barrels

net coil at $R = 2.95\ m$. In the azimuthal plane, each half-barrel is made of 18 individual segments (36 in total for HB), called wedges. Fig. 1.22 shows an image of one of the HCAL half-barrels. The absorber segments of HB are composed of brass. This choice of absorber material is motivated by its relatively short interaction length and non-magnetic properties. The absorber layers are interleaved with plastic scintillator tiles, representing the active medium of the detector. The signal produced by the tiles is collected by wavelength-shifting fibres, placed in machined grooves inside the scintillator, and is measured by hybrid photodiodes [29].

The hadron calorimeter endcaps (HE) cover the pseudorapidity ranges $1.3 < |\eta| < 3$. The technology of the absorber and the sensitive medium is very

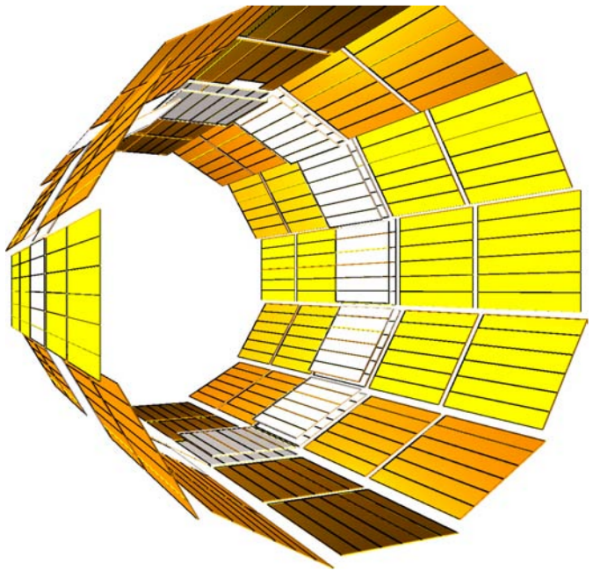


Figure 1.23: Layout of the outer calorimeter

similar to the ones described for the HB.
The outer calorimeter (HO), is placed on the outer surface of the solenoid magnet. The necessity for such a component in the HCAL arises from the fact that in the central pseudorapidity region, the hadron showers cannot be entirely contained by the inner calorimeter. So HO serves as a tail catcher of any hadron shower that extends beyond the solenoid magnet. Its layout is shown in Fig. 1.23. HO consists only of scintillators and fibres and no absorber material.
The forward calorimeter (HF) is located at a distance of $11.2m$ from the interaction point, both on the + and - sides of the CMS detector. HF is faced with extreme particle fluxes as it covers the very high pseudorapidity range. The design choices made during the construction of this component of HCAL were guided mostly by the need to withstand very high absorbed doses during operation. As a result, quartz fibres were chosen as the sensitive medium, due to their satisfactory radiation hardness.

The principle of operation is that of a Cherenkov detector, where charged particles above the Cherenkov threshold, generate light which is collected by fibres and transferred to the readout. The sensitive medium and the fibres are placed in grooves inside a steel absorber. Fig. 1.24 displays a cross-section view of HF and its components.

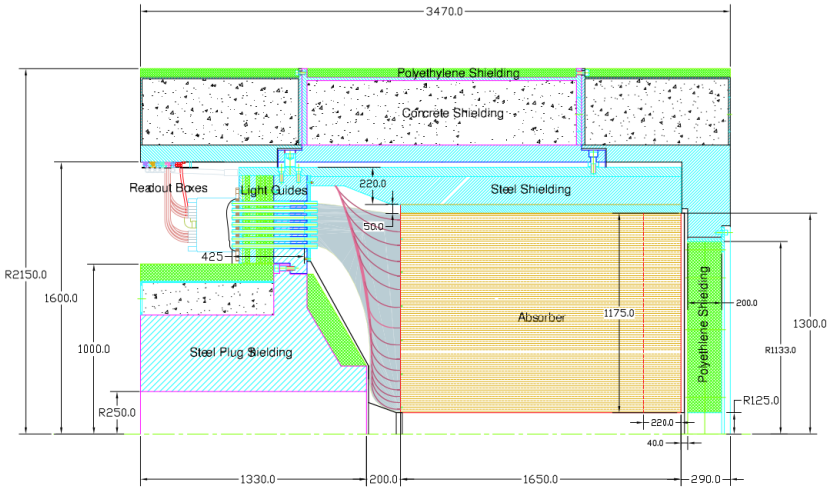


Figure 1.24: Cross-section schematic of the forward calorimeter

1.2.4 Muon system

The muon system is located entirely outside of the solenoid magnet (Fig. 1.25).

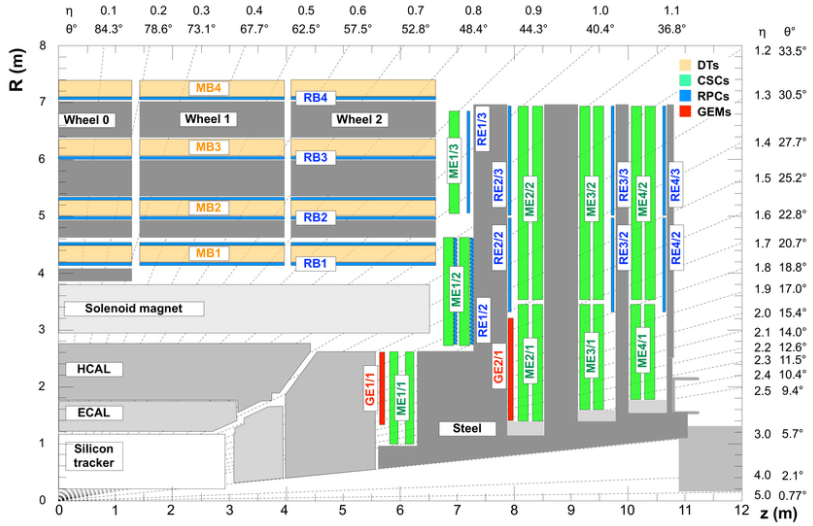


Figure 1.25: Illustration of one quadrant of the muon system of the CMS detector. Figure taken from [30].

As the name suggests, it is designed for the purpose of detecting muons, with three main functions: muon identification, momentum measurement and

triggering. It consists of a cylindrical barrel part and two planar endcap parts, housing the following subsystems:

- Drift Tubes (DT): located in the barrel part
- Cathode Strip Chambers (CSC): located in the endcaps
- Resistive Plate Chambers (RPC): located both in the barrel and endcap parts of the detector, alongside the DT and CSC subsystems, respectively
- Gas-Electron Multipliers (GEM): represent the newest detector technology in the muon system, gradually being integrated in the endcaps since LS2.

Drift tube (DT) chambers cover the $|\eta| < 1.2$ pseudorapidity region and are organized in four barrel stations. Each of the five barrel wheels contains 50 chambers, 4 for each of the 12 ϕ sectors plus additional chambers in sectors 4 and 10 (Fig. 1.26).

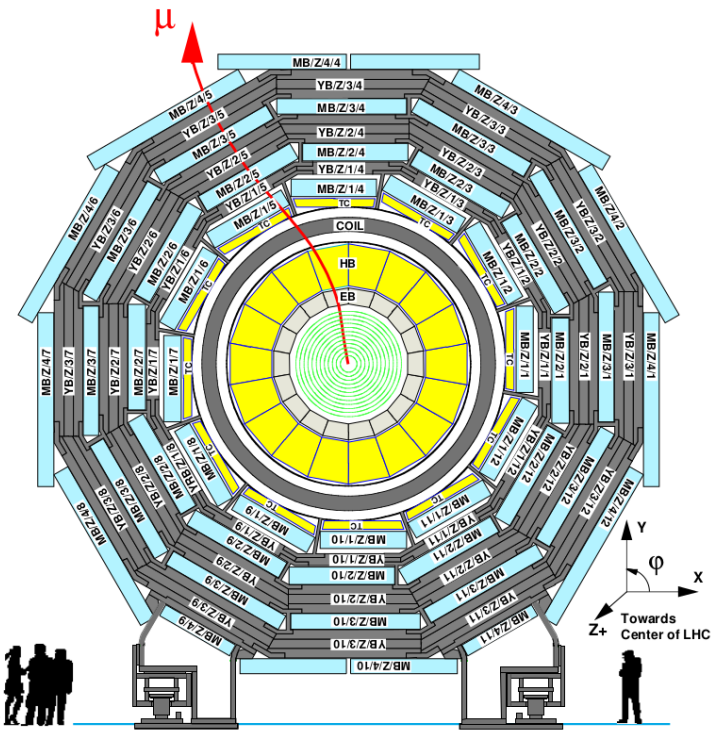


Figure 1.26: Transversal view of the DT position in one of the CMS barrel wheels

DT are gaseous detectors, as are the rest of the detectors in the muon system and they are operated with a Ar/CO_2 gas mixture.

Each DT chamber is internally constructed of separate cells (Fig. 1.27) organized in 2 or 3 superlayers (SL). This modular construction localizes damage and allows neighboring cells to continue functioning properly if a given cell sustains some kind of malfunction (e.g. broken anode wire). There are SL with orientation of wires parallel to the beam axis, providing position measurement in the $r - \phi$ plane and SL where the orientation is perpendicular to the beam axis, providing measurement position in the z-direction also.

Each cell contains field electrodes, made of aluminium tape and a gold-plated stainless steel wire, $50\mu m$ in diameter, acting as the anode and which is extended along the middle of the cell. The cathode strips are placed at the border of two neighbouring cells. The applied voltages on the electrodes are: +3600 V for anode wires, +1800 V for field electrodes and −1200 V for cathode strips. The DT chambers offer a time resolution of a few nanoseconds and spatial resolution of $\sim 150\mu m$.

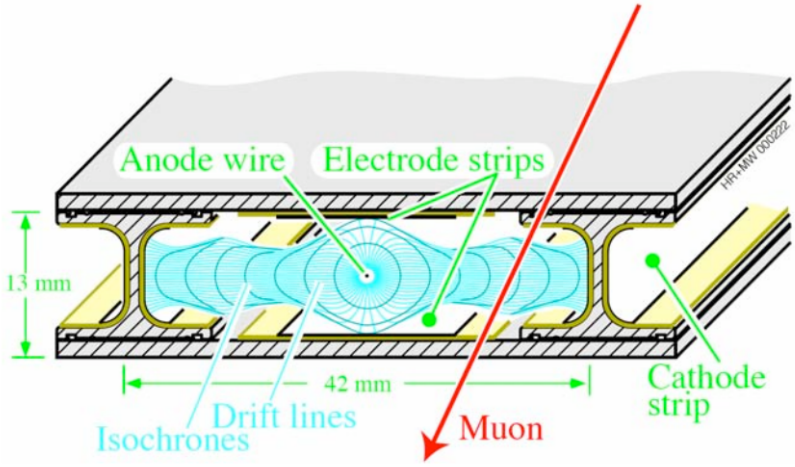


Figure 1.27: A cross-section of a drift cell inside a DT chamber

Cathode Strip Chambers (CSC) are located in the CMS endcaps. A total of 468 CSC chambers, of variying sizes, constitute the CSC subsystem. CSC are multiwire proportional chambers comprised of 6 anode wire planes interleaved with 7 cathode panels (Fig. 1.28). The wires are positioned azimuthally and they determine the radial component of the coordinates of a passing muon's trajectory. On the other hand, the ϕ coordinate is obtained by interpolating charges induced on the strips, running perpendicular to the wires, as is illustrated in Fig. 1.29.

Gas Electron Multipliers (GEM) detectors [31] were introduced as an idea by Fabio Sauli [31] in 1996. GEM consist of thin polymere foil electrodes, metal-coated on both sides, with a high-density of holes ($50 - 100$ mm^{-2}), as illustrated in Fig. 1.31. The pierced electrode is located between a drift and a

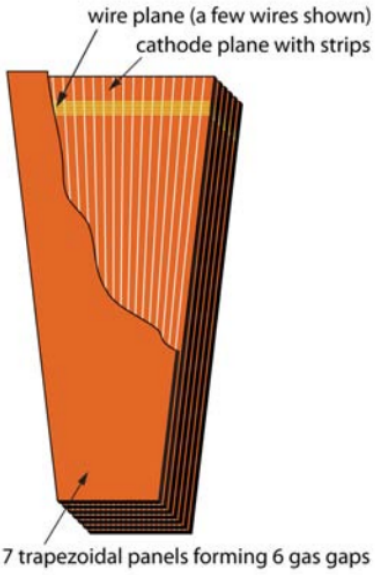


Figure 1.28: Layout of a CSC chamber

charge-collecting electrode. The HV potentials applied on these electrodes lead to the creation of a field with near equipotential field lines inside the microscopic holes (Fig 1.32). A passing particle, such as a muon, interacts with the gas in the working volume of the GEM detector, causing an ionization. The produced electron, drifts towards the holes, where the strong field causes it to ionize other atoms of the working gas, thus initiating an electron avalanche. A large portion of the electrons produced in this amplification region inside the hole leave the electrode and are collected by a layer of anodes, determining both the x and y coordinates of the particle's trajectory.

The first GEM chambers were installed on the CMS detector in 2019. A total of 36 superchambers were integrated in the first disks of both the positive and negative endcaps (Fig. 1.33). At the time of writing, the GEM subsystem has installed additional chambers in the CMS endcaps during the 2023/2024 Year-End Technical-Stop (YETS).

The main motivations for integrating the GEM detectors in the CMS muon system is to increase the muon system capabilities and the need to face the challenges of HL-LHC [33]. The GE1/1 GEM chambers provide redundancy in the $|\eta| > 1.6$ pseudorapidity region, which up to LS2 was covered only by the CSC chambers. In addition, with the upcoming installations of the ME0 GEM chambers, the muon acceptance will be extended into the pseudorapidity range $2.4 < |\eta| < 2.8$ which is beyond the current capabilities of the muon system.

The RPC subsystem, central to one of the works presented in this dissertation, is covered in detail in the following section, dedicated entirely to it.

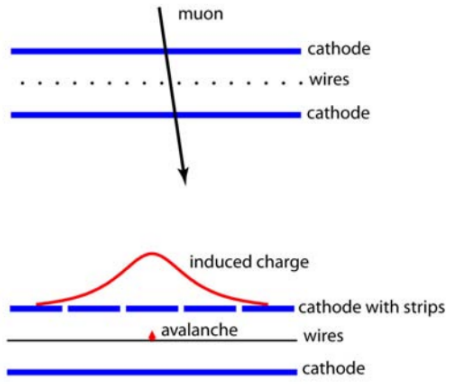


Figure 1.29: Principle of operation of CSC

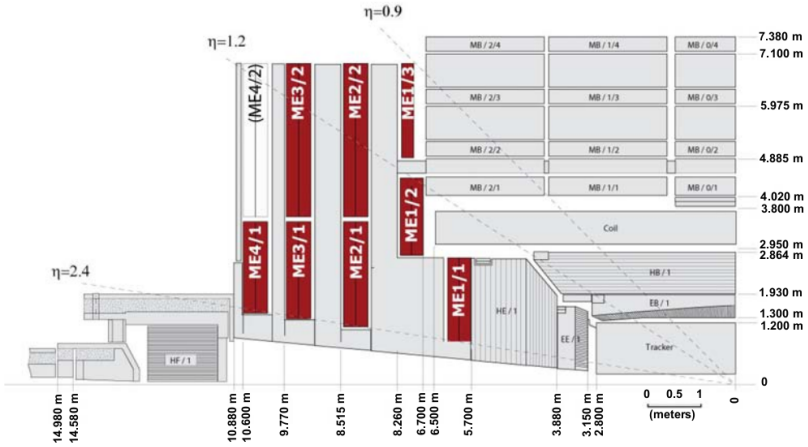


Figure 1.30: A quadrant of the CMS endcap, showing the position of the CSC chambers (highlighted in red)

1.3 The RPC subsystem

The RPC subsystem [34] of the muon system of CMS extends both in the barrel and endcap regions of the detector. It is comprised of a total of 1054 chambers, with 480 chambers in the barrel and 574 in the endcaps (Fig. 1.34).
It covers the pseudorapidity region $|\eta| < 2.1$. In the barrel, RPC are distributed in 4 concentric cylindrical stations. The first two stations contain two layers of RPCs, placed on both sides of the DT chambers, while the third and four stations contain a single RPC layer, placed on the inner side of the DT chambers. This allows for the track reconstruction to always be performed on at least 4

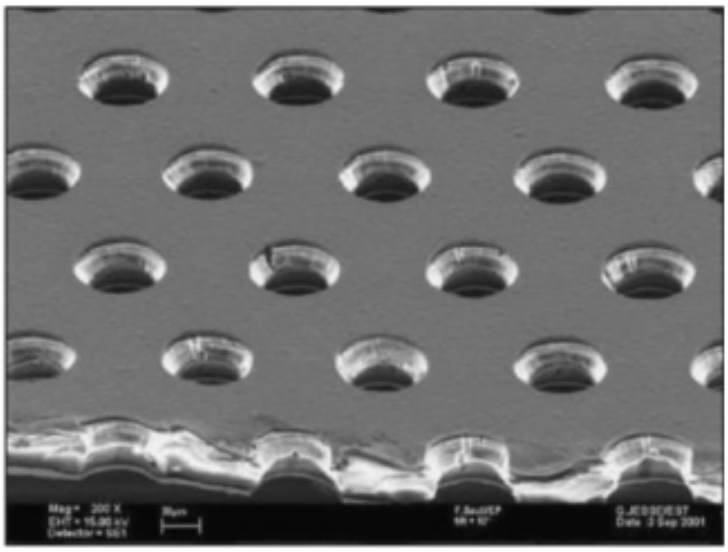


Figure 1.31: A microscopic view of the holes pierced through a GEM electrode. Image from [32].

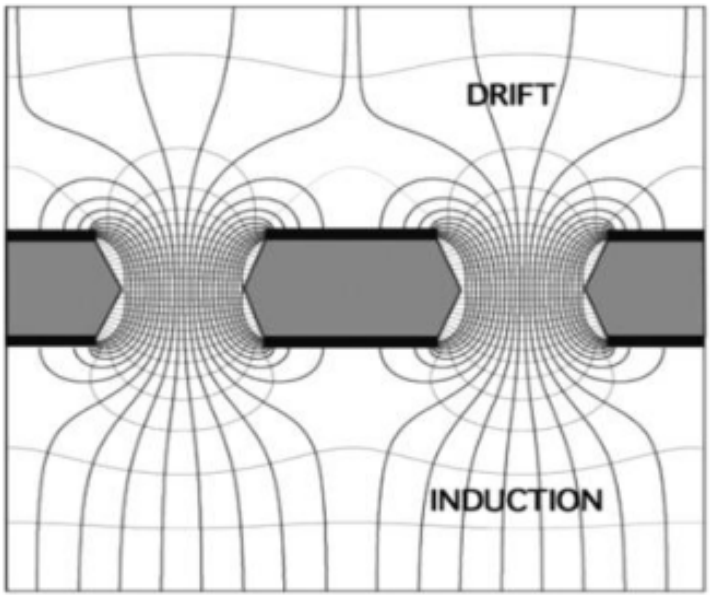


Figure 1.32: The field lines that are established inside the GEM electrode holes. Image from [32].

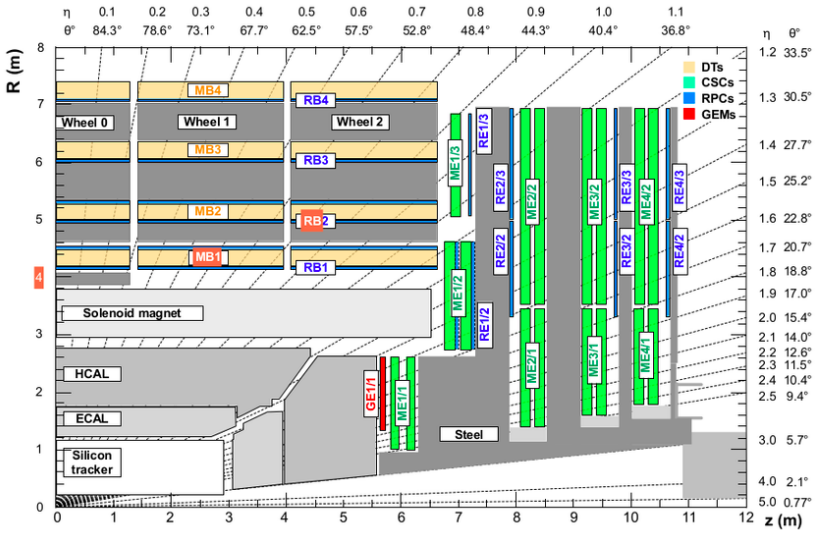


Figure 1.33: A cross-section of a CMS quadrant, after the installation of the GE1/1 GEM superchambers. Image from [30].

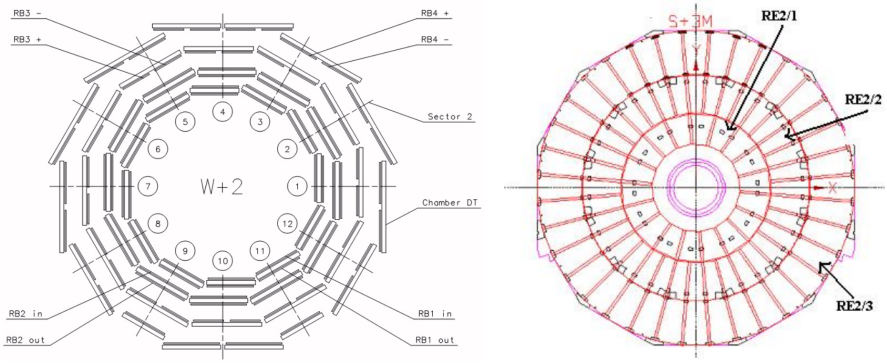


Figure 1.34: Layout of the CMS RPC chambers in the barrel part of the detector (left) and the endcaps (right)

muon hits, even for low p_T muons which are stopped before they reach the third station.

The RPC system can independently serve as a dedicated muon trigger. This trigger performs the following functions: identification of muon tracks, determination of a track's momentum and the association of this track with the corresponding bunch crossing (BX) .

1.3.1 Resistive Plate Chambers

Resistive Plate Chambers (RPC) [35] are gaseous parallel-plate detectors that are mechanically agile and robust and make possible the coverage of large detection areas for a reasonable financial cost, making them a perfect alternative for employment in the large LHC experiments, as is CMS. They are constructed using high-resistivity bakelite electrodes with a volume resistivity ranging from $10^9 - 10^{10}$ Ω cm. The electrodes enclose a thin gas gap 2 mm thick (Fig 1.35). Spacers are placed on on the nodes of a square grid between the two electrodes, in order to maintain the uniformity of the gap thickness across the whole surface. The thin gap allows for the achievement of a very good time resolution,

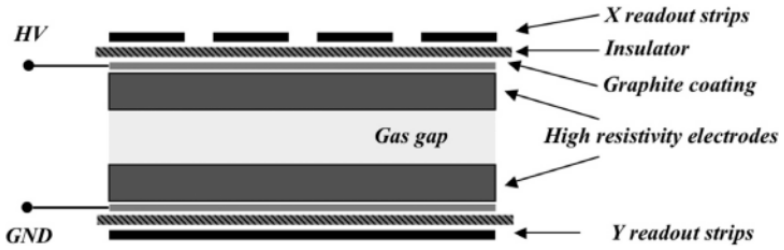


Figure 1.35: Cross section schematic of a standard single-gap RPC. Image from [32].

of the order of ∼ns. This is important, as the chambers are used in the CMS muon trigger and they must be able to associate a particle's track with a given BX, the latter occurring every 25ns.

The RPC electrodes are coated with a thin graphite layer on their outer surfaces, where the High-Voltage (HV) is applied. The low surface resistivity of the graphite minimizes any potential drops due to the signal currents by redistributing the surface charge. The produced signal is detected on a set of copper detecting strips, insulated from the graphite coating by an insulating PET film.

The working principle of an RPC chamber is the following: a passing high energy particle interacts with the working gas, creating a primary ionization cluster. Under the effect of the strong electric field, these primary electrons are accelerated to energies high enough to cause secondary ionization. The process continues exponentially giving rise to an avalanche of electrons [36], as is visualized in Fig. 1.36. The produced charge induces a signal on the detecting strips, constituting the signal of an RPC chamber. The amount of induced charge can, to first approximation, be estimated using the following expression [36]:

$$< q_e >= \frac{k}{\eta d} < Q_e(d) >= q_{el} n_0 \frac{k}{\eta d} \frac{\lambda}{\eta + \lambda} e^{\eta d} \tag{1.5}$$

where $< Q_e(d) >$ is the mean charge in an avalanche and d is the gap width. η is the effective ionization coefficient and is calculated as $\eta = \alpha - \beta$, where α

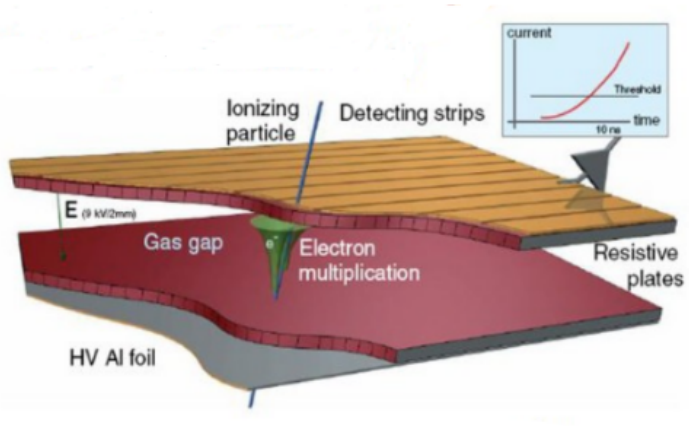


Figure 1.36: Passage of a highly energetic particle through an RPC chamber and the subsequent avalanche formation

is the Townsend coefficient [37], expressing the number of ionizing encounters of an electron in the gas per unit length while β is the attachment coefficient and expresses the number of attaching encounters per unit length. λ is the gas mixture's cluster density[3] and k is a constant, calculated as:

$$k = \frac{\epsilon_r d}{s} \frac{\epsilon_r d}{s+2} \tag{1.6}$$

This coefficient depends on the material parameters: ϵ_r, which is the relative dielectric constant of the electrode and s which is the electrode thickness. Fig. 1.37 shows the average fast charge induced on the detecting strips as a function of η for two different values of λ. It should be noted that in equation 1.5 there is a non-trivial dependence on the gap width. Considering the fact that the gaps used to construct the RPC chambers at CMS are quite thin, the induced charge is not very large. This introduces the need for sensitive electronics at the front-end boards. The signal is amplified in order to overcome the electronics noise threshold.

The operating voltage of an RPC gap determines the working mode of the chamber. There are two modes of operation: avalanche and streamer mode. Avalanche mode occurs at lower operating voltages [36]. A condition for the chamber to be in avalanche mode in terms of the quantities introduced in 1.5 is:

$$\eta d < 20$$

With increasing HV a regime called saturated avalanche regime is reached, when the exponential growth of the number of electrons in the avalanche exceeds $10^6 - 10^7$ and the generated space charge reduces the field produced by the

[3]It expresses the number of primary ionization clusters per unit length produced in the gas and is discussed further in the subsection about the CMS RPC gas mixture

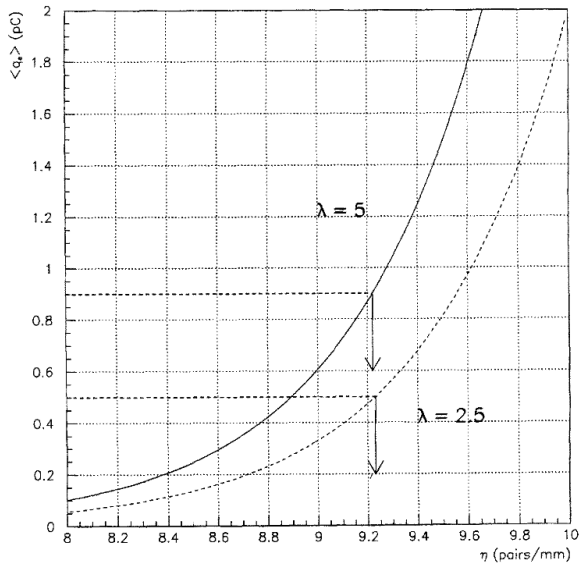


Figure 1.37: Average fast charge as a function of the effective ionization coefficient, for two different values of the cluster density λ. Plot from [32].

electrodes considerably. A further increase of the HV leads to a transition to a streamer regime, where the signals obtained are much larger but which leads to a loss of rate capability and time resolution, as the produced streamer signal needs a longer time to be read and the detector needs a longer time to recover. At CMS, RPC are operated in a saturated avalanche mode and the occurrence of streamers is avoided by fine-tuning the parameters of the chamber.

1.3.2 Gas mixture

The working volume of RPC chambers in the CMS detector is a 3-component non-flammable gas mixture [38] composed of: 96.2 % tetrafluoroethane[4] ($C_2H_2F_4$), 3.5 % isobutane (i-C_4H_{10}) and 0.3 % sulfur esafluoride (SF_6). This gas mixture offers all the desired properties.

The gas cluster density λ, introduced in equation 1.5, is an important quantity that affects the amount of primary charge produced during the passage of an ionizing particle. This reflects in the total avalanche charge and the charge induced on the strips, thus directly affecting the efficiency. $C_2H_2F_4$, which is the predominant component of the RPC gas mixture has a relatively high λ of ~ 5 clusters/mm, allowing RPC to achieve high efficiency whilst avoiding very high working point, which would increase the streamer probability.

SF_6, an electro-negative gas, is a strong quencher, which is a property needed in order to suppress the occurence of streamers. The addition of this gas in small

[4]Also known as freon, with industrial nomenclature R134a

quantities achieves the desired effect of separating the onset of the avalanche regime from the streamer regime, thus making the occurrence of streamers at HV working point far less probable.

A relative humidity of 45% is maintained in the gas mixture by adding water vapour. The humidity also acts as a protective element for the bakelite resistivity.

The gas supply is performed by a closed-loop circulation system, as is illustrated in Fig. 1.38. The main functions of the gas supply system is to mix the

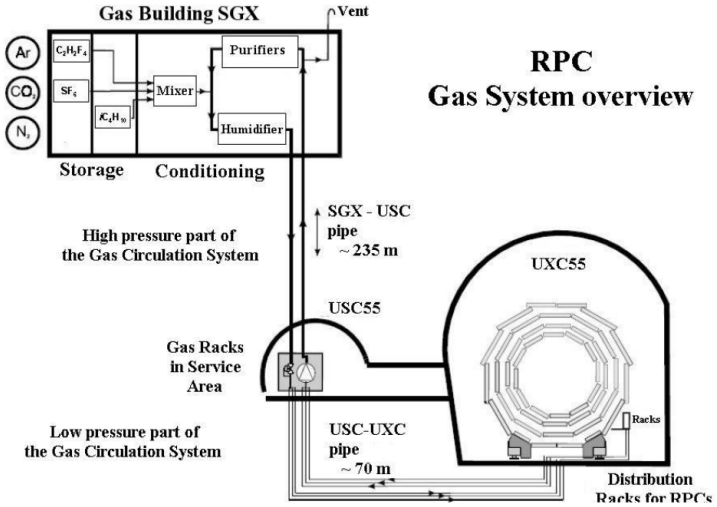


Figure 1.38: Overview of the CMS RPC closed-loop gas circulation system. Image from [19]

gas components in the appropriate ratios and to circulate the obtained mixture in order to supply every chamber in the system. The modules comprising the system are: the primary gas supply, mixer and closed-loop circulation system, gas distributor to the chambers, purifier, pump and gas-analysis station. The mixer prepares the mixture and constantly monitores the flow of component gases. Special attention is needed to ensure that the ratio of isobutane doesn't exceed the flamability limit. The presence of impurities in the gas mixture is also constantly monitored. If the concentration of these impurities is high enough they can influence the detector performance. The role of the purifier is to remove these impurities in order to avoid the long-term deterioration of the performance.

1.3.3 RPC properties

The performance of an RPC chamber has to fulfill several requirements: good time resolution, high rate capability, low streamer probability, high detection efficiency, low cluster size etc. The design of the RPC chambers of CMS was based

on a series of tradeoffs with the intention of finetuning the detector properties such that all the requirements would be satisfied.

The time resolution of a chamber mainly depends on the operating voltage and the gap width. The narrower a gap, the better the time resolution. However, narrowing the gap reduces the size of the working volume of the detector, thus reducing the detection efficiency. At CMS, the solution of using $2mm$ gaps in a double-gap configuration has been implemented. It satisfies the requirements with regards to both resolution (Fig. 1.39) and efficiency. An additional benefit

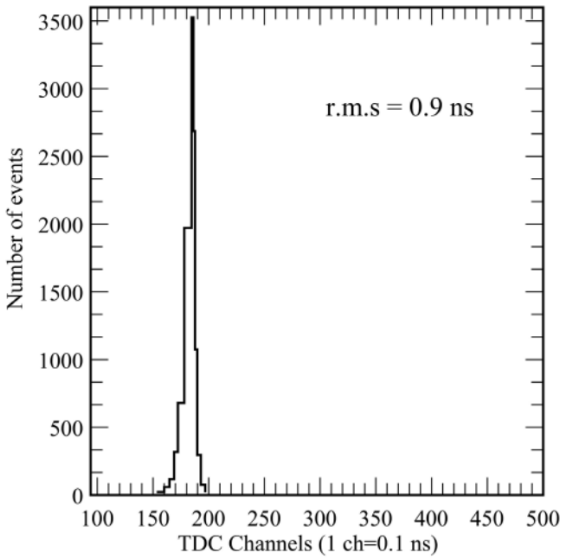


Figure 1.39: Time resolution histogram for an RPC with $2mm$ gaps and low resistivity electrodes. Histogram from [32].

from utilizing thin gaps is that they can be operated at a lower gas gain and therefore at a lower working HV, nevertheless yielding the same performance.

In the hostile environment of the LHC, an RPC chamber must be able to withstand hit rates of the order of 10^3 Hz/cm^2. The rate capability of an RPC chamber is strongly dependent on the resistivity ρ of the bakelite electrodes. This dependence manifests through the voltage drop across the electrodes upon the formation of an electron avalanche. A simple expression giving the voltage drop as a function of the resistivity is:

$$V_d = 2 < Q_e > rs\rho \tag{1.7}$$

where $< Q_e >$ is the mean avalanche charge, r is the rate, ρ is the bulk resistivity and s is the electrode thickness.

It is visible from this expression that the voltage drop is directly proportional to the resistivity. Therefore, using lower resistivity electrodes leads to lower voltage drop and thus the voltage drop can be brought to an acceptable value.

However, the resistivity value is constrained from below by the requirement of maintaining a low cluster size (Fig. 1.40), which translates in a better space resolution. In order to satisfy both these conditions, a bulk resistivity of $8 \times$

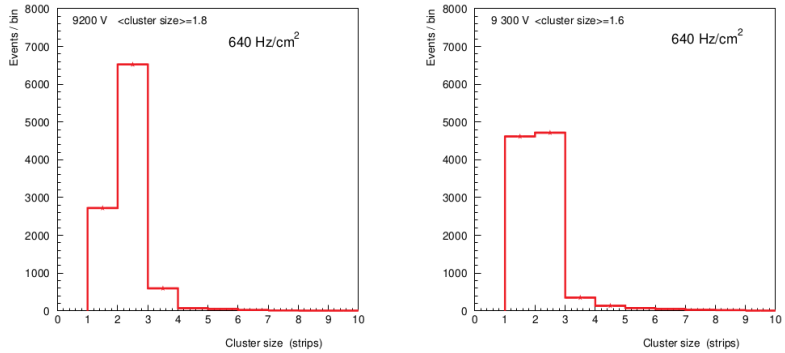


Figure 1.40: Cluster size as a function of working point (HV) for RPC with low resistivity electrodes (left) and higher resistivity electrodes (right)

$10^{10} \Omega m$ is used in the CMS RPC chambers.

An additional requirement on the RPC chambers is the ability to withstand operation in the high-radiation environment of the LHC for extended periods of time. The initially projected lifetime of $\sim$ 10 years for this system has already passed at the time of writing of this dissertation and the system is still operating successfully. This has been achieved by designing radiation resistant components and electronics and regularly maintaining the chambers during the Year-End Technical Stops (YETS) and the Long Shutdowns (LS) between LHC runs.

Periodical aging and longevity studies [39] have been performed on the RPC chambers since the beginning of operation. They have shown no performance deterioration due to prolonged exposition to high radiation doses.

1.3.4 RPC currents monitoring

The most common malfunction of an RPC chamber during operation is a HV trip. It occurs when the dark current[5] that the chamber draws increases disproportionately to the applied HV working point. An example of a chamber going through a HV trip is shown in Fig. 1.41.

Such an event produces an error which is propagated upwards in the CMS detector's hierarchy. This causes the detector to stop the data acquisition process until the problem is addressed. Most of the time, it is resolved quickly and the data acquisition resumes. However, preventing or reducing such events would save valuable time and increase the amount of data collected with the detector.

[5]The current present in the HV channel line supplying the chamber. Not to be confused with the useful signal of the chamber

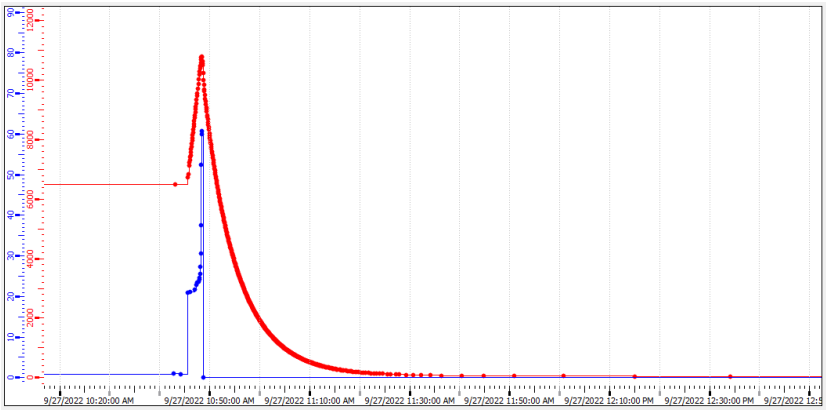


Figure 1.41: The plot shows the evolution of the dark current (blue) and HV (red) of an RPC chamber (W0_S09_RB2in) in the CMS barrel during a HV trip. The immediate increase, once a threshold is reached, triggers a safety mechanism that gradually powers the chamber down, avoiding any potential damage. Plot taken from the RPC Detector Control System (DCS) [40].

Chapter 2

Theoretical overview

2.1 The Standard Model of particle physics

The Standard Model (SM) [41] is the best experimentally tested theoretical model describing three of the four fundamental interactions of nature: the electromagnetic, the weak and the strong interactions. SM is a gauge theory, based on the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. It models interactions of fermions via the exchange of spin-1 gauge bosons.

2.1.1 Particle content of the SM

The particle content of the SM is comprised of 3 generations of spin-$\frac{1}{2}$ particles, called fermions, namely leptons and quarks and spin-1 bosons, which constitute the mediators of the fundamental interactions. There are 8 gluons, mediating the strong interaction, 1 photon for the electromagnetic interaction and the $W^{\pm}$ and Z bozons for the weak interaction. The 3 generations of fermions are:

$$\begin{bmatrix} \nu_e & u \\ e^- & d' \end{bmatrix}, \quad \begin{bmatrix} \nu_\mu & c \\ \mu^- & s' \end{bmatrix}, \quad \begin{bmatrix} \nu_\tau & t \\ \tau^- & b' \end{bmatrix} \tag{2.1}$$

where

$$\begin{bmatrix} \nu_l & q_u \\ l^- & q_d \end{bmatrix} \equiv \begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L, \; l_R^-, \; q_{uR}, \; q_{dR} \tag{2.2}$$

The difference between the behavior of the left-handed and righ-handed comoponents of the fermionic fields under a $SU(2)_L$ transformation, will be discussed below.

2.1.2 Quantum Electrodynamics (QED)

The Lagrangian describing a free Dirac (fermion) field is:

$$\mathcal{L}_0 = i\bar{\psi}(x)\gamma^\mu \partial_\mu \psi(x) - m\bar{\psi}(x)\psi(x) \tag{2.3}$$

$\mathcal{L}_0$ is invariant with respect to U(1) transformations:

$$\psi(x) \xrightarrow{U(1)} \psi'(x) \equiv exp\{iQ\theta\}\psi(x) \tag{2.4}$$

However, if the phase θ is not constant, but is a function of space-time coordinates, $\theta \equiv \theta(x)$, then the Lagrangian of the free Dirac field is not invariant under U(1) transformations:

$$\partial_\mu\psi(x) \xrightarrow{U(1)} exp\{iQ\theta\}(\partial_\mu + iQ\partial_\mu\theta)\psi(x) \tag{2.5}$$

The 'gauge principle' is the requirement that this symmetry needs to hold even for such local phase transformation. In order to ensure this symmetry, an additional term has to be introduced, containing a spin-1 field A_μ, which transforms as:

$$A_\mu \xrightarrow{U(1)} A'_\mu \equiv A_\mu(x) - \frac{1}{e}\partial_\mu\theta \tag{2.6}$$

This naturally invites the definition of the covariant derivative:

$$D_\mu\psi(x) \equiv [\partial_\mu + ieQA_\mu(x)]\psi(x) \tag{2.7}$$

The covariant derivative transforms as follows:

$$D_\mu\psi(x) \xrightarrow{U(1)} (D_\mu\psi)'(x) \equiv exp\{iQ\theta\}D_\mu\psi(x) \tag{2.8}$$

The resulting Lagrangian, with the ordinary derivative being replaced with the newly introduced covariant derivative, is therefore invariant to local phase transformations:

$$\mathcal{L} \equiv i\bar{\psi}(x)\gamma^\mu D_\mu\psi(x) - m\bar{\psi}(x)\psi(x) = \mathcal{L}_0 - eQA_\mu(x)\bar{\psi}(x)\gamma^\mu\psi(x) \tag{2.9}$$

By making the Lagrangian symmetric, a new term has been generated, namely the second term in 2.9. This term represents an interaction between the Dirac field and the boson field A_μ. Fig. 2.1 illustrates this interaction vertex as a Feynamn diagram. An additional term is required to be included in the Lagrangian,

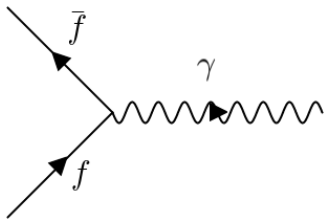


Figure 2.1: Feynman diagram of the Quantum Electrodynamics vertex

in order to account for the propagation of the A_μ field. This kinetic term also has to be gauge-invariant:

$$\mathcal{L}_{kin} \equiv -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \tag{2.10}$$

where $F_{\mu\nu}(x) \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength.

2.1.3 Quantum Chromodynamics (QCD)

The rich spectrum of hadronic states implies a deeper underlying structure. The internal components of hadrons are quarks. In the SM model, there are three generations of two quarks each, for a total of six, and their corresponding antiparticles. The spectrum of baryons is obtained from the combination of three quarks ($B \equiv qqq$) while mesons are bound states of a quark and an antiquark ($M \equiv q\bar{q}$). In order to preserve the Pauli exclusion principle for fermions, it is required that quarks possess an additional internal property, a quantum number called colour. Moreover, this quantum number is required to have three different values ($N_c = 3$). Thus, each quark comes in three different colours, q^α, where $\alpha = 1, 2, 3$ or red, blue and green. Hadrons are colour-singlet states:

$$B = \frac{1}{\sqrt{6}}\epsilon^{\alpha\beta\gamma}|q_\alpha q_\beta q_\gamma >, \qquad M = \frac{1}{\sqrt{3}}\delta^{\alpha\beta}|q_\alpha q_\beta > \tag{2.11}$$

Transformations in colour space are described mathematically by the elements of the $SU(3)_C$ group, which can be written in matrix form as:

$$U \;=\; \exp\left\{ i\frac{\lambda^a}{2}\theta_a \right\} \tag{2.12}$$

where $\frac{1}{2}\lambda^a$, $(a = 1, 2, ..., 8)$ are the generators of the $SU(3)_C$ group and θ_a are scalar parameters.
Considering the free Lagrangian:

$$\mathcal{L}_0 = \sum_f \bar{q}_f(i\gamma^\mu\partial_\mu - m_f)\; q_f \tag{2.13}$$

where f is the quark flavour, it can be verified that this Lagrangian is invariant with respect to the global transformations U. In analogy to the QED approach, the gauge principle is applied and the requirement that this Lagrangian is also invariant with respect to *local* $SU(3)_C$ transformations is imposed. This invites the definition of the covariant derivative:

$$D^\mu q_f \;\equiv\; \left[\partial^\mu + ig_s\frac{\lambda^a}{2}G^\mu_a(x)\right] \equiv \left[\partial^\mu + ig_s G^\mu(x)\right] q_f \tag{2.14}$$

Since the number of $SU(3)_C$ generators is eight, a matching number of gauge bosons, called gluons, are introduced.
The transformation properties of the fields and the covariant derivative are:

$$q^\alpha_f \to (q^\alpha_f)' = q^\alpha_f + i\left(\frac{\lambda^a}{2}\right)_{\alpha\beta}\delta\theta_a q^\beta_f \tag{2.15}$$

$$G^\mu_a \to (G^\mu_a)' = G^\mu_a - \frac{1}{g_s}\partial^\mu(\delta_a) - f^{abc}\delta\theta_b G^\mu_c \tag{2.16}$$

Before writing down the full Lagrangian, a gauge-invariant kinetic term for the gluon fields has to be included:

$$G^{\mu\nu}(x) \equiv -\frac{i}{g_s}\left[D^\mu, D^\nu\right] = \partial^\mu G^\nu - \partial^\nu G^\mu + i g_s \left[G^\mu, G^\nu\right] \equiv \frac{\lambda^a}{2} G_a^{\mu\nu}(x) \quad (2.17)$$

$$G_a^{\mu\nu}(x) = \partial^\mu G^\nu - \partial^\nu G^\mu - g_s f^{abc} G_b^\mu G_c^\nu \quad (2.18)$$

Having all the gauge-invariant components, the full QCD Lagrangian [42] can be written:

$$\begin{aligned}\mathcal{L}_{QCD} = &-\frac{1}{4}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a) + \sum_f \bar{q}_f^\alpha (i\gamma^\mu \partial_\mu - m_f) q_f^\alpha \\ &- g_s G_a^\mu \sum_f \bar{q}_f^\alpha \gamma_\mu \left(\frac{\lambda^a}{2}\right)_{\alpha\beta} q_f^\beta \\ &+ \frac{g_s}{2} f^{abc} \left(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu\right) G_\mu^b G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade} G_b^\mu G_c^\nu G_\mu^d G_\nu^e\end{aligned}$$

The gauge principle excludes the possibility of a mass term for the gluon fields, meaning that the gluons are massless spin-1 particles. The coupling constant g_s characterizes all of the QCD vertices, which are illustrated as Feynman diagrams in Fig. 2.2.

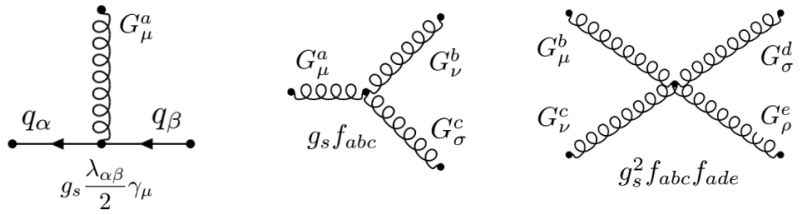


Figure 2.2: Feynman diagrams of the QCD interaction vertices

An interesting feature of the QCD Lagrangian is the presence of terms governing the self-interaction of the gluons. This self-interaction shows that the strong interaction is qualitatively different from the electromagnetic interaction, where the photon does not self-interact. In addition, this feature leads to very interesting properties of the strong interaction, such as confinement and asymptotic freedom.

The magnitude of the strong interaction coupling constant can be evaluated by studying the ratio of the hadronic Z decays:

$$Z \to q\bar{q} \qquad \text{and} \qquad Z \to q\bar{q}G \quad (2.19)$$

Such events, registered at the ALEPH experiment [43] at the former CERN accelerator LEP, are illustrated in Fig. 2.3. From these calculations, the obtained value of the coupling constant is $\alpha_s \equiv g_s^2/(4\pi) \sim 0.12$.

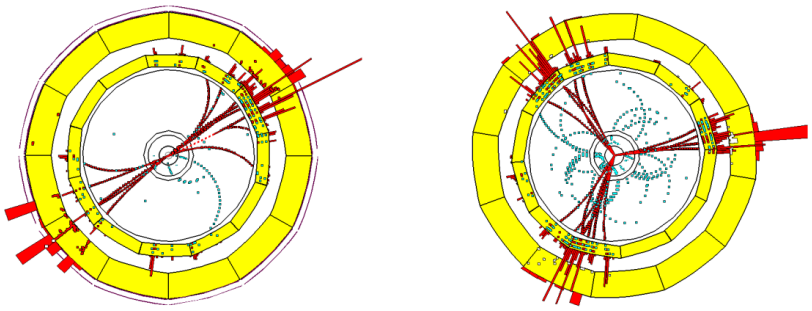


Figure 2.3: Two- and three-jet events from bosonic Z decays

2.1.4 Electroweak unification

In analogy to the QED and QCD cases, we start with the free Lagrangian:

$$\mathcal{L}_0 = i\bar{u}(x)\gamma^\mu\partial_\mu u(x) + i\bar{d}(x)\gamma^\mu\partial_\mu d(x) = \sum_{j=1}^{3} i\bar{\psi}_j(x)\gamma^\mu\partial_\mu\psi_j(x) \tag{2.20}$$

The field transformation properties under global transformations are:

$$\begin{aligned} \psi_1(x) &\xrightarrow{G} \psi_1'(x) \equiv exp\{iy_1\beta\}\, U_L\, \psi_1(x) \\ \psi_2(x) &\xrightarrow{G} \psi_2'(x) \equiv exp\{iy_2\beta\}\psi_2(x) \\ \psi_3(x) &\xrightarrow{G} \psi_3'(x) \equiv exp\{iy_3\beta\}\psi_3(x) \end{aligned} \tag{2.21}$$

To ensure invariance under local $SU(2)_L \otimes U(1)_Y$ transformations, the ordinary derivatives have to be replaced by the covariant derivatives:

$$\begin{aligned} D_\mu\psi_1(x) &\equiv \left[\partial_\mu + ig\tilde{W}_\mu(x) + ig'y_1B_\mu(x)\right]\psi_1(x) \\ D_\mu\psi_2(x) &\equiv \left[\partial_\mu + ig'y_2B_\mu(x)\right]\psi_2(x) \\ D_\mu\psi_3(x) &\equiv \left[\partial_\mu + ig'y_3B_\mu(x)\right]\psi_3(x) \end{aligned} \tag{2.22}$$

The transformation properties of the four generated gauge fields are:

$$B_\mu(x) \xrightarrow{G} B_\mu'(x) \equiv B_\mu(x) - \frac{1}{g'}\partial_\mu\beta(x) \tag{2.23}$$

$$\tilde{W}_\mu \xrightarrow{G} \tilde{W}_\mu' \equiv U_L(x)\tilde{W}_\mu U_L^\dagger(x) + \frac{i}{g}\partial_\mu U_L(x)U_L^\dagger(x) \tag{2.24}$$

where $U_L(x) \equiv \exp\left\{i\frac{\sigma_i}{2}\alpha^i(x)\right\}$.
Thus, we obtain a Lagrangian which is invariant under local $SU(2)_L \otimes U(1)_Y$ transformations:

$$\mathcal{L} = \sum_{j=1}^{3} i\tilde{\psi}_j(x)\gamma^\mu D_\mu\psi_j(x) \tag{2.25}$$

This Lagrangian is still missing the gauge-invariant kinetic term. To construct such a term and add it to the Lagrangian, the following field strengths are introduced:

$$B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.26}$$

$$\tilde{W}_{\mu\nu} \equiv -\frac{i}{g}\left[\left(\partial_\mu + ig\tilde{W}_\mu\right),\ \left(\partial_\nu + ig\tilde{W}_\nu\right)\right] = \partial_\mu\tilde{W}_\nu - \partial_\nu\tilde{W}_\mu + ig\left[\tilde{W}_\mu,\ \tilde{W}_\nu\right] \tag{2.27}$$

$$\tilde{W}_{\mu\nu} \equiv \frac{\sigma_i}{2}W^i_{\mu\nu}, \qquad W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon^{ijk}W^j_\mu W^k_\nu \tag{2.28}$$

Finally, the kinetic part of the Lagrangian is:

$$\mathcal{L}_{kin} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{2}\mathrm{Tr}\left[\tilde{W}_{\mu\nu}\tilde{W}^{\mu\nu}\right] = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu} \tag{2.29}$$

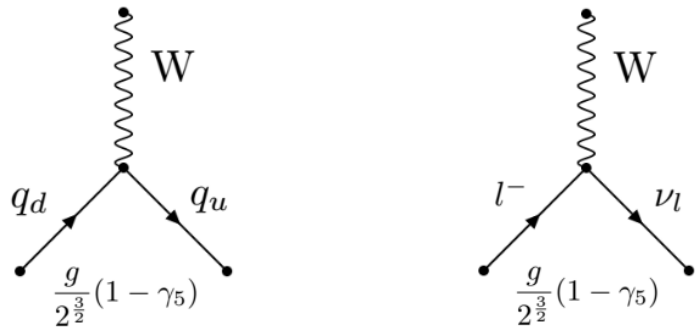


Figure 2.4: Feynman diagrams of electroweak charged current vertices

2.1.5 Spontaneous Symmetry Breaking

The Lagrangian obtained above has the obvious disadvantage of the lack of mass terms for the $W^\pm$ and Z bosons. Therefore, a modification is required that will introduce the needed mass terms without losing the gauge invariance. Such a mechanism is offered by what is called Spontaneous Symmetry Breaking (SSB) [44].

Goldstone theorem [45]

Considering a complex scalar field, with Lagrangian:

$$\mathcal{L} = \partial_\mu\phi^\dagger\partial^\mu\phi - V(\phi) \tag{2.30}$$

where

$$V(\phi) = \mu^2\phi^\dagger\phi + h\left(\phi^\dagger\phi\right)^2 \tag{2.31}$$

it can be verified that this Lagrangian is invariant with respect to global phase transformations of the field:

$$\phi(x) \quad \rightarrow \quad \phi'(x) \equiv \exp\{i\theta\}\,\phi(x) \tag{2.32}$$

If $\mu^2 < 0$, the potential minimum is at:

$$|\phi_0| = \sqrt{\frac{-\mu^2}{2h}} \equiv \frac{v}{\sqrt{2}} > 0, \qquad V(\phi_0) = -\frac{h}{4}v^4 \tag{2.33}$$

There is an infinite number of ground states, differing only by the phase θ: $\phi_0(x) = \frac{v}{\sqrt{2}}\exp\{i\theta\}$. Choosing a particular solution, e.g. $\theta = 0$ spontaneously breaks the symmetry. Excited states can be parametrized as:

$$\phi(x) \equiv \frac{1}{\sqrt{2}}\left[v + \varphi_1(x) + i\varphi_2(x)\right] \tag{2.34}$$

With the new parametrization, the following potential is obtained:

$$V(\phi) = V(\phi_0) - \mu^2\varphi_1^2 + hv\varphi_1\left(\varphi_1^2 + \varphi_2^2\right) + \frac{h}{4}\left(\varphi_1^2 + \varphi_2^2\right)^2 \tag{2.35}$$

φ_1 represents a massive state of mass $m_{\varphi_1}^2 = -2\mu^2$, while φ_2 represent a massless state.

If we introduce a doublet of complex scalar fields:

$$\phi(x) = \begin{pmatrix}\phi^{(+)}(x)\\ \phi^{(0)}(x),\end{pmatrix} \tag{2.36}$$

then the following Lagrangian is symmetric under local $SU(2)_L \otimes U(1)_Y$ transformations:

$$\mathcal{L}_S = (D_\mu\phi)^\dagger D^\mu\phi - \mu^2\phi^\dagger\phi - h(\phi^\dagger\phi)^2 \tag{2.37}$$

$$D^\mu\phi = [\partial^\mu + ig\widetilde{W}^\mu + ig'y_\phi B^\mu]\phi \tag{2.38}$$

The scalar doublet can be expressed as:

$$\phi(x) = \exp\left\{i\frac{\sigma_i}{2}\theta^i(x)\right\}\frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\ +\ H(x)\end{pmatrix} \tag{2.39}$$

in terms of the four real fields $\theta^i(x)$ and $H(x)$. Since the Lagrangian is invariant ot local $SU(2)_L$ transformations, a gauge can be chosen such that the three $\theta^i(x)$ fields are set to zero. Consequently:

$$(D_\mu\phi)^\dagger D^\mu\phi \xrightarrow{\theta^i=0} \frac{1}{2}\partial_\mu H\partial^\mu H + (v+H)^2\left\{\frac{g^2}{4}W_\mu^\dagger W^\mu + \frac{g^2}{8\cos^2\theta_W}Z_\mu Z^\mu\right\} \tag{2.40}$$

The quadratic terms for the $W^\pm$ and the Z mean that the gauge bosons have acquired masses:

$$M_Z \cos\theta_W = M_W = \frac{1}{2} v g \tag{2.41}$$

The introduced doublet of scalar fields and the associated Lagrangian lead to the generation of masses for the carriers of the weak interaction. Adding $\mathcal{L}_S$ to the electroweak Lagrangian (Eq. 2.25), produces a Lagrangian which preserves the gauge invariance but provides mass terms for the gauge bosons, generated via the SSB mechanism. Additionally, it gives a relation (Eq. 2.41) between the masses and the electroweak mixing angle θ_W, which can be verified experimentally.

The Higgs boson

Apart from the quadratic mass terms for the boson, the newly added Lagrangian (Eq. 2.37), introduces a new scalar particle H. Writing $\mathcal{L}_S$ as:

$$\mathcal{L}_S = \frac{1}{4} h v^4 + \mathcal{L}_H + \mathcal{L}_{HG^2} \tag{2.42}$$

where

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{1}{2} M_H^2 H^2 - \frac{M_H^2}{2v} H^3 - \frac{M_H^2}{b v^2} H^4 \tag{2.43}$$

$$\mathcal{L}_{HG^2} = M_W^2 W_\mu^\dagger W^\mu \left\{ 1 + \frac{2}{v} H + \frac{H^2}{v^2} \right\} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu \left\{ 1 + \frac{2}{v} H + \frac{H^2}{v^2} \right\} \tag{2.44}$$

shows explicitly the presence of the H scalar field. The mass of the Higgs particle can be expressed as follows:

$$M_H = \sqrt{-2\mu^2} = \sqrt{2h} v \tag{2.45}$$

As mentioned in Section 1.1, the Higgs boson was discovered jointly by the ATLAS and CMS experiments. The measured mass of the Higgs boson is ([46]):

$$m_H = 125.20 \pm 0.11 \; GeV \tag{2.46}$$

Fermion masses

At this point, the only remaining component to the SM Lagrangian are mass terms for the fermions. Coupling the scalar doublet, introduced in Eq. 2.36, with the fermionic fields, allows us to write a gauge-invariant Lagrangian, known as the Yukawa Lagrangian:

$$\mathcal{L}_Y = -c_1 (\bar{u},\ \bar{d})_L \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} d_R - c_2 (\bar{u},\ \bar{d})_L \begin{pmatrix} \phi^{(0)*} \\ -\phi^{(-)} \end{pmatrix} u_R - c_3 (\bar{\nu}_e,\ \bar{e})_L \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} e_R + \text{h.c.} \tag{2.47}$$

Choosing the unitary gauge, we obtain a simpler form:

$$\mathcal{L}_Y = -\frac{1}{2}(v+H)\left\{c_1\bar{d}d + c_2\bar{u}u + c_3\bar{e}e\right\} \tag{2.48}$$

The SSB generates the ferminoic masses:

$$m_d = c_1\frac{v}{\sqrt{2}}, \quad m_u = c_2\frac{v}{\sqrt{2}}, \quad m_e = c_3\frac{v}{\sqrt{2}} \tag{2.49}$$

The fermionic masses are free parameters of the SM and are determined only experimentally, a consequence of the fact that the c_i coefficients in Eq. 2.49 are unknown.

Factorizing the fermionic masses in the Lagrangian (Eq. 2.48):

$$\mathcal{L}_Y = -\left(1+\frac{H}{v}\right)\left\{m_d\bar{d}d + m_u\bar{u}u + m_e\bar{e}e\right\} \tag{2.50}$$

shows that the Higgs field coupling to the fermionic fields is directly proportional to their masses.

2.2 Flavor physics and CP violation

The Standard Model of particle physics contains three fundamental discrete symmetries: P (parity), C (charge conjugation) and T (time reversal). Two of the three forces described by the SM, namely the electromagnetic and strong interactions, fully preserve these discrete symmetries. However, it has been observed experimentally that the weak interaction fully violates both P and C. This violation leads to some very interesting consequences, to be discussed in this chapter. Combined discrete symmetries, such as CP and CPT, can also be constructed. In the rest of this section, we focus on the CP transformation and outline the theoretical framework for describing the potential sources of violation of this symmetry.

2.2.1 CKM matrix and meson mixing

The CKM matrix of the SM describes the mixing between flavor and mass eigenstates of down-type quarks.

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} \tag{2.51}$$

The individual elements $|V_{ij}|^2$, give the transition probability from the down-type quark of generation i, to the down-type quark of generation j.

The current best values for the CKM elements (PDG 2024 [46]) are given in Eq. 2.52.

$$|V_{CKM}| = \begin{pmatrix} 0.97367 \pm 0.00032 & 0.22431 \pm 0.00085 & 0.00382 \pm 0.00020 \\ 0.221 \pm 0.004 & 0.975 \pm 0.006 & 0.0411 \pm 0.0012 \\ 0.0086 \pm 0.0002 & 0.0415 \pm 0.0009 & 1.010 \pm 0.027 \end{pmatrix} \tag{2.52}$$

The unitarity of the CKM matrix leads to a set of nine constraints: three constraints on the diagonal terms (*weak universality*):

$$\sum_k |V_{ik}|^2 = \sum_i |V_{ik}|^2 = 1, \tag{2.53}$$

and six constraints regarding the off-diagonal terms:

$$\sum_k V_{ik} V_{jk}^* = 0 \quad \text{with fixed } i \text{ and } j, \text{ and} \tag{2.54}$$

$$\sum_i V_{ij} V_{ik}^* = 0 \quad \text{with fixed } j \text{ and } k \tag{2.55}$$

These six vanishing combinations can be represented as triangles in the complex plane. The B_s^0 unitarity triangle:

$$V_{us}V_{ub}^{*} + V_{cs}V_{cb}^{*} + V_{ts}V_{tb}^{*} = 0 \tag{2.56}$$

is of particular interest to the analysis presented in this dissertation. The angles of this triangle can be expressed in terms of the CKM elements as follows:

$$\alpha_s = \arg\left(-\frac{V_{ts}V_{tb}^{*}}{V_{us}V_{ub}^{*}}\right), \quad \beta_s = \arg\left(-\frac{V_{ts}V_{tb}^{*}}{V_{cs}V_{cb}^{*}}\right), \quad \gamma_s = \arg\left(-\frac{V_{us}V_{ub}^{*}}{V_{cs}V_{cb}^{*}}\right) \tag{2.57}$$

2.2.2 Mixing in B mesons

The neutral B mesons produced at particle colliders are flavor eigenstates, that are in turn a linear superposition of mass eigenstates (Eq. 2.58). This leads to a probability of the neutral B meson evolving in time into its antiparticle, with opposite quark content.

$$|\Psi(t)\rangle = |B_s^0(t)\rangle + |\bar{B}_s^0(t)\rangle \tag{2.58}$$

The time evolution of the B_s^0 system is governed by the Schrödinger equation:

$$i\frac{d}{dt}\begin{pmatrix}|B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle\end{pmatrix} = H\begin{pmatrix}|B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle\end{pmatrix} \tag{2.59}$$

The effective Hamiltonian H in Eq. 2.59 can be split in 2×2 Hermitian matrices, one representing mass (M) and one representing decay width (Γ):

$$H \equiv M - \frac{i}{2}\Gamma \tag{2.60}$$

The diagonal terms in the M and Γ matrices describe the flavor-conserving transitions $B_s^0 \leftrightarrow B_s^0$ and $\bar{B}_s^0 \leftrightarrow barB_s^0$, while the off-diagonal terms descrive flavor-changing transitions, $B_s^0 \leftrightarrow \bar{B}_s^0$ and $\bar{B}_s^0 \leftrightarrow B_s^0$. Additionally, the CPT theorem constraints the diagonal terms of the Hamiltonian to be equal, resulting in:

$$\text{CPT} \quad \leftrightarrow \quad \begin{matrix}M_{11} = M_{22} \\ \Gamma_{11} = \Gamma_{22}\end{matrix} \tag{2.61}$$

The two mass eigenstates, referred to as light and heavy, can be expressed in terms of the flavor eigenstates as:

$$\begin{aligned}|B_s^L\rangle &= p|B_s^0\rangle + q|\bar{B}_s^0\rangle \\ |B_s^H\rangle &= p|B_s^0\rangle - q|\bar{B}_s^0\rangle\end{aligned} \tag{2.62}$$

where p and q are complex coefficients. Eq. 2.59 can now be rewritten as:

$$i\frac{d}{dt}\begin{pmatrix}|B_s^L(t)\rangle \\ |B_s^H(t)\rangle\end{pmatrix} = \begin{pmatrix}M_L - \frac{i}{2}\Gamma_L & 0 \\ 0 & M - \frac{i}{2}\Gamma_H\end{pmatrix}\begin{pmatrix}|B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle\end{pmatrix} \tag{2.63}$$

where the indices L and H indicate the light and heavy eigenstates, respectively. The 2×2 Hamiltonian in Eq. 2.63 has been obtained by applying the following diagonalization:

$$Q^{-1}HQ = \begin{pmatrix} M_L - \frac{i}{2}\Gamma_L & 0 \\ 0 & M - \frac{i}{2}\Gamma_H \end{pmatrix}, \quad \text{with } Q = \begin{pmatrix} q & p \\ p & -q \end{pmatrix} \tag{2.64}$$

The ratio of the p and q coefficients can now be expressed as:

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}} \tag{2.65}$$

Time evolution of the mass eigenstates can be written as:

$$|B_s^{H,L}(t)\rangle = e^{-iM_{H,L}t - \frac{1}{2}\Gamma_{H,L}t}|B_s^{H,L}(t)\rangle \tag{2.66}$$

2.2.3 Mass eigenstates

As common metrics of the light and heavy eigenstates, the average mass m_s, the average decay width Γ_s, the mass difference Δm_s, and the decay width difference $\Delta\Gamma_s$ can be defined as:

$$m_s \equiv \frac{M_H + M_L}{2} = M_{11} = M_{22} \tag{2.67}$$

$$\Gamma_s \equiv \frac{\Gamma_H + \Gamma_L}{2} = \Gamma_{11} = \Gamma_{22} \tag{2.68}$$

$$\Delta m_s \equiv M_H - M_L \tag{2.69}$$

$$\Gamma_s \equiv \Gamma_L - \Gamma_H \tag{2.70}$$

In the case of B_s^0 mesons, the approximation $\Gamma_{12} \ll M_{12}$ is valid, leading to:

$$\Delta m_s \simeq 2|M_{12}| \tag{2.71}$$

$$\Delta\Gamma_s \simeq 2|\Gamma_{12}|cos\phi \tag{2.72}$$

$$\frac{q}{p} \simeq \sqrt{\frac{M_{12}^*}{M_{12}}} \tag{2.73}$$

where $\phi = \arg(-M_{12}/\Gamma_{12})$ is the relative phase between the off-diagonal terms. The approximation in Eq. 2.73 yields the important consequence:

$$|\frac{q}{p}| \simeq 1 \tag{2.74}$$

which is supported by both theoretical calculations ($|q/p| = 0.99989 \pm 0.00001$) and experimental results ($|q/p| = 1.0003 \pm 0.0014$). In addition, the phase $\phi_M = \arg(M_{12})$ can be used to express the ratio:

$$\frac{q}{p} \simeq e^{-i\phi_M} \tag{2.75}$$

which can further be expanded in terms of CKM matrix elements:

$$\frac{q}{p} \simeq \frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}} = e^{-i\phi_M} \tag{2.76}$$

2.2.4 Time evolution

Combining equations 2.62 and 2.66 yields the following expression for the time evolution of a B_s^0 state:

$$\begin{aligned} |B_s^0(t)\rangle &= g_+(t)|B_s^0\rangle + \frac{q}{p}g_-(t)|\bar{B}_s^0\rangle \\ |\bar{B}_s^0(t)\rangle &= g_+(t)|\bar{B}_s^0\rangle + \frac{p}{q}g_-(t)|B_s^0\rangle \end{aligned} \tag{2.77}$$

where $g_+(t)$ and $g_-(t)$, have been introduced as:

$$\begin{aligned} g_+(t) &= e^{-im_s t}e^{-\frac{\Gamma_s}{2}t}\left[\cosh\left(\frac{\Delta\Gamma_s t}{4}\right)\cos\left(\frac{\Delta m_s t}{2}\right) - i\sinh\left(\frac{\Delta\Gamma_s t}{4}\right)\sin\left(\frac{\Delta m_s t}{2}\right)\right] \\ g_-(t) &= e^{-im_s t}e^{-\frac{\Gamma_s}{2}t}\left[-\sinh\left(\frac{\Delta\Gamma_s t}{4}\right)\sin\left(\frac{\Delta m_s t}{2}\right) + i\cosh\left(\frac{\Delta\Gamma_s t}{4}\right)\cos\left(\frac{\Delta m_s t}{2}\right)\right] \end{aligned} \tag{2.78}$$

Using Eq. 2.77, the probability that a particle produced as $B_s^0(\bar{B}_s^0)$ at $t = 0$ has oscillated to a $\bar{B}_s^0(B_s^0)$ after a time t can be calculated:

$$\begin{aligned} |\langle B_s^0(0)||\bar{B}_s^0(t)\rangle|^2 &= |\frac{p}{q}|^2|g_-(t)|^2 \\ |\langle \bar{B}_s^0(0)||B_s^0(t)\rangle|^2 &= |\frac{q}{p}|^2|g_-(t)|^2 \end{aligned} \tag{2.79}$$

with

$$|g_-(t)|^2 = \frac{e^{-\Gamma_s t}}{2}\left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) - \cos(\Delta m_s t)\right] \tag{2.80}$$

Eq. 2.79 shows that the oscillation of the flavor eigenstates in the B_s^0 system depends on the mass difference Δm_s, the average decay width Γ_s and the decay width difference $\Delta\Gamma_s$.

2.3 CP violation in B mesons

The following amplitudes are introduced to characterise any transition $B_s^0 \to f$:

$$\begin{aligned} A_f &= \langle f|H|B_s^0\rangle, \;\; \bar{A}_f = \langle f|H|\bar{B}_s^0\rangle \\ A_{\bar{f}} &= \langle \bar{f}|H|B_s^0\rangle, \;\; \bar{A}_{\bar{f}} = \langle \bar{f}|H|\bar{B}_s^0\rangle \end{aligned} \tag{2.81}$$

A useful quantity for the description of CP-violation effects, is the phase-invariant complex parameter λ_f:

$$\lambda_f \equiv \frac{q}{p}\frac{\bar{A}_f}{A_f} \tag{2.82}$$

The time-dependent decay rates, for both states in the B_s^0 system, using the introduced mixing formalism, can be expressed as:

$$\begin{aligned} \Gamma_{B_s^0\to f}(t) =& |A_f|^2 \frac{1}{1+C_f}\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ &+ C_f \cos(\Delta m_s t) - S_f \sin(\Delta m_s t)\Bigg] \end{aligned} \tag{2.83}$$

$$\begin{aligned} \Gamma_{\bar{B}_s^0\to f}(t) =& |A_f|^2 |\frac{p}{q}|^2 \frac{1}{1+C_f}\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ &- C_f \cos(\Delta m_s t) + S_f \sin(\Delta m_s t)\Bigg] \end{aligned} \tag{2.84}$$

where the following definitions are used:

$$C_f \equiv \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \quad S_F \equiv \frac{2\mathrm{Im}(\lambda_f)}{1+|\lambda_f|^2}, \quad D_f \equiv -\frac{2\mathrm{Re}(\lambda_f)}{1+|\lambda_f|^2} \tag{2.85}$$

At this point, we can identify the two conditions needed for CP violation to occur: λ_f or $|q/p|$ being different from one. There are three different categories of CPV, which are explored separately in what follows.

2.3.1 CP violation in decay

Direct CP violation occurs when the decay rate of $B_s^0 \to f$ is different from the decay rate of $\bar{B}_s^0 \to \bar{f}$. In terms of the amplitudes:

$$|\frac{\bar{A}_{\bar{f}}}{A_f}| \neq 1 \tag{2.86}$$

If the final state is a CP eigenstate, λ_f can be expressed as:

$$\lambda_f = \eta_{CP}\frac{q}{p}\frac{\bar{A}_{\bar{f}}}{A_f} \tag{2.87}$$

where:

$$|\bar{f}\rangle = CP|f\rangle = \eta_{CP}|f\rangle, \tag{2.88}$$

$\eta_{CP} = \pm 1$ here is the CP eigenvalue of f. According to Eq. 2.87, if $|q/p| = 1$, direct CPV ind ecay is equivalent to $\lambda_f \neq 1$.

The amplitudes can be expanded in terms of their absolute value A_f, the strong phase $e^{i\theta}$ and the weak phase $e^{i\phi}$:

$$\begin{aligned} A_f &= |A_f| e^{i(\theta+\phi)} \\ \bar{A}_{\bar{f}} &= |A_f| e^{i(\theta-\phi)} \end{aligned} \tag{2.89}$$

Since physical observables are only sensitive to the square of the amplitude $|A_f|^2$, the occurence of CPV requires two interfering amplitudes:

$$\begin{aligned} A_f &= |A_1| e^{i(\theta_1+\phi_1)} + |A_2| e^{i(\theta_2+\phi_2)} \\ \bar{A}_{\bar{f}} &= |A_1| e^{i(\theta_1-\phi_1)} + |A_2| e^{i(\theta_2-\phi_2)} \end{aligned} \tag{2.90}$$

Measuring a difference in the rates would then result in a direct CPV measurement:

$$|A_f|^2 - |\bar{A}_{\bar{f}}|^2 = -4|A_1||A_2|\sin(\theta_1 - \theta_2)\sin(\phi_1 - \phi_2) \tag{2.91}$$

2.3.2 CP violation in mixing

CPV in mixing occurs when the probability of the transition $B_s^0 \to \bar{B}_s^0$ is different from the probability for $\bar{B}_s^0 \to B_s^0$:

$$|\langle B_s^0(0)||2\rangle|^2 \neq |\langle \bar{B}_s^0(0)||2\rangle|^2 \tag{2.92}$$

Using Eq. 2.79, the inequality of transition probabilities can be expressed as:

$$|\frac{p}{q}|^2 |g_-(t)|^2 \neq |\frac{q}{p}|^2 |g_-(t)|^2 \tag{2.93}$$

This leads to the following condition for CPV in mixing:

$$|\frac{q}{p}| \neq 1 \quad \leftrightarrow \quad |q| \neq |p| \tag{2.94}$$

In the absence of direct CPV, the above condition means a deviation from unity of λ_f.

2.3.3 CP violation in the interference between decay and mixing

The third CPV category occurs when the same final state is accessible to both the particle and the antiparticle, and the decay rate fulfill the following condition:

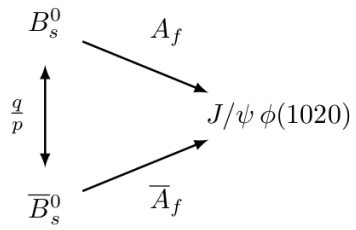


Figure 2.5: Two paths to final state

$$\Gamma(B^0_{s(\rightsquigarrow \bar{B}^0_s)} \to f)(t) \neq \Gamma(\bar{B}^0_{s(\rightsquigarrow B^0_s)} \to f)(t) \tag{2.95}$$

This type of CPV is a result of a non-zero complex phase of λ_f:

$$\mathrm{Im}(\lambda_f) \neq 0 \tag{2.96}$$

The time-dependent CP-assymetry can be expressed by (after plugging-in Eq. 2.83 and Eq. 2.84):

$$a_{CP}(t) = \frac{\Gamma_{\bar{B}^0_s \to f}(t) - \Gamma_{B^0_s \to f}(t)}{\Gamma_{\bar{B}^0_s \to f}(t) + \Gamma_{B^0_s \to f}(t)} = \frac{S_f \mathrm{sin}(\Delta m t) - C_f \mathrm{cos}(\Delta m t)}{\mathrm{cosh}(\frac{\Delta\Gamma t}{2}) - D_f \mathrm{sinh}(\frac{\Delta\Gamma t}{2})} \tag{2.97}$$

where $S_f \propto 2\mathrm{Im}(\lambda_f)$, $D_f \propto -\mathrm{Re}(\lambda_f)$, and $C_f \propto 1 - |\lambda_f|^2$.

Golden Modes Decays where all leading diagrams carry the same CP-violating phase are called *golden modes*. In such decays, $|A_f| = |\bar{A}_f|$ so there is no direct CPV. This simplifies the time-evolving CP-assymetry (Eq. 2.97) to:

$$a_{CP}(t) = \frac{\mathrm{Im}(\lambda_f)\mathrm{sin}(\Delta m t)}{\mathrm{cosh}(\frac{\Delta\Gamma t}{2}) + \mathrm{Re}(\lambda_f)\mathrm{sinh}(\frac{\Delta\Gamma t}{2})} \tag{2.98}$$

In B meson golden modes, the CP-violating phase can be easily computed, as the phase of $\bar{A}/A$ is trivially extracted from the phase of the CKM elements and the phase of q/p can be computed from Eq. 2.75.

2.4 The $B^0_s \to J/\psi\phi(1020)$ decay

The final state of the decay ($J/\psi\phi(1020)$) studied in the physics analysis presented in this dissertation is a CP eigenstate. This final state is accessible to both the B^0_s and its antiparticle (Fig. 2.5). The total branching fraction for this channel is:

$$\mathcal{B}(B^0_s \to J/\psi\phi(1020) \to \mu^+\mu^- K^+ K^-) \approx 3 \cdot 10^{-5} \tag{2.99}$$

The decay amplitude $A_{J/\psi\phi}$ can be written in terms of the tree-level (t) and higher-order "penguin" (p_q) contributions:

$$\begin{aligned} A_{f_C} &= (V_{cb}^* V_{cs})t + \sum_{q=u,c,t} (V_{qb}^* V_{qs}) p_q \\ &= (V_{cb}^* V_{cs})(t - p_c - p_t) + (V_{ub}^* V_{us})(p_u - p_t) \\ &= (V_{cb}^* V_{cs})T + (V_{ub}^* V_{us})P \end{aligned} \tag{2.100}$$

where $T \equiv t - p_c - p_t$ and $P \equiv p_u - p_t$. The penguin contributions are supressed by a factor of $\mathcal{O}(10^{-2})$, which allows us to regard this decay as a *golden mode*. Under this assumption, and considering that the tree-level transitions only contain one phase, we can conclude that there is no direct CPV in this channel:

$$|A_{f_{CP}}| = |\bar{A}_{\bar{f}_{CP}}|, \quad \text{with} \;\; f_{CP} \equiv J/\psi\phi(1020) \tag{2.101}$$

In the rest of this overview, the following simplified notation will be used:

$$\boxed{\lambda_{J/\psi\phi} \equiv \lambda}, \quad f_{J/\psi\phi(1020)} \equiv f \tag{2.102}$$

The tree-level amplitudes A_f and $\bar{A}_f$ can be expressed as:

$$\begin{aligned} A_f &= |A_f| e^{i\theta_D} e^{i\phi_D} \\ \bar{A}_f &= \eta_f |A_f| e^{i\theta_D} e^{-i\phi_D} \end{aligned} \tag{2.103}$$

where θ_D is the strong and ϕ_D the weak decay phase while η_f is the CP eigenstate of the final state.

In the $B_s^0 \to J/\psi\phi(1020)$ decay, the final state does not have a definite CP eigenvalue η_f, as it is a superposition of a CP-odd and CP-even eigenstates. This stems from the process being a decay of a pseudo-scalar meson with spin 0 to two vector mesons with spin 1, allowing for three possible values for the final state orbital angular momentum. The CP eigenvalue of the final state can then be calculated as a function of the orbital momentum l:

$$CP|\text{f.s.}\rangle_l = \eta_f |\text{f.s.}\rangle_l = (-1)^l |\text{f.s.}\rangle_l \tag{2.104}$$

The amplitude ratio is:

$$\frac{\bar{A}_f}{A_f} = \eta_f e^{-2i\phi_D} = \eta_f \frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*} \tag{2.105}$$

where the amplitudes have been expressed in terms of the CKM elements:

$$\begin{aligned} A_f &= \;\; V_{cs} V_{cb}^* \\ \bar{A}_f &= \eta_f V_{cs}^* V_{cb} \end{aligned} \tag{2.106}$$

Combining Eq. 2.82 and Eq. 2.105, we obtain:

$$\lambda = \frac{q}{p}\frac{\bar{A}_f}{A_f}\eta_f \left(\frac{V_{ts} V_{tb}^*}{V_{ts}^* V_{tb}}\right)\left(\frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*}\right) = \eta_f e^{-i\phi_s} \tag{2.107}$$

where $\phi_s = \phi_M + 2\phi_D$ is the phase associated with CPV in the interference between decay and mixing of $B_s^0 \to J/\psi\phi(1020)$. This phase can be calculated as follows:

$$\begin{aligned}
\phi_s &= -\text{arg}\left(\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}\frac{V_{cs}^*V_{cb}}{V_{cs}V_{cb}^*}\right) \\
&= -2\text{arg}\left(\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) \\
&= -2\left[\text{arg}\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) - \pi\right] \\
&= -2\beta_s - 2\pi \\
&= \boxed{-2\beta_s}
\end{aligned} \tag{2.108}$$

where β_s is the angle of the B_s^0-relevant unitairy triangle introduced in Eq. 2.57.

The best theoretical calculations of this phase are provided by the CKMfitter group and the UTfit collaboration:

$$\phi_s^{\text{CKMfitter}} = -2\beta_s^{\text{CKMfitter}} = -36.82^{+0.60}_{-0.86}\,\text{mrad} \tag{2.109}$$

$$\phi_s^{\text{UTfit}} = -2\beta_s^{\text{UTfit}} = -36.8 \pm 1.0\,\text{mrad} \tag{2.110}$$

In the next chapter, the CKMfitter value will be the one used to make comparisons with the value measured in the analysis presented in this dissertation.

In conclusion, the time-evolution of the CP-assymetry can be expressed as (using Eq. 2.98):

$$\begin{aligned}
a_{CP}(t) &= \frac{\Gamma_{\bar{B}_s^0\to J/\psi\phi(1020)}(t) - \Gamma_{B_s^0\to J/\psi\phi(1020)}(t)}{\Gamma_{\bar{B}_s^0\to J/\psi\phi(1020)}(t) + \Gamma_{B_s^0\to J/\psi\phi(1020)}(t)} \\
&= \frac{-\eta_f \text{sin}\phi_s \text{sin}(\Delta m_s t)}{\text{cosh}(\frac{\Delta\Gamma_s t}{2}) + \eta_f \text{cos}\phi_s \text{sinh}(\frac{\Delta\Gamma_s t}{2})}
\end{aligned} \tag{2.111}$$

2.4.1 The $B_s \to J/\psi\phi(1020)$ decay rate

The time dependent decay rates are (using Eq. 2.83):

$$\begin{aligned}
\frac{d\Gamma(B_s^0 \to J/\psi\phi(1020))}{dt} \propto |A_f|^2 \frac{1}{1+C} e^{-\Gamma_s t}\Bigg[&\text{cosh}\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f D \text{sinh}\left(\frac{\Delta\Gamma_s t}{2}\right) \\
&+ C\text{cos}(\Delta m_s t) - \eta_f S \text{sin}(\Delta m_s t)\Bigg], \\
\frac{d\Gamma(\bar{B}_s^0 \to J/\psi\phi(1020))}{dt} \propto |A_f|^2 \frac{1}{1+C} e^{-\Gamma_s t}\Bigg[&\text{cosh}\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f D \text{sinh}\left(\frac{\Delta\Gamma_s t}{2}\right) \\
&- C\text{cos}(\Delta m_s t) + \eta_f S \text{sin}(\Delta m_s t)\Bigg]
\end{aligned} \tag{2.112}$$

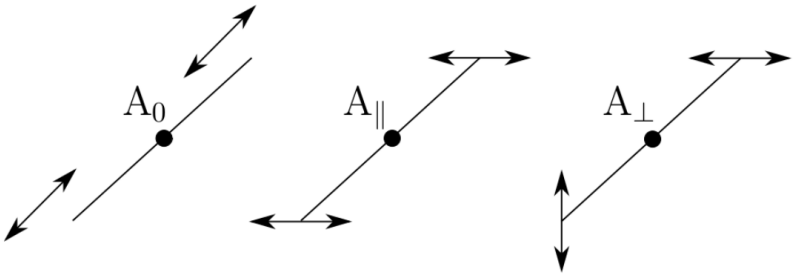


Figure 2.6: Polarization amplitudes of the $J/\psi\phi(1020)$ final state

where the coefficients have been defined as (based on Eq. 2.107):

$$C = \frac{1-|\lambda|^2}{1+|\lambda|^2}, \quad S = -\frac{2|\lambda|\sin(\phi_s)}{1+|\lambda|^2}, \quad D = -\frac{2|\lambda|\cos(\phi_s)}{1+|\lambda|^2} \tag{2.113}$$

The η_f has been explicitly shown and removed from the definition of the coefficients. The decay rates are invariant under the following transformations:

$$\begin{cases} \phi_s & \leftrightarrow & \pi - \phi_s \\ \Delta\Gamma_s & \leftrightarrow & -\Delta\Gamma_s \end{cases} \tag{2.114}$$

The time dependent decay rates (Eq. 2.112) contain the final state CP eigenvalue η_f, which is in turn determined by the orbital momentum of the final state:

$$\eta_f = (-1)^l \quad for \quad l = 0, 1, 2 \tag{2.115}$$

Since the final state orbital momentum for each individual event is not a quantity measured within the CMS experiment, and is therefore, inaccessible to the analysis, a statistical approach is followed: an analysis is performed on the angular distribution of the final-state particles.

In order to perform this analysis, the so-called "transversity" basis is employed. This basis is defined by three different polarization states, each associated to a corresponding amplitude (Fig. 2.6):

- the CP-even amplitude A_0, representing the longitudinal polarization, where the spins of the final-state particles are aligned to their momenta,
- the CP-even amplitude $A_\parallel$, representing the parallel transverse polarization, where the spins are parallel to each other,
- the CP-odd amplitude $A_\perp$, representing the perpendicular transverse polarization, where the spins are perpendicular to each other.

The amplitudes follow the identity:

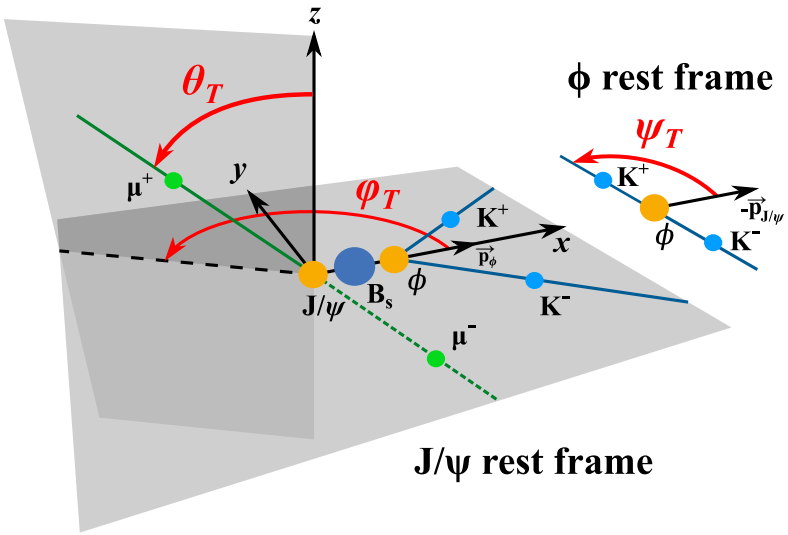


Figure 2.7: Definition of the three angles of the transversity basis θ_T ψ_T ϕ_T

$$|A_0|^2 + |A_\perp|^2 + |A_\parallel|^2 = 1 \tag{2.116}$$

Apart from these three amplitudes, an additional one, A_S (for S-wave) is included to account for the presence in the data sample of the non-resonant channels $B_s^0 \to J/\psi K^+K^-$ and $B_s^0 \to J/\psi f_0(980) \to \mu^+\mu^- K^+K^-$.

The transversity basis introduced above, is parametrized by three angles $\Theta = (\theta_T, \psi_T, \phi_T)$. As shown in Fig. 2.7, they are defined as the polar angle (θ_T) and azimuthal angle (ϕ_T) of μ^+ in the rest frame of the J/ψ meson where the x-axis is the B_s^0 direction and the xy-plane coincides with the plane of the decay $\phi \to$, while ψ_T is defined in the ϕ rest frame as the angle between the K^+ momentum and the direction opposite the J/ψ momentum.

The differential decay rate (Eq. 2.112), can now be written in terms of the transversity angles:

$$\frac{d^4\Gamma(B_s^0)}{d\Theta dt} = f(\Theta, t|\alpha) \propto \sum_{i=1}^{10} O_i(\alpha, t) \cdot g_i(\Theta) \tag{2.117}$$

where

$$\begin{aligned} O_i(\alpha, t) = N_i e^{-\Gamma_s t}\Bigg[& a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ & + c_i \cos(\Delta m_s t) + d_i \sin(\Delta m_s t)\Bigg] \end{aligned} \tag{2.118}$$

are time-dependent functions, with N_i, a_i, b_i, c_i, and d_i are defined in Tab. 2.1. The angular functions, g_i, are given explicitly in Tab. 2.2 while $\alpha =$

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

Table 2.1: Time-dependent terms of the differential decay rate

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	$\frac{4}{3}\sqrt{3}\cos\psi_T(1-\sin^2\theta_T\cos^2\varphi_T)$

Table 2.2: The angular functions g_i

(ϕ_s, $\Delta\Gamma_s$, Γ_s, Δm_s, $|\lambda|$, $|A_0|$, $|A_\parallel|$, $|A_\perp|$, $|A_S|$, δ_0, $\delta_\parallel$, $\delta_\perp$, δ_S) is the set of physical parameters investigated in the physics analysis presented in this dissertation.

2.4.2 Measurements and calculations

The theoretical predictions on the physics parameters listed above, are reported in Tab. 2.3, where available. On the other hand, Tab. 2.4 displays the best experimentally obtained values of those same parameters, containing contributions from experiments at the Fermilab Tevatron [47–51] as well as the ATLAS, CMS, and LHCb experiments [52–63] on the LHC.

These values will be utilized in the next chapter to compare our measurements with existing experimental results and theoretical predictions.

Parameter	Prediction	Ref.
φ_s	-37 ± 1mrad	[64, 65]
$\Delta\Gamma_s$	0.091 ± 0.031 ps^{-1}	[66]
Δm_s	18.77 ± 0.86 $\hbar$ps^{-1}	[66]
$\lvert\lambda\rvert$	≈ 1	SM

Table 2.3: Theoretical predictions of the physics parameters describing the $B_s^0 \to J/\psi\phi(1020)$ decay rate

Parameter	World-average	Ref.
ϕ_s	-49 ± 19mrad	[67]
$\Delta\Gamma_s$	0.084 ± 0.005ps^{-1}	[67]
Γ_s	0.6573 ± 0.0023ps^{-1}	[67]
Δm_s	$17.765 \pm 0.006\hbar$ps^{-1}	[67]
$\lvert\lambda\rvert$	1.001 ± 0.018	[46]
$\lvert A_0\rvert^2$	0.520 ± 0.003	[67]
$\lvert A_\parallel\rvert^2$	0.2222 ± 0.0027	[67]
$\lvert A_\perp\rvert^2$	0.253 ± 0.006	[67]
$\lvert A_S\rvert^2$	0.030 ± 0.005	[67]
$\delta_\parallel$	3.18 ± 0.06rad	[67]
$\delta_\perp$	3.08 ± 0.12rad	[67]
$\delta_S - \delta_\perp$	0.23 ± 0.05rad	[67]

Table 2.4: World-average values for the physics parameters describing the $B_s^0 \to J/\psi\phi(1020)$

Part II

Results

Chapter 3

Precision measurements in $B_s^0 \to J/\psi\phi(1020)$ decays

This chapter covers the first of three separate works performed in the context of this disseration, namely the precision measurement of CP-violation and related parameters in the $B_s^0 \to J/\psi\phi(1020)$ decays. The extensive use of Machine Learning methods is emphasized throughout. Besides a general description of the analysis workflow, from the selection and reconstruction algorithms to the fit model and physics parameters extraction, a more detailed account is provided on several topics, such as: the novel algorithms for neutral B meson flavor tagging, angular efficiency estimation, MC simulated events reweighting, and the estimation of potential background contributions. The final section presents the combined results with the previous CMS measurement of this kind.

3.1 Analysis overview

The analysis is performed using data collected by the CMS experiment at the CERN LHC in 2017-2018 at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity $\mathcal{L}_{int} = 96.5$ fb^{-1}. Tab. 3.1 lists the used datasets, along with the corresponding sizes of each.

MC simulated samples (as listed in Tab. 3.2 and Tab. 3.3), were utilized in order to develop the analysis tools and to validate the fit model. Three types of MC samples are used:

- "Standard" $B_s^0 \to J/\psi\phi(1020)$ samples, represent the signal as expected in data
- $B_s^0 \to J/\psi\phi(1020)$ samples where $\Delta\Gamma_s$ is forced to be equal to 0 ps^{-1}, used for the evaluation of time and angular efficiencies
- $B_s^0 \to J/\psi\phi(1020)$ samples where no acceptance or kinematic filters are aplied and $\Delta\Gamma_s$ is forced to be equal to 0 ps^{-1}

Data set name (DAS)	$\mathcal{L}_{int}$ [fb^{-1}]
Charmonium/Run2017C-UL2017_MiniAODv2-v1/MINIAOD	9.57
Charmonium/Run2017D-UL2017_MiniAODv2-v1/MINIAOD	4.25
Charmonium/Run2017E-UL2017_MiniAODv2-v1/MINIAOD	9.31
Charmonium/Run2017F-UL2017_MiniAODv2-v1/MINIAOD	13.54
Charmonium/Run2018A-UL2018_MiniAODv2-v1/MINIAOD	14.03
Charmonium/Run2018B-UL2018_MiniAODv2-v1/MINIAOD	7.06
Charmonium/Run2018C-UL2018_MiniAODv2-v1/MINIAOD	6.89
Charmonium/Run2018D-UL2018_MiniAODv2-v1/MINIAOD	31.83
tot	96.5

Table 3.1: List of the data sets used for the analysis along with the respective integrated luminosities

Data set name (DAS)	Description	#evt
BsToJpsiPhi_BmuonFilter_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Signal	$95 \cdot 10^6$
BsToJpsiPhi_JpsiPhiFilterDGO_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Signal ($\Delta\Gamma_s = 0$)	$93 \cdot 10^6$
BsToJpsiK_BmuonFilter_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Tagging calibration channel	$24 \cdot 10^6$
BdToJpsiKstar_BmuonFilter_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Control channel	$74 \cdot 10^6$
[1]RunIISummer20UL17MiniAODv2-106X_mc2017_realistic_v9-v2/MINIAODSIM		
[2]RunIISummer20UL17MiniAODv2-106X_mc2017_realistic_v9-v1/MINIAODSIM		

Table 3.2: List of simulated MC samples for 2017 used in this analysis along with the respective number of events and the application case of each

Data set name (DAS)	Description	#evt
BsToJpsiPhi_BmuonFilter_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Signal	$97 \cdot 10^6$
BsToJpsiPhi_JpsiPhiFilterDGO_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Signal ($\Delta\Gamma_s = 0$)	$93 \cdot 10^6$
BsToJpsiPhi_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Signal ($\Delta\Gamma_s = 0$, no filters)	$10 \cdot 10^6$
BsToJpsiK_BmuonFilter_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[2]	Tagging calibration channel	$25 \cdot 10^6$
BdToJpsiKstar_BmuonFilter_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[2]	Control channel	$74 \cdot 10^6$
LambdaBToJpsiKp_SoftQCDnonD_TuneCP5_13TeV-pythia8-evtgen/[1]	Possible background	$97 \cdot 10^6$
[1]RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16-_L1v1-v2/MINIAODSIM		
[2]RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16-_L1v1-v1/MINIAODSIM		

Table 3.3: List of simulated MC samples for 2018 used in this analysis along with the respective number of events and the application case of each

3.2 Selection and reconstruction

The events used in this analysis are filtered by two triggers of the CMS High-Level Trigger (HLT):

- HLT_Dimuon0_Jpsi3p5_Muon2, requires that the events contains a $J/\psi \to \mu^+\mu^-$ candidate plus an additional muon. The third is necessary for the implementation of flavor tagging (Sec. 3.3) in order to infer the initial flavor of the B_s^0.

- HLT_DoubleMu4_JpsiTrkTrk_Displaced, which requires a displaced $J/\psi \to \mu^+\mu^-$ candidate plus a couple of charged tracks with invariant mass in the

region of the $\phi(1020)$ resonance.

The full data set, introduced in Sec. 3.1 is divided in two separate data sets, based on the HLT triggers mentioned above. The first one, with events selected by HLT_Dimuon0_Jpsi3p5_Muon2 trigger, will be referred to as *dataset_JpsiMuon* while the data set containing events selected by the HLT_DoubleMu4_JpsiTrkTrk_Displaced trigger, will be referred to as *dataset_JpsiTrkTrk*. To avoid any overlap between the two, the first trigger is explicitly vetoed in the second data set (*dataset_JpsiTrkTrk*).

The next step in the selection process involves applying a set of offline selection requirements and cuts. The common requirements for both data sets are summarized in Tab. 3.4, while the specific criteria for *dataset_JpsiMuon* and *dataset_JpsiTrkTrk* are detailed in Tab. 3.5 and Tab. 3.6, respectively.

Candidate	Variable	Requirement	Optimised
Event	PV found	Yes	
	Good data-taking conditions	Yes	
	Trigger objects matched	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 μm	
$J/\psi \to \mu^+\mu^-$	Muon reconstruction	*Global*	
	Muon identification	*Loose*	
	$\lvert\eta(\mu)\rvert$	< 2.4	Yes
	$\left\lvert m(\mu^+\mu^-) - m^{\text{w.a.}}_{J/\psi}\right\rvert$	< 150 MeV	
$\phi(1020) \to K^+K^-$	Track quality	*High purity*	
	$\lvert\eta(K)\rvert$	< 2.5	Yes
	$\left\lvert m(K^+K^-) - m^{\text{w.a.}}_{\phi(1020)}\right\rvert$	< 10 MeV	
	Number of tracker hits	≥ 4	

Table 3.4: Common selection criteria for the two data sets: *dataset_JpsiMuon* and *dataset_JpsiTrkTrk*

Variable	Requirement	Optimised
HLT	HLT_Dimuon0_Jpsi3p5_Muon2	
$p_T(\mu)$	> 3.5 GeV	Yes
$p_T(K)$	> 1.15 GeV	Yes
$ct(B_s^0)$	> 60 μm	Yes

Table 3.5: Selection criteria for *dataset_JpsiMuon* specifically

Once the offline selection criteria are applied, a total of 623121 $B_s^0 \to J/\psi\phi(1020) \to K^+K^-\mu^+\mu^-$ remain for analysis, divided between *dataset_JpsiMuon*

Variable	Requirement	Optimised
HLT	HLT_DoubleMu4_JpsiTrkTrk_Displaced	
HLT veto:	HLT_Dimuon0_Jpsi3p5_Muon2	
$p_T(\mu)$	> 4 GeV	Yes
$p_T(K)$	> 0.9 GeV	Yes
$p_T(\mu^+\mu^-)$	> 6.9 GeV	
$ct(B_s^0)$	> 100 μm	Yes
$ct(B_s^0)/\sigma_{ct}$	> 3	

Table 3.6: Selection criteria for *dataset_JpsiTrkTrk* specifically

(67908) and *dataset_JpsiTrkTrk* (555213). Invariant mass distributions for the obtained J/ψ, $\phi(1020)$, and B_s^0 are shown in Figs. 3.1, 3.2, and 3.3, respectively.

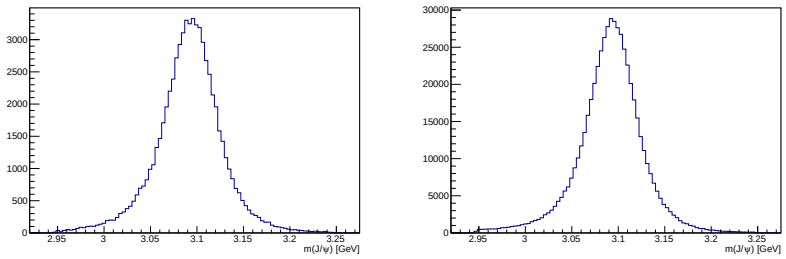


Figure 3.1: Invariant mass of the selected J/ψ candidates, from *dataset_JpsiMuon* (left) and *dataset_JpsiTrkTrk* (right)

3.3 Flavor tagging

Flavor tagging algorithms encompass a variety of methods used in analysis to identify the flavor of a meson, distinguishing it from its antiparticle. In our analysis, the flavour of the B_s^0 meson at production is an essential input for the measurement of CP-violation in $B_s^0 \to J/\psi\phi(1020)$ decays. Additionaly, in this particular case, the flavor of the initial meson cannot be deduced by the charge of the final-state particles. Instead, an indirect approach is followed. This approach leverages the fact that b-quarks are predominantly born in $b\bar{b}$ pairs, forming the basis for the following four algorithms (Fig. 3.4):

- **opposite-side muon tagger:** uses semileptonic $b \to \mu X$ decays of the other b-hadron in the event to infer the flavor of the B_s^0
- **opposite-side electron tagger:** uses semileptonic $b \to eX$ decays of the other b-hadron in the event to infer the flavor of the B_s^0
- **opposite-side jet tagger:** uses charge assymetries in the jet produced by the other b-hadron in the event to infer the flavor of the B_s^0

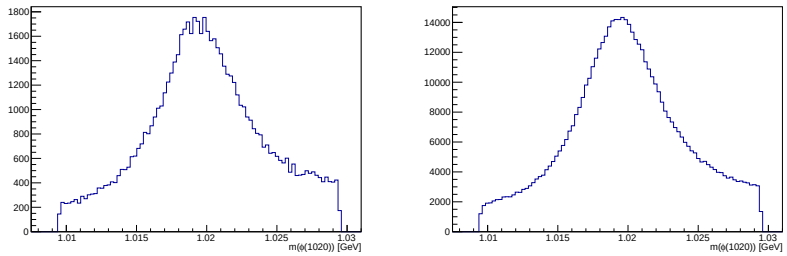


Figure 3.2: Invariant mass of the selected $\phi(1020)$ candidates, from *dataset_JpsiMuon* (left) and *dataset_JpsiTrkTrk* (right)

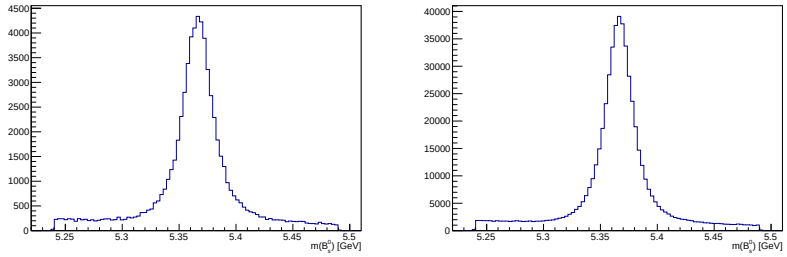


Figure 3.3: Invariant mass of the selected B_s^0 candidates, from *dataset_JpsiMuon* (left) and *dataset_JpsiTrkTrk* (right)

- **same-side tagger**: uses charge assymetries in the products of the B_s^0 hadronization

Tab. 3.7 and 3.8 give an overview of the logic employed in combining the tagging algorithms.

At the output of a tagging algorithm, each event is labelled by a tagging decision ξ_{tag}, defined as:

$$\xi_{tag} = \begin{cases} +1, & \text{for } B_s^0\ (\bar{b}s) \\ -1, & \text{for } \bar{B}_s^0\ (b\bar{s}) \\ 0, & \text{if no tagging decision is made} \end{cases} \tag{3.1}$$

Since not all events are tagged, the tagging efficiency is a valuable metric for quantifying the percentage of events that are successfully tagged:

$$\varepsilon_{tag} = \frac{N_{tag}}{N_{tot}}, \tag{3.2}$$

where N_{tag} is the number of tagged events and N_{tot} is the total number of events. Some events may be incorrectly tagged however, which prompts the introduction of the mistag probability:

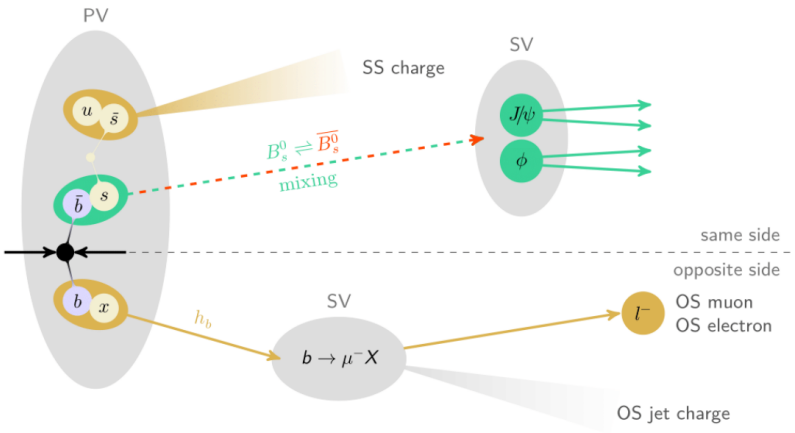


Figure 3.4: An overview of the flavor tagging algorithms employed in this analysis

$$\omega_{tag} = \frac{N_{WT}}{N_{tag}}, \tag{3.3}$$

where N_{WT} is the number of the wrongly-tagged events. In addition, a rescaled efficiency, referred to as "tagging power" is defined:

$$P_{tag} = \varepsilon_{tag}(1 - 2\omega_{tag})^2 = \varepsilon_{tag}\mathcal{D}_{tag}^2, \tag{3.4}$$

where $\mathcal{D}_{tag}$ is known as the *tagging dilution*, a metric for the tagging performance degradation due to mistagged events.

In the implementation of the algorithms, Deep Neural Networks (DNN) are trained to determine the mistag probability. Due to the fact that the models are trained on simulated events but intended to be used on data, there is a miscalibration which arises from the mismodeling in the simulated events. Additionally, large DNN are known to exhibit a certain degree of intrinsic miscalibration [68].

To correct for these effects, the ω_{tag}^{DNN} output by the DNN, is calibrated in data, using a sample of self-tagging $B^+ \to J/\psi K^+$ decays. In this decay, the flavor at production time is directly determined by the charge of the kaon. This allows to directly measure the mistag fraction ω_{tag}^{bin}. The resulting calibration plots (Figs. 3.5, 3.6, 3.7, 3.8 and 3.9) are fitted and the obtained calibration function is used to improve the estimation of the mistag probability. The SS-tagger undergoes some additional steps including a calibration in

$$B^+ \to J/\psi K^+$$

and $B_s^0 \to J/\psi\phi(1020)$ simulated events, with the correpsonding calibration curves shown in Fig. 3.10.

The final tagging performances are evaluated in the data sample used for the CPV measurement, by integrating the $\mathcal{D}_{tag}^2$ distribution. The results are

presented in Tab. 3.9. The combined performance of the tagging framework achieves a tagging power of $\hat{\mathcal{P}} = (5.59 \pm 0.02)\%$ with a tagging efficiency of $\varepsilon_{tag} = (55.9 \pm 0.1)\%$ and $\mathcal{D}^2_{tag} = 0.100$ which is approximately 4 times the one achieved during the Run-I CMS analysis [57] and among the highest tagging performances ever achieved for neutral B meson tagging in a hadronic collider experiment.

3.3.1 Tagging in the decay model

The tagging information is included in the decay model by modifying the O_i functions (Eq. 3.5) of the differential decay rate (Eq. 2.117), as follows:

$$O_i(\alpha, t) = N_i e^{-\Gamma_s t}\Bigg[a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) + c_i \xi_{tag}(1-2\omega_{evt})\cos(\Delta m_s t) + d_i \xi_{tag}(1-2\omega_{evt})\sin(\Delta m_s t)\Bigg] \tag{3.5}$$

where the tagging decision ξ_{tag} and the per-event tagging dilution $\mathcal{D}_{evt}$ are applied to each of the c_i and d_i terms.

In order to constrain the mistag probability ω_{tag} within the [0, 0.5] range, the following transformation is applied:

$$(\xi_{tag},\ \omega_{evt}) = \begin{cases} (\xi_{tag},\ \omega_{evt}) & if\ \ \omega_{evt} \leq 0.5 \\ (-\xi_{tag},\ 1-\omega_{evt}) & if\ \ \omega_{evt} > 0.5 \end{cases} \tag{3.6}$$

Data set	OS muon	OS electron	OS jet	SS
dataset_JpsiMuon	✓			
dataset_JpsiTrkTrk	✓	✓	✓	✓

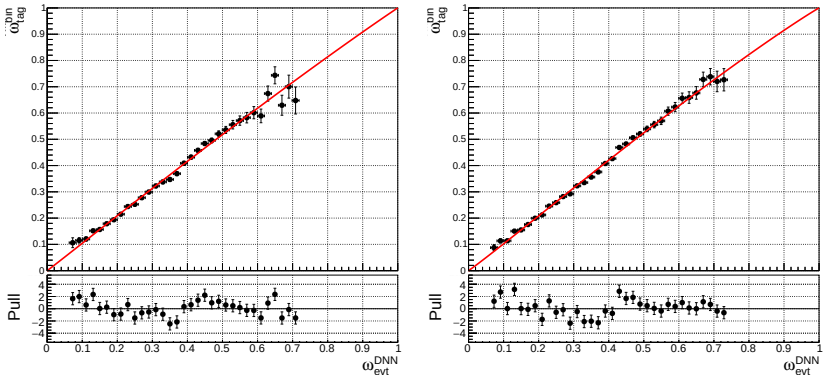
Table 3.7: Overview of the available tagging algorithms for each data set

Overlap	OS muon	OS electron	OS jet	SS
OS muon		×	×	✓
OS electron	×		×	✓
OS jet	×	×		✓
SS	✓	✓	✓	

Table 3.8: Overview of the available tagging algorithms for each data set

Category	ε_{tag} [%]	$\hat{\mathcal{D}}^2_{tag}$	$\hat{P}_{tag}$
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

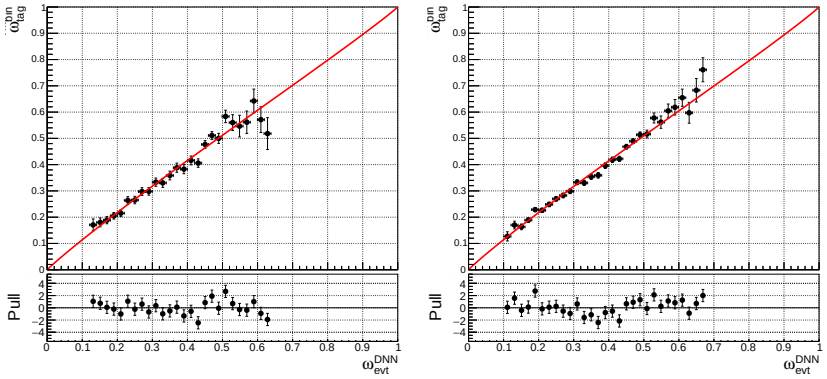
Table 3.9: Summary of tagging performance



(a) OS-muon calibration plot for *dataset_JpsiMuon* (2017)

(b) OS-muon calibration plot for *dataset_JpsiMuon* (2018)

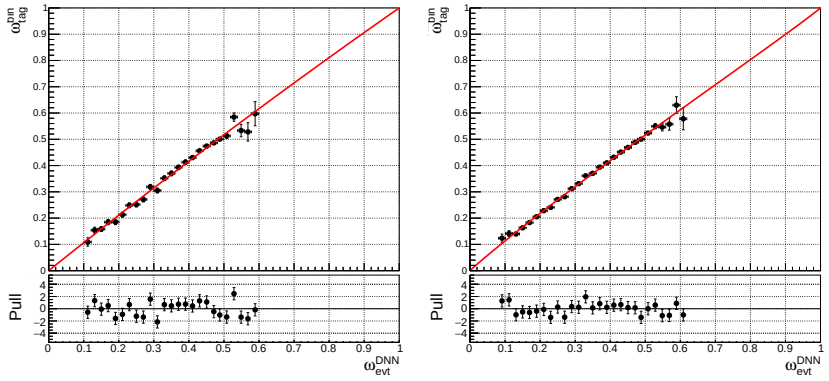
Figure 3.5: Result of the mistag probability calibration fit for the OS-muon tagger on $B^+ \to J/\psi K^+$ decays from data, for *dataset_JpsiMuon*. The solid red line represents the calibration fit to data (solid markers)



(a) OS-muon calibration plot for *dataset_JpsiTrkTrk* (2017)

(b) OS-muon calibration plot for *dataset_JpsiTrkTrk* (2018)

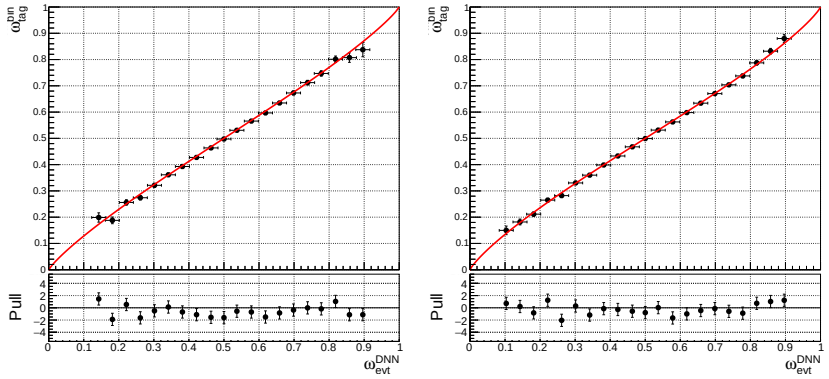
Figure 3.6: Result of the mistag probability calibration fit for the OS-muon tagger on $B^+ \to J/\psi K^+$ decays from data, for *dataset_JpsiTrkTrk*. The solid red line represents the calibration fit to data (solid markers)



(a) OS-electron calibration plot for *dataset_JpsiTrkTrk* (2017)

(b) OS-electron calibration plot for *dataset_JpsiTrkTrk* (2018)

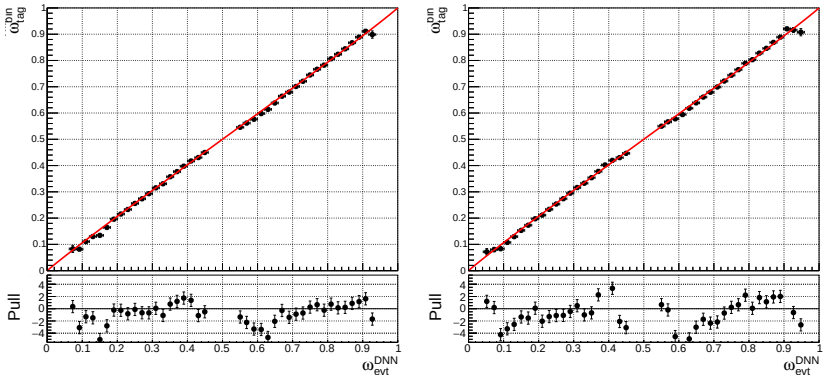
Figure 3.7: Result of the mistag probability calibration fit for the OS-electron tagger on $B^+ \to J/\psi K^+$ decays from data. The solid red line represents the calibration fit to data (solid markers)



(a) OS-jet calibration plot for *dataset_JpsiTrkTrk* (2017)

(b) OS-jet calibration plot for *dataset_JpsiTrkTrk* (2018)

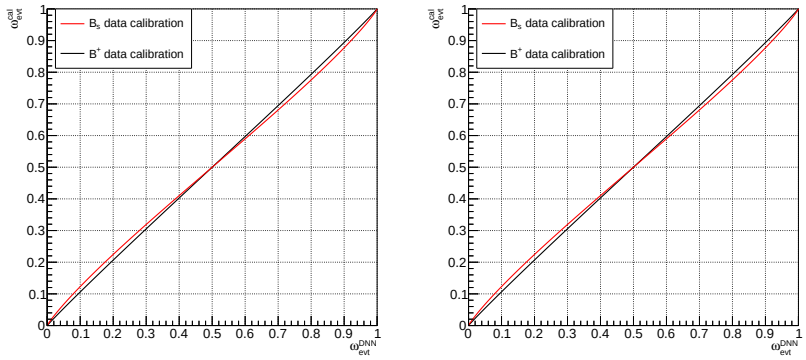
Figure 3.8: Result of the mistag probability calibration fit for the OS-jet tagger on $B^+ \to J/\psi K^+$ decays from data. The solid red line represents the calibration fit to data (solid markers)



(a) SS calibration plot for *dataset_JpsiTrkTrk* (2017)

(b) SS calibration plot for *dataset_JpsiTrkTrk* (2018)

Figure 3.9: Result of the mistag probability calibration fit for the SS tagger on $B^+ \to J/\psi K^+$ decays from data. The solid red line represents the calibration fit to data (solid markers)



(a) Comparison of B^+ and B^0_s calibration curves in 2017 data

(b) Comparison of B^+ and B^0_s calibration curves in 2018 data

Figure 3.10: Comparison of the calibration curves before (black) and after (red) the correction to account for the difference between $B^+ \to J/\psi\phi(1020)$. The red function is the one used as calibration for the analysis sample

3.4 Angular efficiency

3.4.1 Kernel Density Estimation

Kernel Density Estimation (KDE) is a non-parametric method to estimate the probability density function (PDF) of a random variable based on kernels used as weights. The method is described in detail in Silverman, 1982 [69].
If $X_1, ..., X_n$ are real observations from a probability density f, the kernel estimate f_n of f is defined by:

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \tag{3.7}$$

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 later 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 (Silverman rule [69], Sheather-Jones algorithm [70] etc.). 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:

$$h = 1.06\sigma n^{-1/5} \tag{3.8}$$

In this case, the σ is evaluated from the data sample.

Although the method guarantees a very good estimate of the required probability density function, its main downside is the computational complexity in the multivariate case [71].

In the CERN ROOT framework, used for this analysis, the method [72] works with an acceptable computational time only in the one-dimensional case. For this reason, we implement an in-house ROOT class for optimal performance in the multivariate case, by using the Fast Fourier Transform (FFT) [73].

3.4.2 KDE with FFT

We implement a KDE method for probability density function estimation with FFT convolution and mirrored boundary conditions. The method is applied for the purpose of estimating the angular efficiency function for the reconstruction of the $B_s^0 \to J/\psi\phi \to \mu^+\mu^- K^+ K^-$ decay.

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. 2.7.

In order to perform the discrete Fourier Transform, the data is distributed on a 3-D grid. The estimated function is then obtained by means of interpolation. The steps of the algorithm are as follows:

- Construct a three-dimensional histogram of the data
- Determine the bandwidth h by using either the Silverman rule (also referred to as the rule-of-thumb method [74]) or the Sheather-Jones algorithm [70]
- Apply mirrored boundary conditions
- Calculate the Fourier transforms of the data and of the Gaussian kernel centered at the middle of the histogram range
- Calculate the product of the two Fourier images
- Calculate the Inverse Fourier transform of the product
- Shiftback the resulting histogram to account for the shift introduced by the centered Gaussian kernel
- Return the resulting PDF on request

3.4.3 Obtaining the efficiency

The above algorithm is applied for both the Gen and Reco MC samples. Thus, as a result we have the following KDE approximated distributions:

$$\tilde{\mathrm{N}}_{\mathrm{reco}}(cos(\theta), cos(\psi), \phi) \tag{3.9}$$

$$\tilde{\mathrm{N}}_{\mathrm{gen}}(cos(\theta), cos(\psi), \phi) \tag{3.10}$$

To obtain the efficiency, we simply divide the two distributions:

$$\tilde{\epsilon}(cos(\theta), cos(\psi), \phi) = \frac{\tilde{\mathrm{N}}_{\mathrm{reco}}(cos(\theta), cos(\psi), \phi)}{\tilde{\mathrm{N}}_{\mathrm{gen}}(cos(\theta), cos(\psi), \phi)} \tag{3.11}$$

3.4.4 CERN ROOT implementation

The in-house class KDEFFTConv [75] is written for the implementation of this method in CERN ROOT [76]. There are several constructors of the class, allowing for use in one-, two- and three-dimensional analysis:

```
KDEFFTConv(const char* name, TH1* hist, int bw_option=0);

KDEFFTConv(const char* name, TH2* hist, int bw_option=0);

KDEFFTConv(const char* name, TH3* hist, int bw_option=0);
```

The constructor initializes the object and then calls the following method internally:

```
calcKDE_3D_mirrored(hx,hy,hz);
```

Here hx, hy and hz are the bandwidths for the corresponding dimensions. In the 1D and 2D cases, the `calcKDE\_1D\_mirrored(hx)` and the `calcKDE\_2D\_mirrored(hx,hy)` methods are called respectively. The bandwidths themselves are also calculated internally by the following methods:

```
BWSelector_SILVMN(hx,hy,hz)
```

when `bw\_option=0` (the default value), which evokes the Silverman method for calculating the bandwidth or:

```
BWSelector_SJ(hx,hy,hz)
```

when `bw\_option=1`, which invokes the Sheather-Jones algorithm for calculating the bandwidth.
In turn, `calcKDE\_3D\_mirrored()` performs the shift which corrects for the centered Gaussian kernel, by invoking:

```
fftshift(TH3* histo);
```

The resulting estimated distribution is extracted both as `RooDataHist` and `RooHistPdf` (data types from the `RooFit` statistical package of ROOT [77]) by calling:

```
cloneHistAndPdf(const char* histpdf_name, const char* histpdf_title,
    RooHistPdf** histpdf, const char* hist_name, const char*
    hist_title, RooDataHist** datahist, RooArgSet pdf_arg_set, Int_t
    order);
```

`histpdf` and `datahist` are the objects where the resulting estimate is stored while `pdf\_arg\_set` is the object where the variables of the distribution are specified.
In addition, the class offers a method for calculation of uncertainties. This is done by invoking:

```
calcKDE_3Dpbc_mirrored_average_error_and_histo(const char*
    av_histo_name, TH3D** average_and_error_kde, int nsamples, UInt_t
    bs_seed);
```

where the bootstrap method [78] is implemented.

3.5 Monte Carlo reweighting

In modern high-energy physics analysis, Monte Carlo (MC) generated samples of data are used to train discriminators which can distinguish signal from background events, in the search for new physics etc [79]. Generated distributions are however imperfect, which can be due to mismodelling of the detector or due to inaccuracies in the theoretical model calculation of cross sections and branching ratios. Therefore, to ensure the correctness of the results we obtain when using MC samples, a procedure has to be applied which corrects the MC and makes it as similar to experimental data as possible.

Reweighting MC consists of applying scale factors (weights) to the MC events in order to minimize the differences with experimental data. In what follows, we present a method based on multivariate analysis, using Boosted Decision Trees (BDT) [80], which we implement using the `CERN ROOT TMVA` package [81].

Before moving on to the reweighting itself, as a first step the experimental data have to be prepared. This consists of subtracting the background and preparing a sample with only signal events.

3.5.1 Background subtraction

Several methods are used for background subtraction, including sPlot [82] and sideband subtraction [83]. Here we present an alternative method, called the fit-and-sample method. First, a fit is performed on the B_s^0 invariant mass distribution $m(B_s^0)$. The fitting function is a sum of a Johnson distribution [84] describing the peak, and an exponential function, describing the background (Eq. 3.13). The Johnson distribution is modeled using a `RooJohnson` object while the exponential with a `RooExponential` object. The Johnson distribution is a four-parameter function given as:

$$f_{Johnson}(x) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+\left(\frac{x-\xi}{\lambda}\right)^2}} exp\left[-\frac{1}{2}\left(\gamma + \delta sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right)^2\right] \tag{3.12}$$

$$f_{fit}(x) = f_{Johnson}(x) + Nexp(cx) \tag{3.13}$$

After these functions have been declared, the fit is performed in a straightforward manner in `ROOT`, by:

```
RooFitResult *fitRes = massPdf.fitTo(*dataSet, Extended(),
    PrintLevel(-1), Save());
```

In the code, by `massPdf` we refer to the fitting function $f_{fit}(x)$, with components `sgnPdf` and `bkgPdf`. Once the fit is done, we can extract the separate components of the fit function, that is, the signal and the background. In Fig. 3.11 the B_s^0 mass distribution is shown along with the fit and its separate components.

Once the fitting function is obtained, the following steps are executed to create the background subtracted sample:

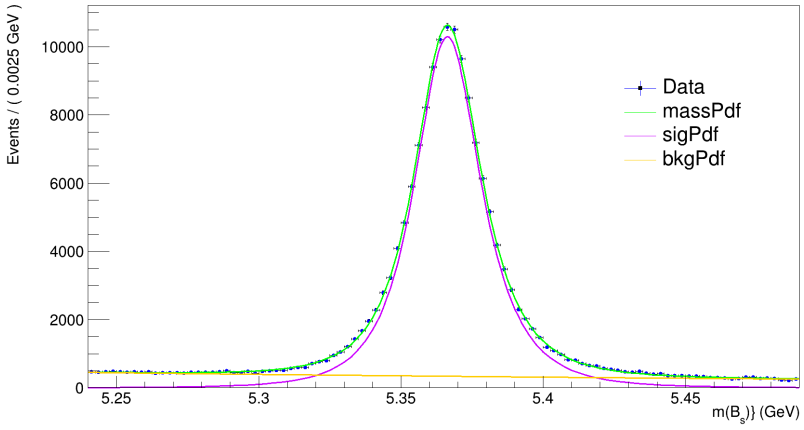


Figure 3.11: B_s^0 invariant mass distribution fit

- Loop through each event in the data sample
- Calculate the following integrals for `sgnpdf` and `bkgpdf`:

$$F(sgn) = \int_{m^i_{B_s}-\epsilon}^{m^i_{B_s}+\epsilon} \text{sgnPdf}(m_{B_s})dm \tag{3.14}$$

$$F(bkg) = \int_{m^i_{B_s}-\epsilon}^{m^i_{B_s}+\epsilon} \text{bkgPdf}(m_{B_s})dm \tag{3.15}$$

  where ϵ is a predefined parameter (e.g. $\epsilon = 0.005$)

- Calculate the ratios:

$$p(sgn) = \frac{F(sgn)}{F(sgn)+F(bkg)} \tag{3.16}$$

$$p(bkg) = \frac{F(bkg)}{F(sgn)+F(bkg)} \tag{3.17}$$

- Test if

$$p(sgn) + p(bkg) = 1 \tag{3.18}$$

- Generate a random number n_{rand} in the interval $[0, 1]$
- In case of $n_{rand} < p(bkg)$ classify the event as "background", otherwise classify it as "signal"
- Iterate over the next event

The B_s^0 mass distribution after background subtraction is shown in Fig. 3.12.

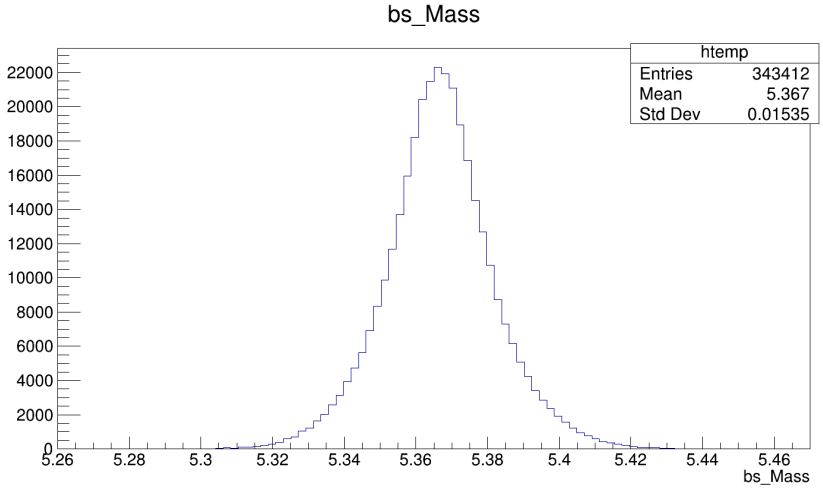


Figure 3.12: B_s^0 invariant mass distribution after background subtraction

3.5.2 MC-Data classifier

Having subtracted the background from the experimental data samples, we can perform a comparison between the MC and data distributions in terms of the different variables characterizing the events. In Fig. 3.14 the comparisons of some of the distributions are shown.

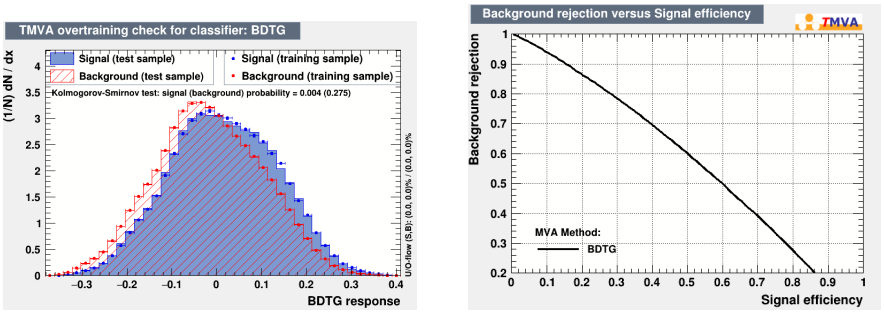


Figure 3.13: Classifier performance characteristics

The reweighting procedure consists of building a classifier that can distinguish MC and data events. The implementation of the classifier is done using the `CERN ROOT TMVA` package.

The first step is to build a BDT. The background separated data are given to the BDT as the input signal while the MC data are given as the input background. The tree is then trained (grown) to be able to classify each event as either signal or background. The classifier performance characteristics are shown in Fig. 3.13.

The MC sample that is to be reweighted is then processed event by event. Each event is propagated through the tree until it ends up on an end leaf. Each end leaf is labeled with a number from the interval [-1, 1].
The following transformation is applied on the label:

$$label \to \frac{label + 1}{2} \tag{3.19}$$

This accomplishes the interval mapping $[-1, 1] \to [0, 1]$, thus giving the label a meaning of probability, namely, $p(Data)$ quantifies how much an event is "data-like".

From a set of events that exist in a given multidimensional bin (representing a region of parameter space), the probability that a given event is sampled from data is simply:

$$p(Data) = \frac{N_{Data}}{N_{Data} + N_{MC}} \tag{3.20}$$

The weight that we need to apply to the same bin in the multidimensional MC histogram is:

$$w(i) = \frac{N_{Data}}{N_{MC}} \tag{3.21}$$

Combining Eq. 3.20 and Eq. 3.21, we obtain the following expression for the per-event weight:

$$w(i) = \frac{1 + p(Data)}{1 - p(Data)} \tag{3.22}$$

After calculating the per-event weights as in Eq. 3.22, the MC distributions follow much more closely the background subtracted experimental data distributions, as is shown in Fig. 3.15.

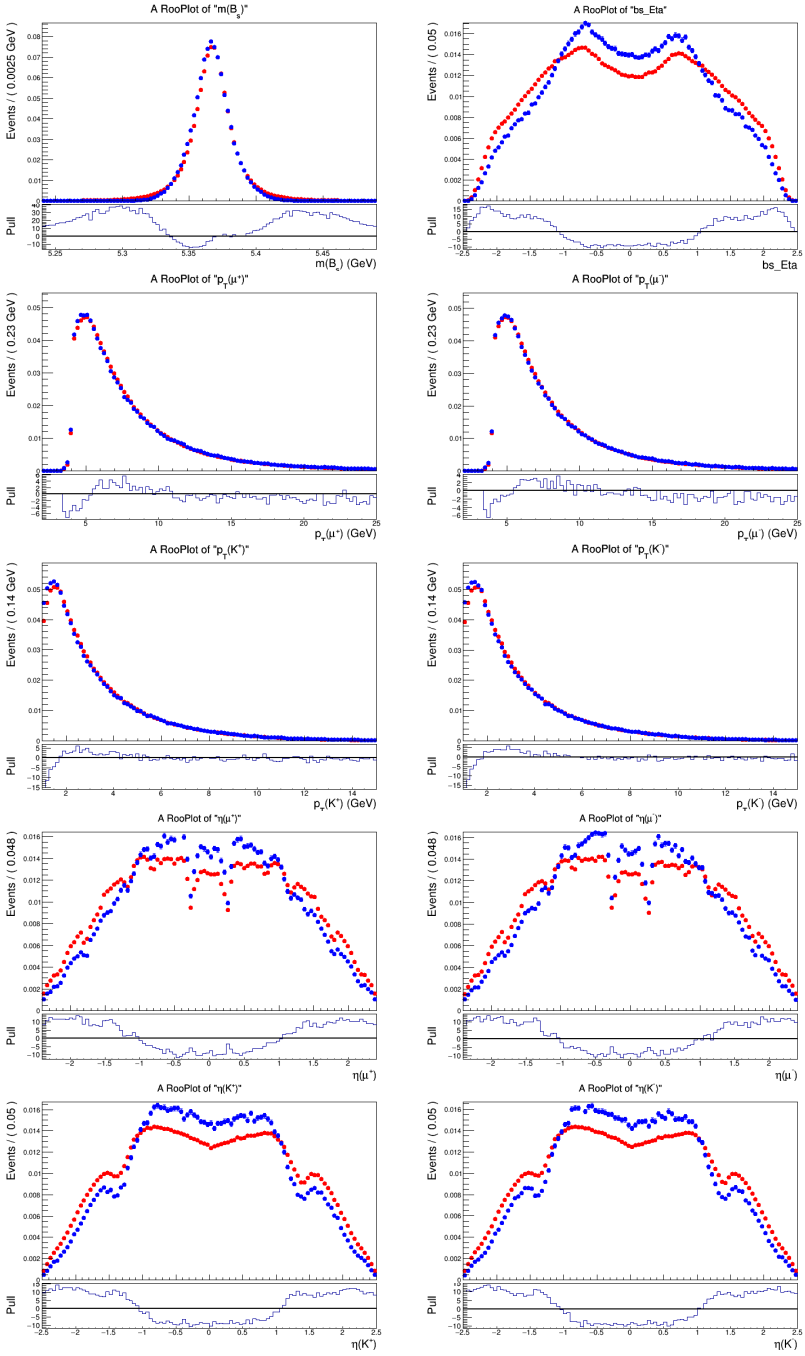


Figure 3.14: MC-Data comparisons before reweighting

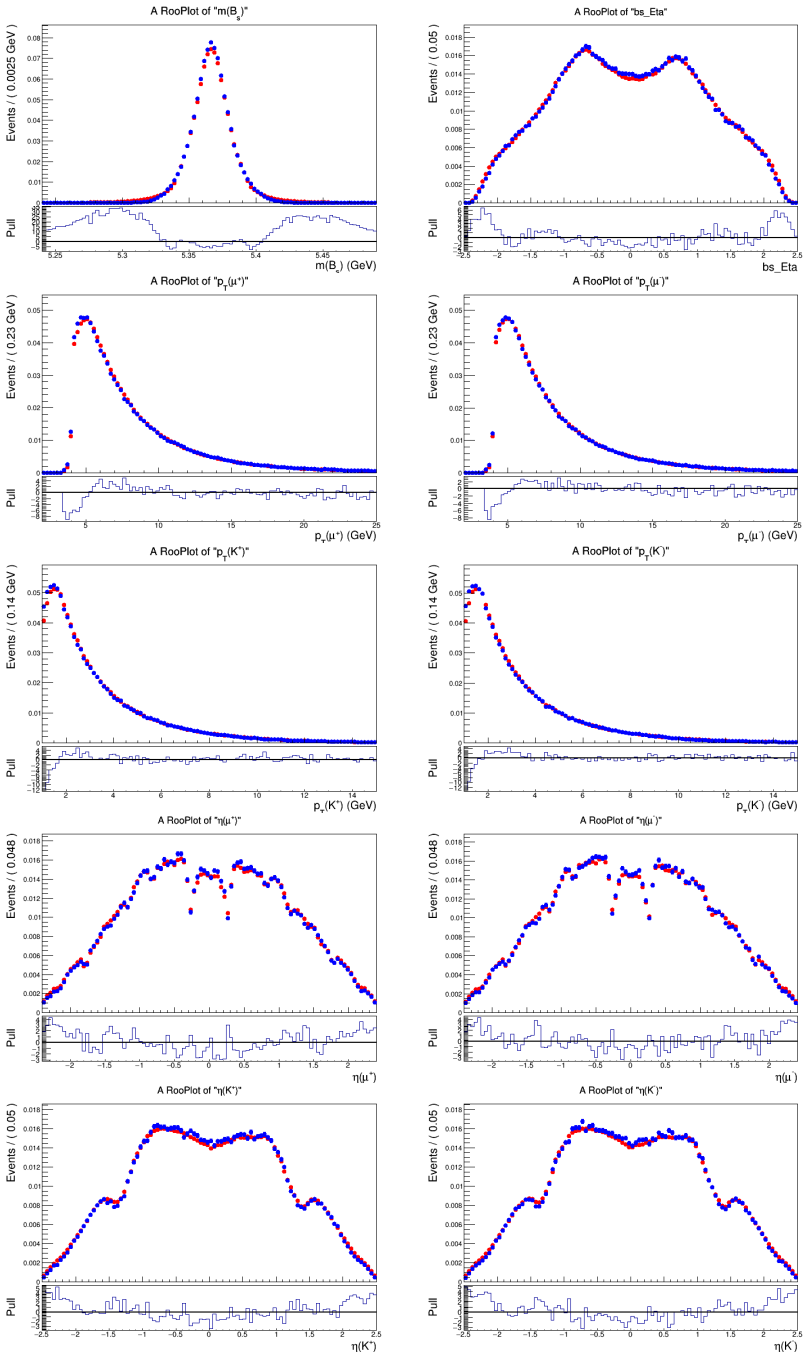


Figure 3.15: MC-Data comparisons after reweighting

3.6 Λ_b background contribution

The final state of the decay

$$\Lambda_b \to J/\psi K p \tag{3.23}$$

differs from the final state of the decay

$$B_s \to J/\psi \phi \tag{3.24}$$

only by one of two charged tracks (considering that ϕ in (3.24) is reconstructed from a K^+K^- pair). The CMS experiment does not offer hadronic particle identification (PID) and the mass of the tracks in the data events is assigned by hand during data analysis. Therefore, misidentifying a proton track as a kaon (assigning it a kaon mass) during the selection and reconstruction algorithm leads to a false reconstruction of a B_s instead of a Λ_b. The mass of such particles will appear as background in the final mass distribution of the B_s candidate events.

Herein we describe and apply a procedure for the quantitative estimation of this background contribution and the subsequent evaluation of its statistical significance. In order to determine the shape of the Λ_b mass distribution that is anticipated to be found in data, we use a sample of MC generated Λ_b events.

On the other hand, the full 2017 + 2018 data sample is used. The B_s mass distribution is reflected in the Λ_b mass region by randomly assigning a proton mass to one of the two charged tracks that were initially identified as kaons. A fit model is constructed, including a term to represent the Λ_b contribution with a shape fixed from MC. The yield of this contribution is however left free to float. Determining the yield by performing the fit then translates to an estimation of the Λ_b background contribution.

3.6.1 MC sample

The MC generated events are reconstructed using the B_s reconstruction algorithm. In Fig. 3.16, the mass histogram is shown in the B_s mass range. The next step is to obtain the mass histogram in the Λ_b mass range. In order to do that, for each event, one of the tracks is chosen randomly on the basis of the output of a Random Number Generator (RNG) [85]. The proton mass is assigned to the track and the invariant mass is recomputed. Fig. 3.17 shows the obtained mass distribution.

The Λ_b peak is best fitted by a Crystal Ball (CB) distribution [86], as is shown in Fig. 3.18. The shape of the obtained CB distribution is saved for use in the next step.

3.6.2 Data sample

The data sample constitutes the full 2017+2018 data after applying the analysis selection and reconstruction algorithm.

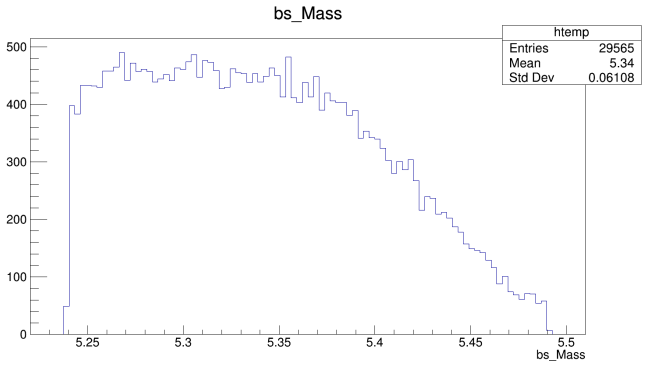


Figure 3.16: MC Λ_b particles reconstructed using the B_s reconstruction algorithm

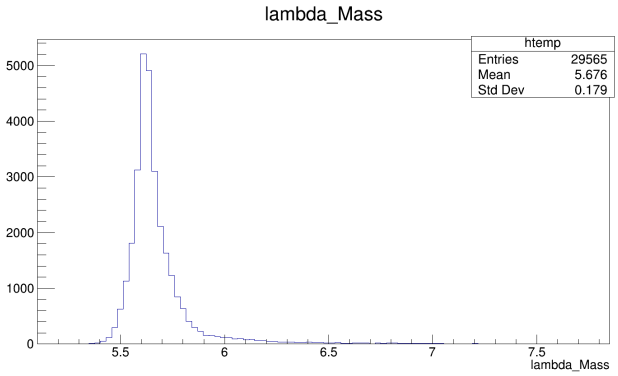


Figure 3.17: MC Λ_b mass distribution histogram

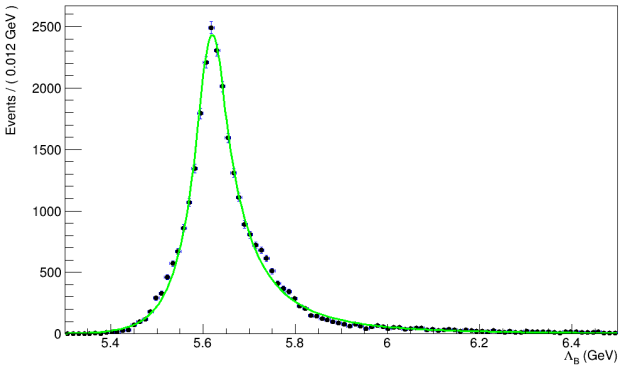


Figure 3.18: MC Λ_b mass histogram fit with a Crystal Ball distribution

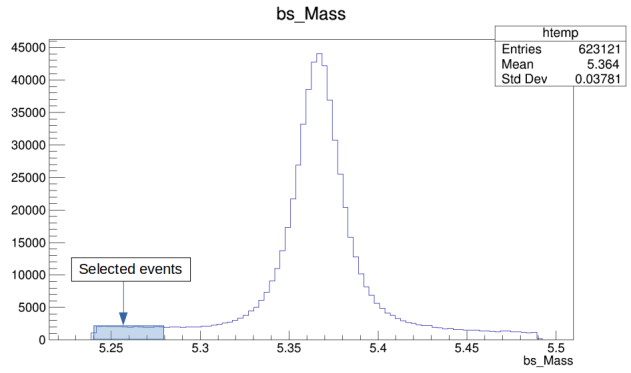


Figure 3.19: B_s candidates mass distribution in data

The mass histogram of the candidate events is shown in Fig. 3.19. The sideband indicated by the shaded region is selected to be reflected in the Λ_b mass range by applying the same random assignment of proton mass to one of the two charged tracks as in the MC case. The range of the sideband is the interval $[5.24, 5.28]$ GeV. The reason for selecting this sideband is because the contribution from Λ_b's is more significant in this region as no signal B_s events are expected to be found there. The only other contribution would be the B_s combinatorial background.

The obtained distribution after applying the cut and performing the random proton mass assignment is shown in Fig. 3.20

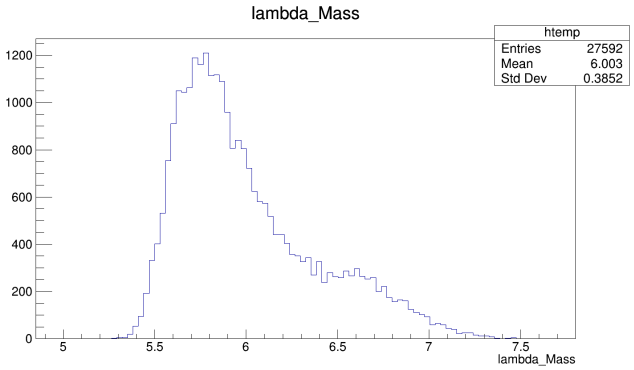


Figure 3.20: Sideband B_s candidates reflected in the Λ_b mass range

3.6.3 Simultaneous fit

We use a simultaneous fit approach. The sideband region in the B_s mass range and its reflection in the Λ_b mass range are fitted simultaneously. The corre-

sponding models used to perform the fit are, respectively:

$$P_{m_{B_s}} = (1 - f_{\Lambda_b})P_{comb}(m_{B_s}) + f_{\Lambda_b}P_{\Lambda_b}(m_{B_s}) \tag{3.25}$$

$$P_{m_{\Lambda_b}} = (1 - \frac{c_{mc}}{c_{data}}f_{\Lambda_b})P_{comb}(m_{\Lambda_b}) + \frac{c_{mc}}{c_{data}}f_{\Lambda_b}P_{\Lambda_b}(m_{\Lambda_b}) \tag{3.26}$$

The terms included in the fit model have the following significance:

- $P_{comb}(m_{B_s})$ - exponential, models the combinatorial background in the sideband region
- $P_{\Lambda_b}(m_{B_s})$ - exponential, models the Λ_b contribution in the sideband region (fixed in shape from the MC fit in the sideband)
- $P_{comb}(m_{\Lambda_b})$ - a sum of two Johnson distributions, models the combinatorial background reflected in the Λ_b mass range
- $P_{\Lambda_b}(m_{\Lambda_b})$ - CB distribution, models the Λ_b peak (fixed in shape from the MC fit)
- f_{Λ_b} - the parameter of interest, represents the Λ_b yield; it is a common parameter of $P_{m_{B_s}}$ and $P_{m_{\Lambda_b}}$

In addition, c_{mc} and c_{data} are introduced to account for the fact that the fit in the Λ_b mass range is not performed over the whole range, but only up to $m = 6.5$ GeV. They are defined as:

$$c_{mc} = \frac{N_{\Lambda_b}\left[(min_{m_{\Lambda_b}fit} < m_{\Lambda_b} < max_{m_{\Lambda_b}fit})\ AND\ (SB_L < m_{B_s} < SB_R)\right]}{N_{\Lambda_b}\left[SB_L < m_{B_s} < SB_R\right]} \tag{3.27}$$

$$c_{data} = \frac{N_{tot}\left[(min_{m_{\Lambda_b}fit} < m_{\Lambda_b} < max_{m_{\Lambda_b}fit})\ AND\ (SB_L < m_{B_s} < SB_R)\right]}{N_{tot}\left[SB_L < m_{B_s} < SB_R\right]} \tag{3.28}$$

with c_{mc} computed in the MC sample while c_{data} in the data sample.

Before the fit is performed, the data sample is randomly split in two equal datasets. Each one of the two functions defined above is fitted to one of the datasets. The fit is performed using the `CERN ROOT RooFit` library's `RooSimultaneous` class [87]. The result of the fit is shown in Fig. 3.21

In the procedure described so far, we have utilized the output of a RNG in both the random selection of a track to assign the proton mass to and the splitting of the data sample in two equal in size datasets. To avoid any bias introduced by the RNG, we perform the fit on multiple samples and calculate the Λ_b yield in each iteration, as is described in the next section.

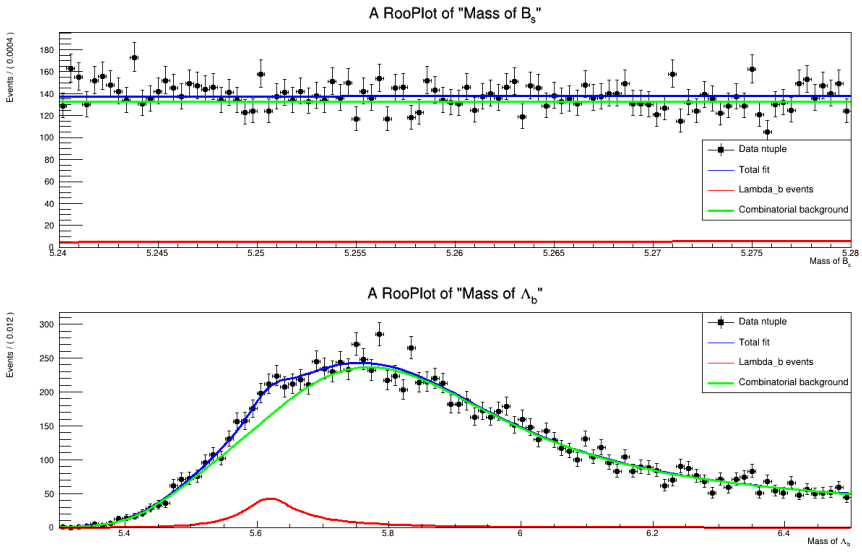


Figure 3.21: Simultaneous fit result

3.6.4 Λ_b yield

The simultaneous fit gives us a value for the f_{Λ_b} parameter in the fit model, representing the yield. In order to extract the estimated total number of lambdas in the sideband, we have to consider the total number of entries in the sideband:

$$N_{\Lambda_b}(SB) = N_{tot}(SB) \times \frac{c_{mc}}{c_{data}} \times f_{\Lambda_b} \tag{3.29}$$

Finally, we can extrapolate to the full mass range, having calculated the ratio of the number of Λ_b events in the SB to the number in the full mass range in the MC sample:

$$N_{\Lambda_b}(full) = \frac{N_{\Lambda_b}(MC\_full)}{N_{\Lambda_b}(MC\_SB)} \times N_{\Lambda_b}(SB) \tag{3.30}$$

As mentioned in the previous section, the fit is performed on multiple samples, with $N_{\Lambda_b}(full)$ being calculated each time. The histogram of $N_{\Lambda_b}(full)$ over 800 iterations is shown in Fig. 3.22, along with a Gaussian fit.

We estimate the final result for $N_{\Lambda_b}(full)$ as the mean of this Gaussian fit:

$$N_{\Lambda_b} = 5982.5 \pm 40.8 \tag{3.31}$$

This constitutes $\approx 1\%$ of the total number of events in the data sample and $\approx 5\%$ of the total background. The impact of this background contribution in the broader context of the analysis is discussed in the next section.

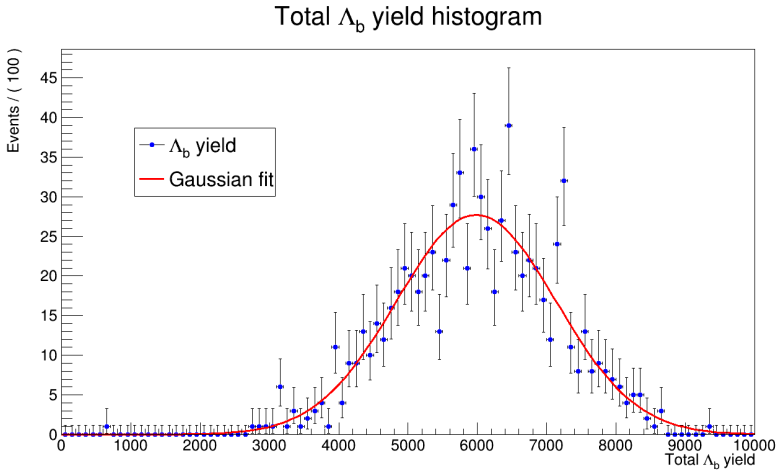


Figure 3.22: The total Λ_b yield calculated over 800 iterations

3.6.5 Impact on the analysis

Having obtained an estimation of the Λ_b contribution to the total B_s background, the systematic uncertainty originating from not including the Λ_b explicitly in the fit model was evaluated. The results are shown in Tab. 3.10.

Table 3.10: Impact of the Λ_b background on the final results

Parameter	Stat. Uncer.	Λ_b^0 syst
ϕ_s[mrad]	21	-
$\Delta\Gamma_s$[ps^{-1}]	0.0039	-
Γ_s[ps^{-1}]	0.0015	0.0003
Δm_s[$\hbar$ps^{-1}]	0.038	-
$\lvert\lambda\rvert$	0.017	-
$\lvert A_0\rvert^2$	0.0016	0.0006
$\lvert A_\perp\rvert^2$	0.0022	0.0005
$\lvert A_S\rvert^2$	0.0035	-
$\delta_\parallel$[rad]	0.045	0.015
$\delta_\perp$[rad]	0.081	0.009
$\delta_{S\perp}$[perp]	0.17	-

We can conclude that the effect from this background is too small to produce a measurable effect on the final results and does not require an explicit modeling.

3.7 Fit model

The fit model consists of probability density functions (pdf) describing the B_s^0 signal (P_{sig}), the main or combinatorial background (P_{bkg}) and the background arising from the decay $B^0 \to J/\psi K^*(892)0$ when a pion is misidentified as a kaon (P_{bkgB^0}). The pdfs are formed as a product of functions that model the invariant mass distributions and the time-dependent decay rates:

$$P = f_{sig}P_{sig} + f_{bkg}P_{bkg} + f_{bkgB^0}P_{bkgB^0} \tag{3.32}$$

The individual terms can be written explicitly as:

$$P_{sig} = \varepsilon(\Theta)\left[\tilde{f}(\Theta, ct|\alpha, \xi_{tag}, \omega_{tag}) \otimes G(ct,\ \sigma_{ct})P_{sig}(m_{B_s^0})P_{sig}(\sigma_{ct})P_{sig}(\omega_{tag})\right], \tag{3.33}$$

$$P_{bkg} = \left[P_{bkg}(ct) \otimes G(ct,\ \sigma_{ct})\right] P_{bkg}(\Theta)P_{bkg}(m_{B_s^0})P_{bkg}(ct)P_{bkg}(\omega_{tag}), \tag{3.34}$$

$$P_{bkgB^0} = \left[P_{bkg}(ct) \otimes G(ct,\ \sigma_{ct})\right] P_{bkgB^0}(\Theta)P_{bkgB^0}(m_{B_s^0})P_{bkgB^0}(ct)P_{bkgB^0}(\omega_{tag}), \tag{3.35}$$

where $\Theta(\cos\theta_T, \cos\psi_T, \varphi_T)$ are the three angles of the transversity basis (as defined in Fig. 2.7), $\varepsilon(\Theta)$ is the angular efficiency function calculated in Section 3.4, and $G(ct,\ \sigma_{ct})$ is the ct resolution function. The differential decay rate $\tilde{f}(\Theta, ct|\alpha, \xi_{tag}, \omega_{tag})$ (first introduced in Eq. 2.117), has been augmented to include tagging information via the tagging decision variable ξ_{tag} (defined in Eq. 3.1) and the mistag probability (Eq. 3.3). The pdfs used to parametrize the signal and background distribution in Eq. 3.32 are given in Tab. 3.11. The effective fractions of the components of the fit model are given by:

$$f_l = \frac{N_l}{N_{tot}}, \quad \text{with } l = (\text{sig, bkg, bkg}B^0) \tag{3.36}$$

Variable	Signal	Combinatorial 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	Same as signal

Table 3.11: Summary of the pdfs used to model the signal and background distributions in the fit model

An unbinned maximum-likelihood (UML) fit is performed on the model outlined above in order to extract the physics parameters of interest. The obtained results, representing the final results of this analysis, are presented in the next section.

3.8 Results and conclusions

The final results of this CP violation analysis in $B_s^0 \to J/\psi\phi$ decays are reported in Tab. 3.12. The obtained value for the CP-violating phase ϕ_s is:

$$\phi_s = -73 \pm 23(\text{stat}) \pm 7(\text{syst}) \text{ mrad} \tag{3.37}$$

It is consistent (within 1.5 standard deviations) with the SM prediction of $\phi_s \approx -37 \pm 1$ mrad. This value is different from zero by 3.1 standard deviations, indicating strong evidence for CPV in decay/mixing interference in $B_s^0 \to J/\psi\phi$ decays.

The decay width difference between the two B_s^0 mass eigenstates is measured to be:

$$\Delta\Gamma_s = 0.0761 \pm 0.0043(\text{stat}) \pm 0.0019(\text{syst})\text{ps}^{-1}, \tag{3.38}$$

again consistent with the latest SM theoretical calculation of 0.091 ± 0.031 ps^{-1}. Additionally, both of these measurements are consistent with the latest world-average values:

$$\phi_s^{w.a.} = -49 \pm 19 \text{ mrad} \tag{3.39}$$

$$\Delta\Gamma_s^{w.a.} = 0.084 \pm 0.005 \text{ ps}^{-1} \tag{3.40}$$

The average decay width value is obtained to be $\Gamma_s = 0.6613 \pm 0.0015(\text{stat}) \pm 0.0028(\text{syst})$ ps^{-1}, consistent with the world-average value. The CP observable $|\lambda| = 1.011 \pm 0.014(\text{stat}) \pm 0.012(\text{syst})$ is consistent with both the world average value and with no direct CP violation in $B_s^0 \to J/\psi\phi$.

The unceratinty in some of the measured parameters, such as: ϕ_s, $\Delta\Gamma_s$, Δm_s, $|\lambda|$, $|A_S|^2$, $\delta_\parallel$, $\delta_\perp$, and $\delta_{S\perp}$ is dominated by the statistical component. In the others: Γ_s, $|A_0|^2$, and $|A_\perp|^2$ the uncertainty is dominted by the systmatics. Tab. 3.13 provides a summary of all the results in comparison to the world-average values and the theoretical predicions, where available.

The results are combined with the Run-I CMS measurement at 8 TeV [56] using the Best Linear Unbiased Estimate (BLUE) method [88, 89], to produce the values $\phi_s = -74 \pm 23$ mrad^{-1} and $\Delta\Gamma_s = 0.0780 \pm 0.0045$ ps^{-1}, both consistent with both the SM and the world-average values. It is important to be noted, that the measured value of ϕ_s is different from zero by 3.1 standard deviations. This result constitutes the first evidence for CPV in decay/mixing interference in the B_s^0 system.

This analysis resulted in a paper [90], expected to be published in the Physical Review Letters (PRL) journal.

Parameter	Fit value	Stat. uncer.	Syst. uncer.
ϕ_s [mrad]	-73	± 23	± 7
$\Delta\Gamma_s[\mathrm{ps}^{-1}]$	0.0761	± 0.0043	± 0.0019
$\Gamma_s[\mathrm{ps}^{-1}]$	0.6613	± 0.0015	± 0.0028
$\Delta m_s[\hbar\mathrm{ps}^{-1}]$	17.757	± 0.035	± 0.017
$\lvert\lambda\rvert$	1.011	± 0.014	± 0.012
$\lvert A_0\rvert^2$	0.5300	$^{+0.0016}_{-0.0014}$	± 0.0044
$\lvert A_\perp\rvert^2$	0.2409	± 0.0021	± 0.0030
$\lvert A_S\rvert^2$	0.0067	± 0.0033	± 0.0009
$\delta_\parallel$	3.145	± 0.074	± 0.025
$\delta_\perp$	2.931	± 0.089	± 0.050
$\delta_{S\perp}$	0.48	± 0.15	± 0.05

Table 3.12: Summary of the analysis results. The values of the measured physics parameters are given alongside theit statistical and systematic uncertainties

Parameter	Measured value	World-average value	Theory prediction
ϕ_s [mrad]	-73 ± 24	-49 ± 19	-37 ± 1
$\Delta\Gamma_s[\mathrm{ps}^{-1}]$	0.0761 ± 0.0047	0.084 ± 0.005	0.091 ± 0.013
$\Gamma_s[\mathrm{ps}^{-1}]$	0.6613 ± 0.0032	0.6573 ± 0.0023	–
$\Delta m_s[\hbar\mathrm{ps}^{-1}]$	17.757 ± 0.039	17.765 ± 0.006	18.77 ± 0.86
$\lvert\lambda\rvert$	1.011 ± 0.018	1.001 ± 0.018	1
$\lvert A_0\rvert^2$	0.5300 ± 0.0047	0.520 ± 0.003	–
$\lvert A_\perp\rvert^2$	0.2409 ± 0.0037	0.253 ± 0.006	–
$\lvert A_S\rvert^2$	0.0067 ± 0.0034	0.030 ± 0.005	–
$\delta_\parallel$	3.145 ± 0.078	3.18 ± 0.06	–
$\delta_\perp$	2.931 ± 0.102	3.08 ± 0.12	–
$\delta_{S\perp}$	0.48 ± 0.16	0.23 ± 0.05	–

Table 3.13: Comparison of the results of this measurement with world-average values and theory predictions. For measured values, the quoted uncertainty is the quadratic sum of the statistical and systematic components

Chapter 4

Machine Learning based monitoring tool

As mentioned in section 1.3.4, during detector operation, the dark current that each individual RPC chamber in the system draws is one of the most important monitorables, as this quantity is directly correlated to the possibility of operational failure events known as HV trips.

One of the works included in this dissertation is an automated tool which utilizes Machine Learning based techniques for current monitoring with a built-in logic for anomaly detection and warning of the end-user.

In this chapter I describe the automated tool, starting with the physics motivation behind the modeling approach, the software implementation and finally its performance in real-life conditions.

4.1 RPC currents modeling

The dark current of an RPC chamber cannot be analytically modeled and is clearly a quantity influenced by a multitude of parameters. These parameters may exhibit complex, non-linear inter-dependencies, making traditional modeling techniques insufficient. Consequently, an alternative Machine Learning (ML) based approach becomes essential, as it can effectively capture intricate patterns and relationships within the data. In the next section, an account of the two types of ML methods we utilize to model the RPC currents is provided.

4.2 ML approaches

Two types of ML approaches are used: Generalized Linear Models (GLM) [91] and Autoencoders [92]. In the GLM case, a set of parameters such as environmental conditions, LHC parameters and detector working points are used to characterize the behavior of the current. In the autoencoder case, the full set of

the RPC HV system currents is used as an input and the autoencoder network is trained to reproduce these inputs onto the output neurons. Both approaches show very good predictive capabilities that are the basis for the monitoring tool. All the developed tools are integrated in a framework that can be easily accessed and controlled by a specially developed Web User Interface that allows the end-users to work with the monitoring tool in a simple manner. It is being deployed for use during the CERN LHC Run-3 data-taking period.

4.2.1 Generalized Linear Model

The GLM depicted in Fig. 4.1 is a generalization of a simple linear regression used to model the current as a function of the following sets of parameters:

- Environmental conditions: temperature (T), relative humidity (RH) and pressure (P)
- LHC parameters: instantaneous luminosity (L) and integrated luminosity (L)
- Applied HV
- Combined terms: $L \times exp(HV/P)$ and $(\Sigma HV/P)\delta t$, where δt is the length of the time period with no luminosity

The L term replaces the term used in the initially proposed model. The improvement is inspired by [4]. The first combined term is to account for the exponential increase of gas multiplication with the raising of HV while the second one is to account for the chamber relaxation and the drop of the current baseline during cosmic data taking, when there is no beam luminosity and the chambers are at their working point. All the remaining terms and the motivation for including them are discussed in [3].

4.2.2 Autoencoder

In contrast to the GLM approach, where we use detailed knowledge for the physical processes taking place in a particular type of detector in order to build the ML model, in this section we take a more general approach, namely develop an ML model based on cross-correlation between different detector modules, thus applicable for detector systems consisting of a large number of RPC chambers. We develop an ML algorithm based on an autoencoder model. Autoencoders are neural networks that are trained to encode the input into a number of neurons that is lower than the number of inputs themselves and then decode that same information onto the output layer (Fig. 4.2). During the learning stage, the autoencoder is supposed to learn the collective behavior of all the RPC chambers. Such an autoencoder could be used later on to spot an anomalous behavior of a single or a small subset of RPC chambers. In this work, the set of RPC currents at a given moment in time is given as an input to the autoencoder and the network is trained to reproduce them on its output layer. The number

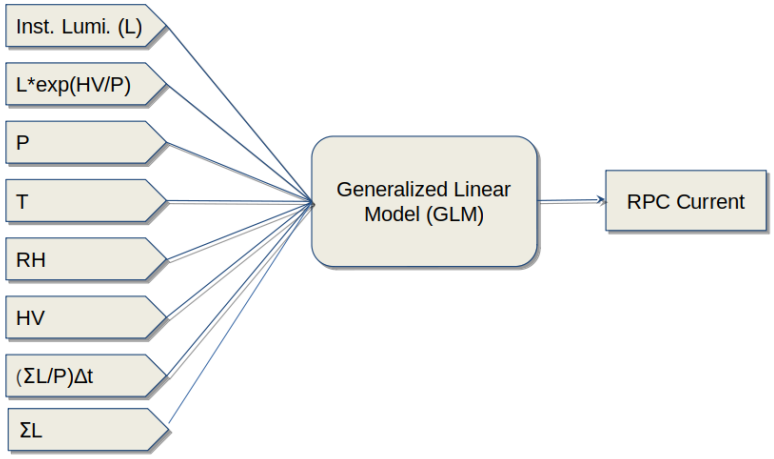


Figure 4.1: GLM topology

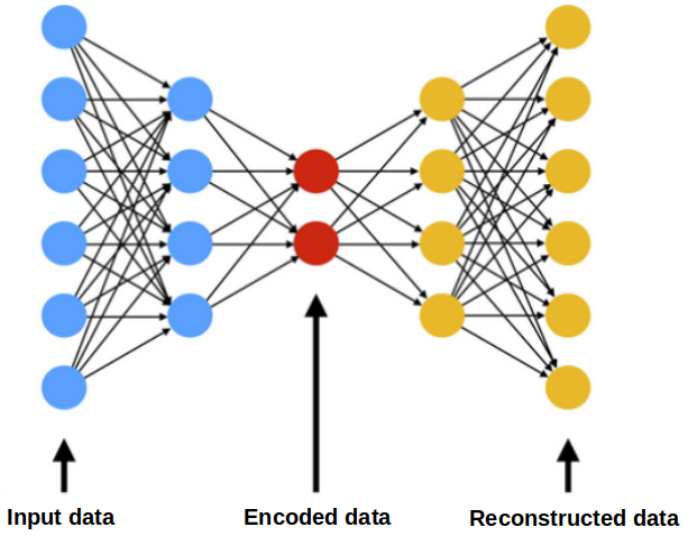


Figure 4.2: Illustration of autoencoder topology

of input and output neurons is 773, which corresponds to the number of HV channels in the RPC system. The hidden layers count respectively: 512, 128, 64, 128 and 512 neurons.

4.2.3 Hybrid network

As discussed above, GLM describes individual RPC chamber behavior while the autoencoder describes collective correlations of the whole system. In order to use their best qualities, we combine the two approaches into a model, referred to as a hybrid network. In this model, a set of GLM equal in size to the number of HV channels provide as output the currents for a given moment in time. These currents are then used as inputs for an autoencoder, as shown in Fig. 4.3. The hybrid network is tested in a distant prediction scenario, where the end of the training period is separated in time (e.g. 1 year) from the beginning of the prediction period. Its performance in such a scenario (Section 7) shows that it can be used as indication for current values that we could expect on a system level for some specified conditions (e.g. the luminosity of the High-Luminosity LHC).

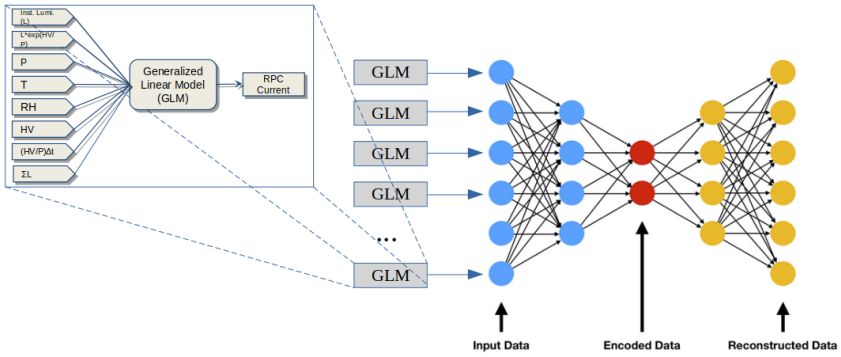


Figure 4.3: Illustration of hybrid network topology

4.3 Monitoring tool

The accurate predictions of the currents performed by both the GLM and autoencoder can be used to detect anomalies in the RPC detector current performance. The implemented tool follows the workflow presented in the flowchart in Fig. 4.4. Raw data coming from the CMS non-physics event bus, referred to as online condition data, are written in the database copy. For each point in time for which data is available, the tool performs comparisons between the measured and predicted RPC currents. If differences higher than some predetermined threshold values are detected for a given HV channel, a flag is raised and the case of that particular channel is followed.

There are two thresholds, the lower one inducing a warning and the higher one inducing an error. After a specified number of points in time, the running average of the differences is calculated and if this average exceeds the thresholds,

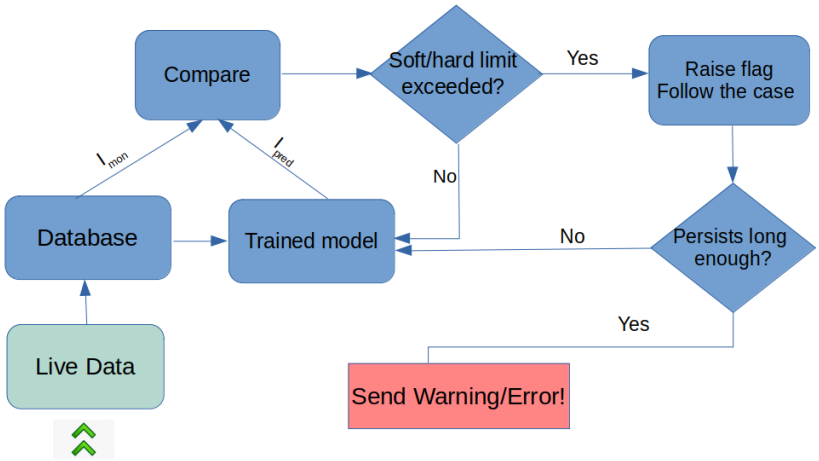


Figure 4.4: Flowchart illustrating the decision logic of the current monitoring tool

a warning or an error is sent to the end-users. This allows for the detection of problematic HV channels before they result in an HV channel trip.

4.4 Software implementation

The monitoring tool is programmed in Python. Tensorflow is used for the implementation of ML. The software is conceptualized and implemented with modularity in mind (Fig. 4.5). All modules communicate back-and-forth with a database. The "Trainer Module" reads the training data from a table and after performing the training, writes back the ML model parameters in another database table. The "Estimator Module" loads the models and performs predictions, which are also stored into the database. Finally, the "Notifications Manager" searches for anomalies in the current values, as described in the previous section and provides notifications.

4.5 Performance results

ML model performance validation is done for three different training scenarios:

- Short-term training (ST), with data from May to September 2018. Such models are able to spot a rapid increase in the RPC currents
- Mid-term training (MT), with data from July 2017 to July 2018, appropriate for describing the seasonal behavior of the currents
- Long-term training (LT), with data from May 2016 to July 2018, appropriate for modeling the overall RPC currents evolution

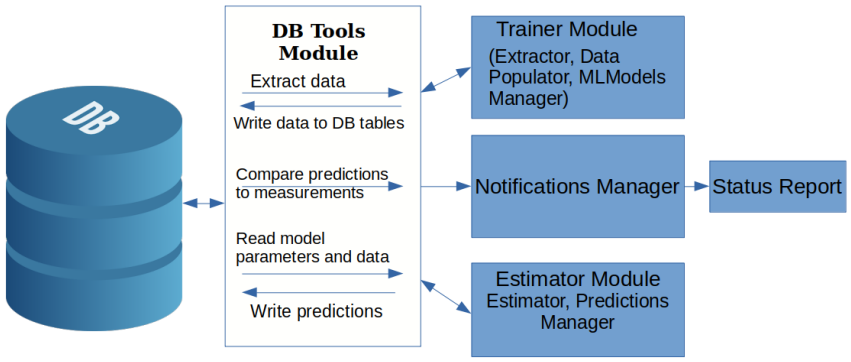


Figure 4.5: Illustration of hybrid network topology

The Mean Absolute Error (MAE) and Mean Squared Error (MSE), are used as performance metrics, and are defined as:

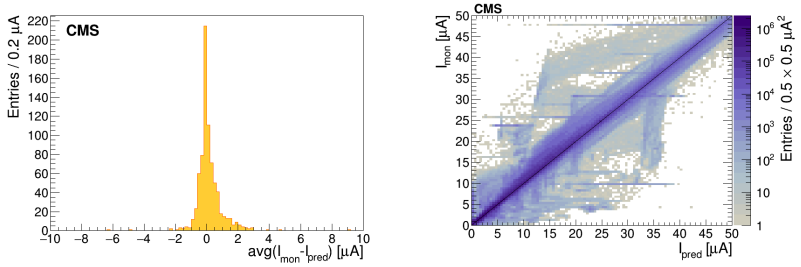
$$\mathrm{MAE} = \sum_{i=1}^{N} \frac{|\mathrm{I}^i_{mon} - \mathrm{I}^i_{pred}|}{\mathrm{N}} \tag{4.1}$$

$$\mathrm{MSE} = \sum_{i=1}^{N} \frac{(\mathrm{I}^i_{mon} - \mathrm{I}^i_{pred})^2}{\mathrm{N}} \tag{4.2}$$

All models are tested against the RPC currents measured in the two-month period between September and October of 2018.

Short-term training

The autoencoder approach performs best in the short-term training scenario. Its predictions are depicted in Fig. 4.6.

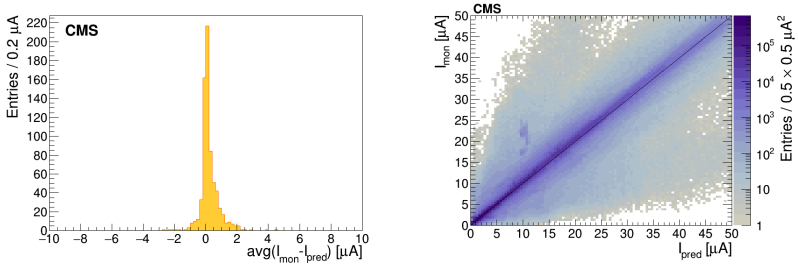


(a) Mean = 0.14 μA; σ = 0.83 μA (b) MAE = 0.49 μA; MSE = 1.39μA^2

Figure 4.6: Autoencoder ST performance, the difference between the measured and predicted value of the current is shown grouped by hardware HV channel (left) and in an exploded $2D$ view, of all individual data points (right)

Long-term training

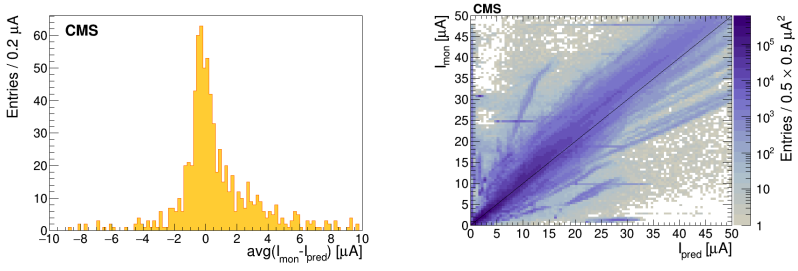
The long-term training of the models, instead favors the GLM, whose performace is depicted in Fig. 4.7.



(a) Mean = 0.21 μA; σ = 0.59 μA (b) MAE = 0.72 μA; MSE = 3.24μA^2

Figure 4.7: GLM LT performance, the difference between the measured and predicted value of the current is shown grouped by hardware HV channel (left) and in an exploded $2D$ view, of all individual data points (right)

Additionally, the HN is evaluated in a tailored scenario, delivering promising results, as shown in Fig. 4.8. This approach demonstrates potential for predicting currents far into the future, specifically in scenarios where there is a significant time gap between the end of the training period and the start of the prediction period.



(a) Mean = 0.60 μA; σ = 2.49 μA (b) MAE = 2.09 μA; MSE = 23.19μA^2

Figure 4.8: Hybrid Network performance, the difference between the measured and predicted value of the current is shown grouped by hardware HV channel (left) and in an exploded $2D$ view, of all individual data points (right)

Both in the GLM and the autoencoder approach, the sigma of the distribution of the performance histograms is $< 1\ \mu A$, which shows excellent predictive capcabilities consistent with the uncertainty in the current measurement which is also of that same order. All performance results are shown in detail in Table 4.1.

Table 4.1: Performance results for each tested scenario

Model class	Training period	Prediction period	1D histo mean [μA]	1D histo sigma [μA]	2D histo MAE [μA]	2D histo MSE [μA^2]
GLMv2	18-05-01 to 18-09-01	18-09-01 to 18-10-30	-0.02	1.65	1.23	7.62
GLMv2	17-07-01 to 18-07-01	18-09-01 to 18-10-30	0.33	1.66	1.23	7.42
GLMv2	16-05-01 to 18-07-01	18-09-01 to 18-10-30	0.21	0.59	0.72	3.24
Autoencoder	18-05-01 to 18-09-01	18-09-01 to 18-10-30	0.14	0.83	0.49	1.39
Autoencoder	17-07-01 to 18-07-01	18-09-01 to 18-10-30	0.69	1.44	0.96	4.18
Autoencoder	16-05-01 to 18-07-01	18-09-01 to 18-10-30	0.42	1.40	0.85	3.16
GLMv2	16-05-01 to 17-07-01	18-09-01 to 18-09-30	-0.24	2.59	1.92	18.69
Autoencoder	16-05-01 to 17-07-01	18-09-01 to 18-09-30	0.06	2.51	2.14	22.57
Hybrid	16-05-01 to 17-07-01	18-09-01 to 18-09-30	0.60	2.49	2.09	23.19

4.6 Tool integration and deployment

The tool has been integrated within the RPC automation framework, which is comprehensively discussed in Chapter 5. Additionally, a front-end interface for the tool has also been developed and deployed on CERN's *Platform-as-a-Service* (*PaaS*), a containerized deployment environment based on OKD [93], the community edition of OpenShift.

The work on this tool was finalized with a paper published in the Nuclear Instruments and Methods Section A (NIMA) journal [94].

Chapter 5

CMS RPC Automation Framework

RPC automation is a framework for the periodic aggregation, synchronization and automated analysis of condition data relevant to the operation of the RPC subsystem chambers at the CMS experiment. The RPC system is part of the CMS muon spectrometer and as described in detail in section 1.3, consists of 1056 chambers, with a total number of 123432 digital readout channels, read out synchronously with a 40 MHz clock.

Main RPC parameters such as currents and rates are continuously correlated with running conditions such as LHC luminosity, environmental humidity, pressure and temperature. All raw data is stored in Oracle databases by the WinCC OA framework. The RPC Automation leverages periodic data streaming within discrete time intervals to optimize the execution of database queries. Our methodology encompasses robust data streaming, real-time analysis of non-physics data, synchronization of data points with different timestamps, and the subsequent storage of processed data in new tables. Furthermore, the automation framework includes modules for preprocessing data intended for training machine learning-based models used in detector operation monitoring via anomaly detection, as explained in great detail in Chapter 4.

5.1 Automation Framework

The RPC automation framework comprises over 40 automata which are categorized as main and auxiliary, with each one designed for a specific task. This framework is directly tied to RPC operation and the Detector Control System (DCS) [40] which monitors and archives condition data in the CMSONR production database [95].

An overview of the configuration and the data flow is presented in Fig. 5.1. The first stage is the emulator, located in the development environment, which is responsible for preparing and setting configuration parameters for the detector.

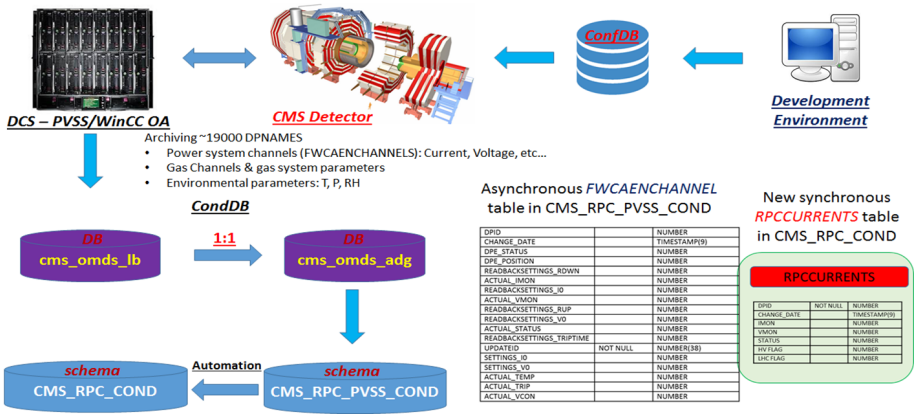


Figure 5.1: Overview of the RPC configuration and condition data flow, starting with the emulator, which prepares and stores configuration parameters. The RPC Detector Control System (DCS) monitors detector parameters and archives non-physics event data. The automation framework retrieves and processes this data, storing the results in structured formats for analysis and fast visualization.

These parameters are stored in the Configuration database. In the next stage, the role of the DCS is crucial, as it monitors all detector parameters and archives non-physics event data in the Condition database.

In turn, the CMS Next Generation Archiver (NGA) [96] registers raw data outside a predefined bandwidth into the *cms_omds_lb* read-only database. A real-time copy of this data is transferred to the *cms_omds_adg* read-only database. The RPC automation framework retrieves this raw data and following a processing through a data streaming analysis pipeline, stores the final results in the *CMS_RPC_COND* schema within the production database. This streamlined process ensures the accurate and efficient handling of detector condition data, supporting the operational integrity and research objectives of the CMS RPC system.

One of the tasks of the main automata is to synchronize the asynchronous data from the *CMS_RPC_PVSS_COND* schema and store them into well-structured tables within the *CMS_RPC_COND* schema on the production database. This includes raw data for RPC current, LHC machine and beam modes, environmental conditions, RPC gas flows, and RPC link board histograms (processed from ROOT files) used for the study of RPC rate data.

All complementary tasks are performed by the auxiliary automata, starting with the segmentation of the LHC filling cycle into four discrete blocks with the purpose of standardizing condition data for analysis. Additionaly, the auxiliary automata facilitate an in-depth investigation into multiple parameters, such as: accumulated integrated charge, current evolution, high voltage (HV) conditioning, and dependencies of RPC currents and rates on LHC luminosity

and charge per hit. The analyses are also extended to newly introduced virtual objects, which represent coarser-granularity detector components such as regions, wheels, disks, stations, and sectors.

The framework operates on a 4-hour cycle and has retroactively processed vast datasets from the CMS experiment. The adoption of virtual objects and predictive modeling has enhanced our understanding of detector behavior, by uncovering performance trends and providing forecasting capabilities.

5.2 Automation methodology

The design of the RPC automation framework required both the utilization of established data-handling methods and the development of approaches tailored to the specificities of the CMS experiment and the overall LHC environment. These include: data streaming, multithreading, data synchronization and tagging, current probe, block averaging method, the LHC block concept, and luminosity methods. In the rest of this section, each method is discussed in detail.

Current Probe Method

The "Current Probe" method was a key innovation that underlied the entire development of the automation framework. It is a lightweight process with minimal CPU load requirements and was crucial in tackling one of the major challenges: the management of the large size of the raw current data table (*FWCAENCHANNEL*). Conventional use of SQL queries on the full volume of data proved impractical and the solution was to segment the data in small, manageable chunks by applying the quieries for narrow time windows, processing data one day at a time.

This approach significantly reduced the database workload, exploiting the database optimization in dealing with high-demand operations.

Data Synchronization

The data from the DCS archiving manager (NGA) are stored asynchronously due to the change-based format, meaning that a value of a parameter is logged only upon update, with other unchanged fields left void for that particular timestamp. The parameters in question are: monitored current (Imon), monitored voltage (Vmon) and status.

Performing analysis on this data requires that each parameter has a value available for each timestamp, which imposes the need to populate the missing values. This was achieved by utilizing Oracle analytical and aggregation functions leading to the required data synchronization, as depicted in the the table excerpt in Fig. 5.2.

DPID	CHANGE_DATE	VMON	IMON	STATUS
315	25-JUN-17 11.58.29.083000000 PM	6501		
315	25-JUN-17 11.59.31.745000000 PM	6498		
315	26-JUN-17 12.03.46.240000000 AM		12.89999962	
315	26-JUN-17 12.03.46.240000000 AM			3
315	26-JUN-17 12.03.46.241000000 AM	8114		
315	26-JUN-17 12.03.49.194000000 AM	8240		
315	26-JUN-17 12.03.59.618000000 AM	8463		
315	26-JUN-17 12.04.03.358000000 AM	8592		
315	26-JUN-17 12.04.12.933000000 AM	8815		
315	26-JUN-17 12.04.16.631000000 AM	8878		
315	26-JUN-17 12.04.26.126000000 AM		13.30000019	
315	26-JUN-17 12.04.26.127000000 AM	9133		
315	26-JUN-17 12.04.29.838000000 AM		12.10000038	
315	26-JUN-17 12.04.29.838000000 AM	9226		
315	26-JUN-17 12.04.29.838000000 AM			1
315	26-JUN-17 12.04.39.462000000 AM		1	
315	26-JUN-17 12.04.39.463000000 AM	9231		
315	26-JUN-17 12.07.22.387000000 AM		0.699999988	

DPID	CHANGE_DATE	VMON	IMON	STATUS
315	25-JUN-17 11.58.29.000000000 PM	6501	0.5	1
315	25-JUN-17 11.59.31.000000000 PM	6498	0.5	1
315	26-JUN-17 12.03.46.000000000 AM	8114	12.9	3
315	26-JUN-17 12.03.49.000000000 AM	8240	12.9	3
315	26-JUN-17 12.03.59.000000000 AM	8463	12.9	3
315	26-JUN-17 12.04.03.000000000 AM	8592	12.9	3
315	26-JUN-17 12.04.12.000000000 AM	8815	12.9	3
315	26-JUN-17 12.04.16.000000000 AM	8878	12.9	3
315	26-JUN-17 12.04.26.000000000 AM	9133	13.3	3
315	26-JUN-17 12.04.39.000000000 AM	9231	1	1
315	26-JUN-17 12.07.22.000000000 AM	9231	0.7	1

Figure 5.2: The raw data storage format in the *FWCAENCHANNEL* (left) and the corresponding synchronized data stored in the *RPCCURRENTS* table (right)

Data Tagging

The data tagging process operates by associating each entry in the *RPCCURRENTS* table with specific operational conditions, including HV states, the state of the CMS magnetic field and the LHC beam mode. The tag, referred to as *flag* is stored as a separate column in the *RPCCURRENTS* table, as shown in Fig. 5.3.

DPID	CHANGE_DATE	VMON	IMON	STATUS	FLAG
315	25-JUN-17 11.58.29.000000000 PM	6501	0.5	1	20
315	25-JUN-17 11.59.31.000000000 PM	6498	0.5	1	20
315	26-JUN-17 12.03.46.000000000 AM	8114	12.9	3	16
315	26-JUN-17 12.03.49.000000000 AM	8240	12.9	3	16
315	26-JUN-17 12.03.59.000000000 AM	8463	12.9	3	16
315	26-JUN-17 12.04.03.000000000 AM	8592	12.9	3	16
315	26-JUN-17 12.04.12.000000000 AM	8815	12.9	3	16
315	26-JUN-17 12.04.16.000000000 AM	8878	12.9	3	16
315	26-JUN-17 12.04.26.000000000 AM	9133	13.3	3	16
315	26-JUN-17 12.04.39.000000000 AM	9231	1	1	24
315	26-JUN-17 12.07.22.000000000 AM	9231	0.7	1	24

Figure 5.3: An excerpt from the *RPCCURRENTS* table where the new parameter *FLAG* is stored as a separate column

The *FLAG* parameter stores information about the combined state of the aforementioned operational conditions by encoding them as a decimal representation of six binary condition bits, see Fig. 5.4.

The least significant bit, OFF, indicates whether the HV channel is turned off, the value being 0 when ON. The second bit, OFFSET, tags currents recorded when the voltage is set to 1000V. The third bit, STDB, signifies whether currents are at the standard standby value for the detector. The fourth bit, named RPC ON, indicates chambers being at the operational HV working point. These HV condition flags are computed internally in the process of writing records to the

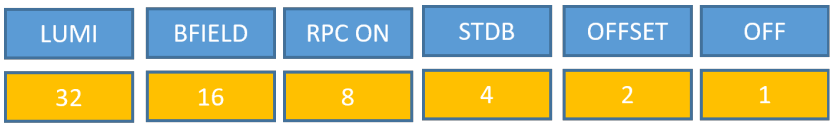


Figure 5.4: The six bit *FLAG* encoding

RPCCURRENTS table.

The remaining two bits are sourced from external schemas, with the fifth bit, named BFIELD, being set when the CMS solenoid's magnetic field exceeds $1.9T$ and the sixth and most significant bit, named LUMI, set when there is presence of collisions inside the machine. The values of these bits are retrieved from the *CMSFWMAGNET* table in the *CMS_DCS_ENV_PVSS_COND* schema and the *LUMISECTIONS* table in the *CMS_OMS* schema [97], respectively.

Block Averaging Method

The value of the voltage applied on an RPC HV channel can in many ocassions, be maintained at a constant value for a period of time. Parameters representing external conditions are also kept constant for the time points in question. Such a period of time is referred to as a "block". For the duration of a block, other monitorables such as the current (Imon), might nonetheless be non-constant.

In order to avoid the repeated storage of the same constant value for some parameters, and at the same time account for the variation in the current, an averaging procedure is employed for the latter, and the entirety of a block is written as a single record.

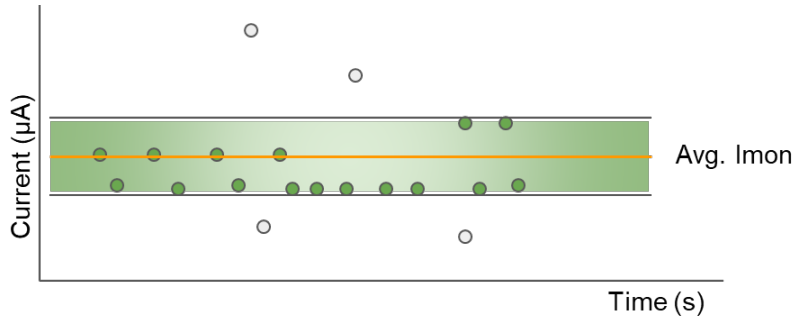


Figure 5.5: The mean Imon value is shown (orange line) along with the $\pm 4\sigma$ bands (green) and the two sets of points: the retained ones inside the bands and the removed outliers, outside the bands

The averaging procedure consists of first calculating the mean of the Imon values for the duration of the block, then removing the outliers, defined as the Imon values that lie outside the $\pm 4\sigma$ boundaries, as illustrated in Fig. 5.5. The procedure is applied iteratively, until no more than 10% of the data points have

been removed. The final mean Imon value is associated with the block and recorded in the table.

Data Streaming and Multithreading Techniques

Data streaming involves the real-time processing of data in small, continuous batches as it is generated, in contrast to conventional batch processing, where the data is processed all at once in large accumulated datasets. This method, used routinely in modern big-data applications, was suitable in the case of RPC automation in the context of LHC, as we are dealing with large amounts of data arriving at a high rate. The technical details of data streaming are discussed separately in subsection 5.3.3.

Multithreading is an additional feature of the developed framework. In our case, it was necessary to ensure the quick processing of data, running of multiple processes concurrently, and the overall efficient use of resources.

In combination, the two techniques are implemented extensively within the RPC automation framework, allowing for efficient processing of condition data within the necessary time frame without exhausting the compute resources available.

LHC Block Concept

LHC machine modes, such as Proton Physics, Ion Physics, Machine Development, and Shutdown, define the overall operational state of the LHC, varying based on the type of experiment or maintenance being conducted. Beam modes, on the other hand, specify the phases of beam preparation, acceleration, collision, and dump during LHC operation.

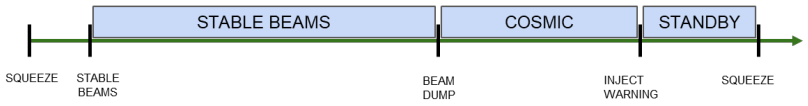


Figure 5.6: The segmentation of the LHC cycle into three separate blocks: COSMIC, STANDBY and STABLE BEAMS

LHC blocks are introduced as a way of dividing the full LHC cycle into three segments (see Fig. 5.6), each containing distinct RPC operation conditions:

- COSMIC - from Beam Dump to Inject Warning, the HV is ON and there is no beam in the machine
- STANDBY - from Injection to Squeeze, HV is STANDBY and there is a beam in the machine, but without collisions
- STABLE BEAMS - from Stable Beams to Beam Dump, HV is ON and there are physics collisions

The segment from Squeeze to Stable Beams is excluded from the analysis, as it represents a state where the detectors are turning on and the beam in the machine is not yet undergoing stable collisions, leading to current fluctuations and spikes.

A fourth additional LHC block, named "COOL-DOWN", was introduced as a subdivision of the "COSMIC" block and encompasses the interval between the Beam Dump and the end of Ramp Down.

LHC Luminosity Method

A luminosity section or lumisection is a CMS data-taking interval, having a rather arbitrary definition of 2^{18} LHC orbits, corresponding to approximately $23.01s$. The length for a lumisection was chosen such that it is small enough to be used as the "atomic" unit of data but large enough to facilitate the measurement of average instantaneous luminosity and the monitoring of subdetectors' status.

The study of RPC currents, as one of the main focus points of the RPC automation framework, had to be conformed to this paradigm. In order to assign a current value per lumisection for each HV channel, a novel streaming technique was developed.

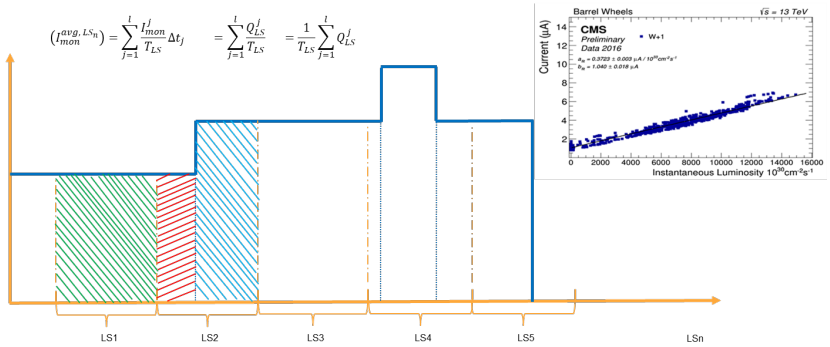


Figure 5.7: The current is integrated over lumisections to calculate the charge per lumisection

As illustrated in Fig. 5.7, it consists of integrating the charge per lumisection and then dividing by the lumisection duration, ending up with the average current per lumisection.

5.3 Automata

The following automata constitute the main group of automata: *UXC Environment*, *LHCLOG*, *RPCCURRENTS*, *RPCRATES*, *RPC Gas Flow Channel*

UXC Environment

The first automaton in the automation chain, *UXC environment* hadles data regarding the environmental parameters inside the UXC experimental cavern of CMS. It retrieves the values of the following parameters: pressure, temperature, relative humidity, and dewpoint and stores them in a table, in synchronized timestamps (Fig. 5.8).

CHANGE_DATE	PRESSURE	TEMPERATURE	RELATIVE_HUMIDITY	DEWPOINT
01-JAN-14 02.37.05.000000000 AM	964.62	19.94	26.5	2.22
01-JAN-14 04.37.06.000000000 AM	964.03	19.94	26.5	2.22
01-JAN-14 06.37.07.000000000 AM	963.74	19.94	26.5	2.22
01-JAN-14 08.37.08.000000000 AM	963.52	19.94	26.5	2.22
01-JAN-14 10.37.08.000000000 AM	962.2	19.94	26.5	2.22
01-JAN-14 12.37.09.000000000 PM	960.66	19.94	26.5	2.22
01-JAN-14 02.37.10.000000000 PM	959.71	19.94	26.5	2.22
01-JAN-14 04.37.11.000000000 PM	958.97	19.94	26.5	2.22
01-JAN-14 06.37.11.000000000 PM	958.02	19.94	26.5	2.22
01-JAN-14 08.37.12.000000000 PM	957	19.94	26.5	2.22
01-JAN-14 10.37.13.000000000 PM	956.41	19.94	26.5	2.22
02-JAN-14 12.37.13.000000000 AM	955.38	19.94	26.5	2.22
02-JAN-14 04.37.14.000000000 AM	954.14	19.94	26.5	2.22
02-JAN-14 06.37.16.000000000 AM	955.38	19.94	26.5	2.22
02-JAN-14 08.37.16.000000000 AM	956.63	19.94	26.5	2.22
02-JAN-14 10.37.17.000000000 AM	957.58	19.94	26.5	2.22
02-JAN-14 12.37.18.000000000 PM	958.24	19.94	26.5	2.22
02-JAN-14 02.37.18.000000000 PM	958.46	19.94	26.5	2.22
02-JAN-14 04.37.20.000000000 PM	958.83	19.94	26.5	2.22

Figure 5.8: Synchronized UXC cavern environmental data

LHCLOG

The LHCLOG automaton extracts the timestamps of changes in machine, beam and handshake modes from the CMS_DCS_ENV_PVSS_COND schema. This is achieved via predefined patterns and the obrained information is stored in an appropriate format.

RPCCURRENTS

This is the main automaton in the entire framework. It retrieves raw currents from the CMS_DCS_PVSS_COND schema, synchronizes them, and associates to each record the following parameters: RPC HV, CMS magnetic field, and LHC instantaneous luminosity information.

After the execution of the main automaton, two auxiliary automatons ("*RAMPUP*" and "*RAMPDOWN*") are applied which smooth the data further, via the removal of ramp-up and ramp-down currents.

The motivation for these auxiliary steps stems from the nature of RPC detectors: due to their large surface area, they behave as large capacitors which leads to a delay in the stabilization of the currents once the hardware channel status switches to "1" (Fig. 5.9 (left)). This transient behavior leads to artificial cosmic current spikes which prompted the employment of this "cleaning" action.

DPID	CHANGE_DATE	VMON	IMON	STATUS	FLAG	TIME_DELTA
315	24-NOV-09 05.39.31.000000000 PM	833	18	3	0	1
315	24-NOV-09 05.39.32.000000000 PM	865	18	3	0	2
315	24-NOV-09 05.39.34.000000000 PM	898	18	3	0	3
315	24-NOV-09 05.39.38.000000000 PM	997	9.899999619	1	2	2
315	24-NOV-09 05.39.40.000000000 PM	997	5.599999905	1	2	4
315	24-NOV-09 05.39.44.000000000 PM	997	1.799999952	1	2	9
315	24-NOV-09 05.39.53.000000000 PM	997	1.399999976	1	2	354
315	24-NOV-09 05.45.47.000000000 PM	997	0.899999976	1	2	4
315	24-NOV-09 05.45.51.000000000 PM	997	1.200000048	1	2	49

DPID	CHANGE_DATE	VMON	IMON	STATUS	FLAG	TIME_DELTA
315	24-NOV-09 05.39.31.000000000 PM	833	18	3	0	1
315	24-NOV-09 05.39.32.000000000 PM	865	18	3	0	2
315	24-NOV-09 05.39.34.000000000 PM	898	18	3	0	18
315	24-NOV-09 05.39.53.000000000 PM	997	1.399999976	1	2	354
315	24-NOV-09 05.45.47.000000000 PM	997	0.899999976	1	2	4
315	24-NOV-09 05.45.51.000000000 PM	997	1.200000048	1	2	49
315	24-NOV-09 05.46.40.000000000 PM	997	1.600000024	1	2	3
315	24-NOV-09 05.46.43.000000000 PM	997	1.100000024	1	2	208
315	24-NOV-09 05.50.11.000000000 PM	997	1.5	1	2	19

Figure 5.9: Filtering out current decay due to the capacitive detector effect after voltage ramp-up (left) ensures stable status "1" data by reassigning removed rows to the ramp-up phase, combining them with the last status "3" entry (right)

The "*RAMPUP*" automaton consists of a method similar to the one described in Section 5.2. It identifies stable current values and removes any outliers outside the stability bands. This filtering could potentially lead to gaps in the timeline, an effect which the automaton avoids by reassigning the filtered current decay records to the ramp-up phase, tagging them with status "3" (Fig. 5.9 (right)).

Similarly, the "RAMPDOWN" automaton deals with the ramp-down phase, by eliminating the erroneous entries with status "1". A direct transition from status "5" (ramp-down) to status "0" (off) is applied, as highlighted in Fig. 5.10 (left), and the duration of the removed status "1" row is appended to the final status "5" entry (Fig. 5.10 (right)).

DPID	CHANGE_DATE	VMON	IMON	STATUS	FLAG	TIME_DELTA
315	01-OCT-09 06.36.42.000000000 PM	1287	0	5	0	4
315	01-OCT-09 06.36.46.000000000 PM	1274	0	5	0	4
315	01-OCT-09 06.36.50.000000000 PM	1263	0	5	0	2
315	01-OCT-09 06.36.52.000000000 PM	1252	0	5	0	4
315	01-OCT-09 06.36.56.000000000 PM	1240	0	1	0	2
315	01-OCT-09 06.36.58.000000000 PM	1240	0	0	1	2
315	01-OCT-09 06.37.00.000000000 PM	1228	0	0	1	4
315	01-OCT-09 06.37.04.000000000 PM	1216	0	0	1	4

DPID	CHANGE_DATE	VMON	IMON	STATUS	FLAG	TIME_DELTA
315	01-OCT-09 06.36.42.000000000 PM	1287	0	5	0	4
315	01-OCT-09 06.36.46.000000000 PM	1274	0	5	0	4
315	01-OCT-09 06.36.50.000000000 PM	1263	0	5	0	2
315	01-OCT-09 06.36.52.000000000 PM	1252	0	5	0	6
315	01-OCT-09 06.36.58.000000000 PM	1240	0	0	1	2
315	01-OCT-09 06.37.00.000000000 PM	1228	0	0	1	4
315	01-OCT-09 06.37.04.000000000 PM	1216	0	0	1	4
315	01-OCT-09 06.37.08.000000000 PM	1197	0	0	1	6

Figure 5.10: Filtering out erroneous status "1" during ramp-down ensures a direct transition from status "5" (ramp-down) to status "0" (off), with the removed row's duration added to the last status "5" entry

One additional novelty developed in the context of this automaton, which was then re-used as well in other automata of the chain, was the concept of virtual objects. This advancement enabled the study of lower-granularity objects, ranging from Barrel wheels, stations, and sectors, as well as Endcap disks, rings, and sectors, to even broader components, such as the entire Barrel and Endcaps. This hierarchical approach allowed for a more comprehensive analysis of the detector's behavior at varying levels of granularity, facilitating insights at both localized and global scales. For instance, currents are now being studied as a function of the CMS beampipe axis (z), radial distance from the beampipe (R) and azimuthal angle (ϕ) of the CMS coordinate system.

Virtual objects are assigned unique data point identifiers (DPID) in the *CMS_COND_RPC* schema, despite not being representative of any physical

hardware channels. Records for the virtual objects are aggregated based on the detector geometrical partitioning, as defined by a DPID map. The aggregated averaged current (Imon) value is recorded for each virtual DPID at every timestamp when any of the associated physical hardware DPID channels is modified. Virtual currents are stored in the same manner as standard HV channel currents.

RPCRATES

Along with currents, RPC trigger rates [98] are the second non-physics event parameter central to the RPC automation framework and closely examined in RPC correlational studies. The raw trigger rate data are stored in *ROOT* [76] files which the automaton processes in order to generate records to be stored in a table in the *OMDS* schema, as illustrated in Fig. 5.11.

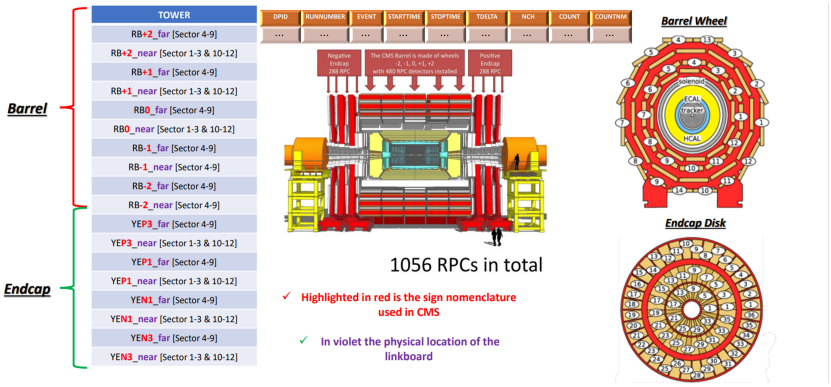


Figure 5.11: Raw data from 2748 link boards, collected from 1056 Resistive Plate Chambers operated in Run-II and Run-III, are transferred from 18 ROOT files per run into a structured database table format in the RPC automation schema on *OMDS*, as shown in the top middle section of the figure. The physical locations of the trigger towers, indicating the sectors from which data is collected, are also presented

The RPC trigger is composed of 10 towers in the Barrel and 8 in the Endcap regions. One ROOT file per trigger tower is stored in each run, resulting in a total of 18 *ROOT* files per run, containing the trigger rates. The trigger rates are monitored only when the RPC subsystem is either included in a global Data Acquisition System (DAQ) [99] run or when it is in a local run.

Therefore, trigger rates are stored on a per-run basis rather than continously over time. They can be categorized in collision and cosmic rates. However, correlation studies performed in the framework focus solely on the collision rates, eliminating the need to explicitly tag the records for LHC beam mode or CMS magnetic field state.

The content of the *ROOT* files provides the number of counts within a given

time readout window, referred to as an "event". The data is read from the Link Boards (LB), part of the RPC link board system, which provides an interface between the detector front-end electronics and the trigger electronics in the CMS Underground Service Cavern (USC). The length of an event is not constant and is in fact defined by a Mutual Exclusion (mutex) operator. The role of this operator is to determine when a given LB is read out. The database records the start and stop times of each event in a run, along with the counts per LB per event. However, trigger rates are not directly stored in the database but are instead calculated during data streaming by dividing the counts by the event duration in seconds, a process handled by the study automata.

In analogy to the case of RPC currents, RPC trigger rates are also associated with the virtual objects introduced previously. A DPID map groups link boards into virtual DPID channels, representing larger-granularity detector components, such as Barrel and Endcap regions and their respective wheels, disks, stations, rings and sectors. Averaged trigger rate values are stored whenever any of the associated hardware component values undergoes a change.

RPC Gas Flow Channel

RPCGASFLOW is the last automaton in the automation chain. It fetches asynchronous data from the *CMS_RPC_PVSS_COND* schema and synchronizes them using the method described in the previous sections. The resulting records are stored in a dedicated table (Fig. 5.12).

DPID	CHANGE_DATE	FLOWIN	FLOWOUT
16310	25-JUN-18 03.37.40.000000000 PM	31.6	30.7
16310	25-JUN-18 10.03.54.000000000 AM	31.5	30.3
16310	25-JUN-18 04.29.55.000000000 AM	31.5	30.5
16310	24-JUN-18 10.56.05.000000000 PM	31.5	30.7
16310	24-JUN-18 05.22.15.000000000 PM	31.3	30
16310	24-JUN-18 11.48.26.000000000 AM	31.4	30.4
16310	24-JUN-18 06.14.31.000000000 AM	31.4	30.2
16310	24-JUN-18 12.41.05.000000000 AM	31.4	30.4
16310	23-JUN-18 07.07.18.000000000 PM	31.7	30.5
16310	23-JUN-18 01.28.08.000000000 PM	31.6	30.2
16310	23-JUN-18 07.54.15.000000000 AM	31.6	30.3
16310	23-JUN-18 02.20.40.000000000 AM	31.4	30.2

Figure 5.12: Excerpt from the *RPCGASFLOW* table in the *CMS_RPC_COND* schema, where the main gas parameters, gas flow-in and flow-out are stored by the *RPCGASFLOW* automaton in a synchronous manner

5.3.1 Auxiliary automata

Once the main automata have performed the necessary structuring and synchronization of raw data, the automation chain proceeds with the execution of the auxiliary automata. It is at this stage that the correlational studies of the fundamental RPC parameters are performed.

In addition to correlational studies, there are several automata in the chain that perform miscellaneous studies such as: HV Conditioning, Current Evolution, Integrated Charge, Operation Mode Identification and Active Channels, covering a multitude of aspects of RPC operations. In what follows, a short overview of each auxiliary automaton is provided.

Integrated Charge

One of the earliest automata, RPC Integrated Charge was designed with the goal of assisting RPC aging studies. The automaton has since provided reference values [100] for the GIF++ iradiation facility tests [101].

The automaton computes the integrated currents over the time intervals between consecutive records in the *RPCCURRENTS* table and stores the daily total for each HV channel (DPID. For each DPID, two types of daily values are recorded: the integrated charge for *COLLISION* and for *RPC_ON*. The *COLLISION* type integrates only currents tagged with flag 56 (described in Section 5.2), while *RPC_ON* accounts for all currents recorded when HV is at working point, thus including both cosmic and collision currents.

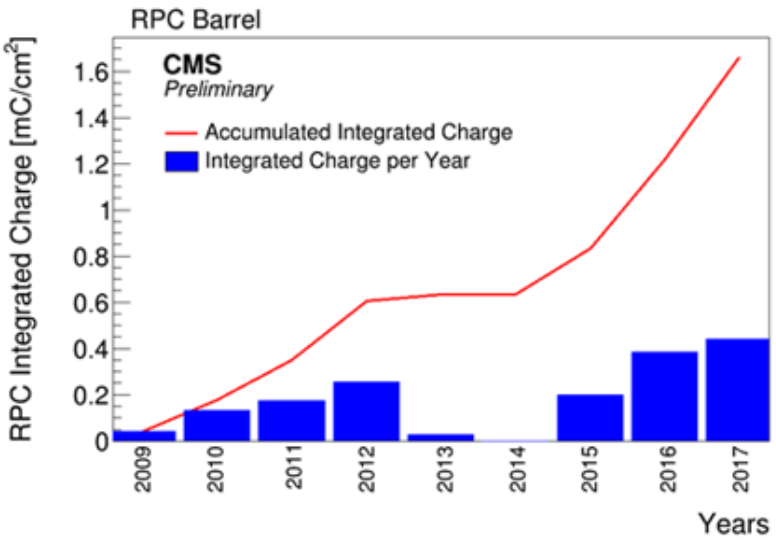


Figure 5.13: The integrated charge per year along with the accumulated integrated charge for the virtual object "Barrel" is shown in the interval 2009-2017

The integration process has been applied retroactively, starting from the onset of RPC operations in 2009. Fig. 5.13 presents a plot of the yearly integrated charge and the accumulated charge over time, for the virtual object "Barrel". The versatility of virtual objects in the study of this specific parameter is evident, as it is with other parameters tracked by the auxiliary automata.

Accumulated Integrated Charge per HV channel

This auxiliary automaton is located just after the RPC Integration Charge in the automation chain. It records a single daily value for the integrated charge per HV channel, in a table whose structure is shown in Fig. 5.14. This coarser-grained view of accumulated charge allows for its fast and efficient plotting in the graphical interfaces associated with RPC operations and facilitates the use of accumulated charge as a proxy for time in correlation analyses.

Figure 5.14: The structure of the database table where the daily per DPID accumulated charges are stored, for both hardware HV channels and virtual objects

Accumulated Integrated Charge per Chamber

The HV channels in the Endcap regions of CMS power two RPC chambers each. In October 2014, the mapping of HV channels to chambers was restructured to optimise performance. A transition to a chamber-centric paradigm was therefore essential in order to preserve a coherent continuinity of the objects for which the integrated charges and the accumulated charges were computed. For this purpose, a new label was introduced, "*CHAMBER_ID*", which allowed for the quantities associated with a specific chamber to be traced over time, even after the remapping.

The *Accumulated Integrated Charge per Chamber* automaton operates on a per-chamber basis, conputes the accumulated charges and records them in a separate database table. Fig. 5.15 shows the structure of the database table where the per *CHAMBER_ID* records are stored.

Figure 5.15: The structure of the database table where the daily per chamber accumulated charges are stored

HV Conditioning

RPC HV Conditioning is a periodic current scan performed 3-4 times a year. The values of the currents at a set of incremental values of the applied HV are

read out: 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 8500, 9000, 9100, 9200, 9300, 9400, 9500, and 9600 V.

For each HV point, a number of current values is measured and stored. In order to obtain a single represeentative value of the current for each HV point, the *HV Conditioning* automaton applies the block-averaging method described in Section 5.2. The resulting records are grouped in a set spanning all the conditioning HV values and are stored in a separate table in the database.

HV Conditioning Fit

For each HV conditioning set, the *HC Conditioning Fit* automaton processes the HV points in the interval between 1000 and 7000 V, for both hardware HV channel and virtual objects. It performs a linear fit in this so-called "ohmic"

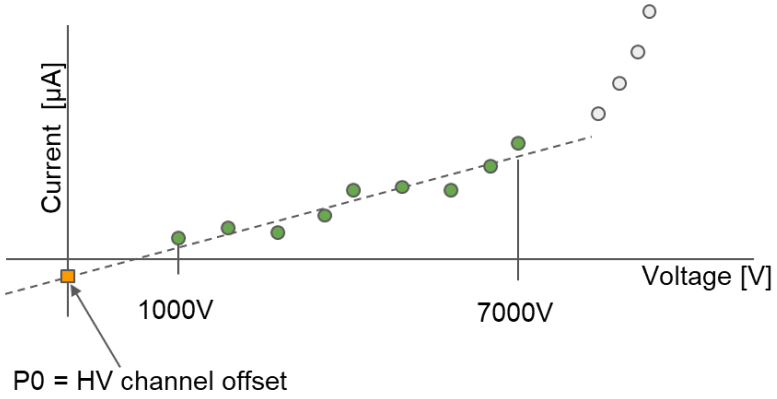


Figure 5.16: RPC HV Conditioning data points and a linear fit in the ohmic region (1000 − 7000 V)

region (Fig. 5.16) and the fit is further optimized before the final parameters are extracted and recorded in a the *RPCCURRENTS_HVCOND_FIT* table in the database. The recorded parameters are: p_0 (HV board channel offset) and p_1 (interpreted as inverse resistance in the ohmic region). These data facilitate fast plotting of parameters, such as the resistance $(1/p_1)$ whose evolution accross multiple HV conditioning scans is of interest since it can be indicative of long-term changes in chamber properties

Operation Mode

RPC chambers in the CMS experiment are constructed using thin gaps in order to achieve the desired time reolution, spatial resolution and cluster size. However, this design feature reduces the working volume, which comes at the cost of decreased efficiency. To address this limitation, the chambers are operated in Double-Gap (DG) mode. In this configuration, two separate gaps are stacked

on either side of a common readout plane. The two HV layers associated with each gap, are commonly labelled as Up and Down.

In DG mode, both layers are operational and the signal induced on the readout plane is a sum of the signals induced by each strip. The *Operation Mode* automaton determines whether a chamber is actually operating in DG mode or in Single-Gap (SG) mode. The SG mode is a consequence of one of the one layers being upowered, leading to the current being approximately halved during the ramp-up phase.

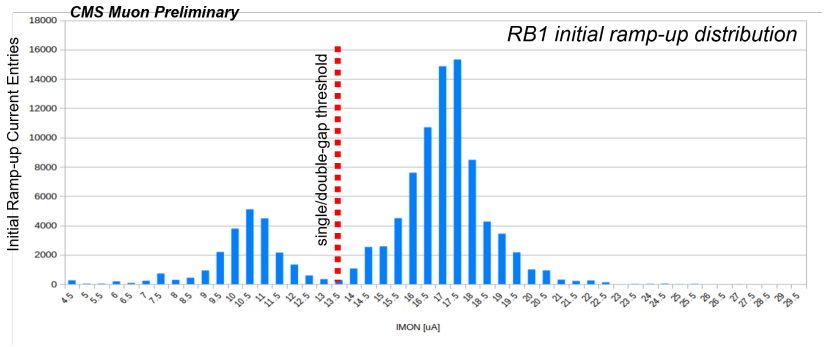


Figure 5.17: Count of monitored current values for all chambers of type RB1, recorded during the initial ramp-up phase, during the initial ramp-up phase, throughout the entire operational history. The two visible peaks correspond to the two operation modes: Single-Gap (SG) and Double-Gap (DG). The threshold of 13.5 μA (red dashed line) separates the two operation modes

The amplitude of the ramp-up current is directly proportional to the gap surface area: larger gaps draw larger currents. To account for this variability, different threshold are set for each chamber type, namely RB1, RB2/2, RB2/3, RB3, RB4 − 1500, RB4 − 2500, RE1/2, RE1/3 and REn/2 or REn/3, where n denotes the station number, with $n \epsilon [2, 4]$. The threshold vary from 500 V to 1030 V and are used in the automaton to determine the operation mode of each chamber.

Lumi Fit automata

The Lumi Fit automata analyze the correlations of fundamental RPC characteristics, such as currents and trigger rates as a function of LHC instantaneous luminosity. The framework includes the following automata performing studies of such character:

Current Lumi Fit

This automaton associates instantaneous luminosity values, measured by luminometers and provided via the non-physics event data bus, to the current data for HV channels and virtual objects. As described in Section 5.2, the currents are averaged per lumisections, which guarantees synchronisity and alignment

with the timestamps of luminosity data. Furthermore, the automaton performs a linear fit of the currents to the luminosity values (Fig. 5.18 (left)) and stores the obtained fit parameters in a database table: the intercept, representing the current at the end of ramp-down, i.e. zero luminosity, and the slope, representing the sensitivity of currents to change in luminosity.

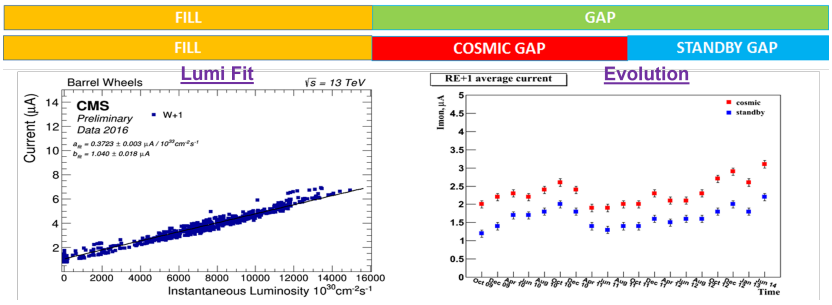


Figure 5.18: Evolution of current as function of instantaneous luminosity for virtual object "Barrel W+1" (left), and evolution of current for virtual object "Endcap RE+1", as function of time (right)

Rate Lumi Fit A similar methodology is emplyed by the *Rate Lumi Fit* automaton, to assign trigger rates to instantaneous luminosities and to perform a linear fit (Fig. 5.19). The obtained fit parameters, the intercept and the slope, represent the rate at zero luminosity and the trigger rate sensitivity to luminosity changes, respectively.

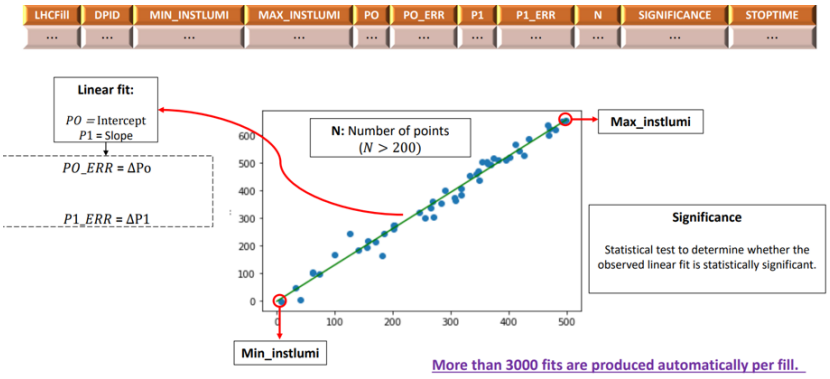


Figure 5.19: Rates evolution with instantaneous luminosity

Evolution

The *Evolution* automaton combines the power of the *LHC block concept* and the *Block averaging method*, both introduced in section 5.2. The purpose of

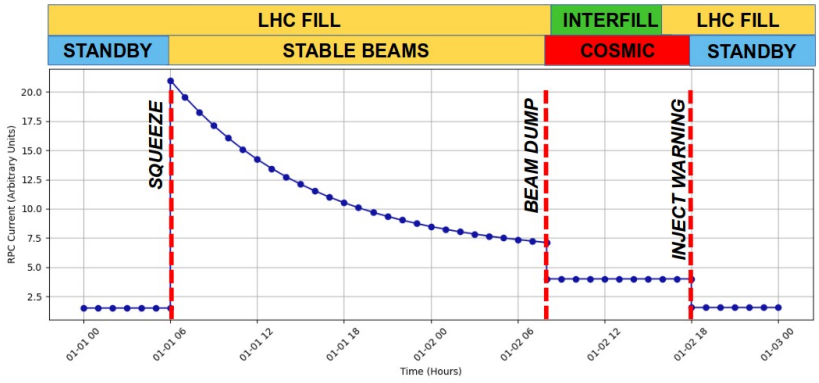


Figure 5.20: Typical evolution of RPC currents during an LHC fill cycle. The LHC blocks segmentation of the cycle is indicated by the red dashed vertical lines

this automaton is to keep track of the entire operational history of an RPC HV channel without needing to store all the recorded raw data points, which would constitute a vast amount of data.

Firstly, the automaton segments the current data in three "types", which leads to the definition of three distinct current blocks: offset (measured at 1000V), standby (measured at 6500V) and cosmic (measured at working point but in no-beam conditions). It then performs block averaging of the raw data points, resulting in only a single record for each block. This reduces drastically the amount of data that needs to be saved, while at the same time retaining a value which characterizes the behavior of the HV channel.

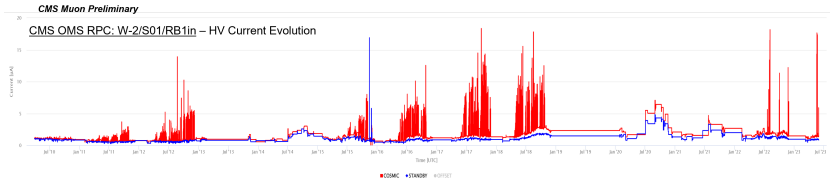


Figure 5.21: Current evolution for the HV channel supplying Barrel chamber *W-2/S01/RB1in* after the application of block averaging

Although the overall trend of the gets described correctly even with a reduced number of data points, certain limitations persist at this post-block-averaging stage, as can be seen in Fig. 5.21. These limitations arise in the form of erroneous current spikes, whose origin can typically be traced to Van-der-Meer scans or luminosity leveling. To deal with this issue, the automaton performs a

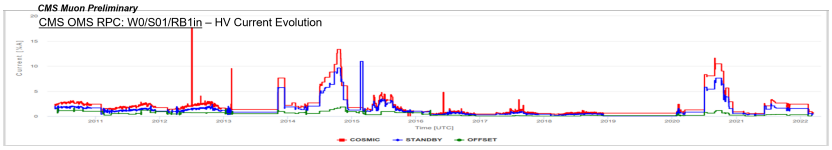


Figure 5.22: Current evolution for the HV channel supplying Barrel chamber *W-2/S01/RB1in* after the application of both block averaging and the alignment of the current blocks with LHC blocks

second step, during which it matches the current blocks with the LHC blocks: standby, cosmic, and stable beams (discussed in section 5.2 and depicted in Fig. 5.20). The result is the removal of the current spikes (Fig. 5.22), which classifies the data as "clean" and allows for efficient plotting and analysis.

Active Channels

During Run-II and Run-III of the LHC, the RPC system has been operated using 123432 readout channels. Each electronic channel corresponds to a single copper strip of the readout planes, located between the two gas gap layers in an RPC chamber.

In operation, there are several reasons that can lead to a channel being masked: noise, hardware issues or malfunctions. Masking a faulty channel is imposed by means of switching to "0" of a single bit, named the *enable* bit. This bit is stored in the configuration database and is loaded in hardware to determine the detector configuration.

The *Active Channels* automaton monitors the status of the readout channels, logging the timestamp of any change in their *enable* bit. This parameter serves as a valuable metric in RPC performance studies, as it indicates the proportion and distribution of system channels that were operational or active during a given period.

5.3.2 Machine Learning automata

The Machine Learning tool for current monitoring and anomaly detection, detailed in Chapter 4, has been seamlessly integrated into the automation framework. Leveraging the developed data-handling tools, it enhances the framework by introducing predictive and trend-revealing capabilities. These features are critical for anticipating operational failures within the RPC subsystem, adding a new dimension to automation studies.

The tool's features have been incorporated into three distinct automata, which are discussed in the remainder of this section.

ML Training Data

The first automaton in the ML part of the automation chain, *ML Training Data* prepares the data needed to train the ML models used to characterize the behavior of detector currents. It retrieves current and environmental data from other tables of the database, already populated by previous automata in the chain. In addition, it retrieves luminosity data from an external schema. The data is recorded in an appropriate format in the *MLTrainingData* table (Fig. 5.23) in the database.

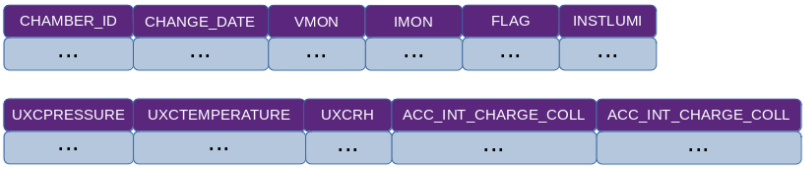


Figure 5.23: Structure of the *MLTrainingData* table

ML Predicted Currents

The *ML Predicted Currents* automaton is executed immediately after the training data have been prepared. It generates predictions using a pre-trained set of models. The training of these models is conducted independently of the automation processes, as it occurs much less frequently and is executed manually by the experts overseeing the automation. The resulting predictions are stored in a separate database table, whose structure is depicted in Fig. 5.24.

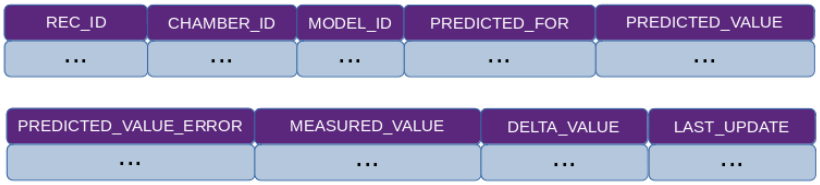


Figure 5.24: Structure of the *MLPredictedCurrent* table

ML Notifications

The last of the ML automata, *ML Notifications* goes through the predictions recorded by the previous automaton and performs a chack for each chamber, following the logic explained in Section 4.3. The resulting notifications about anomalous chamber behavior, if any, are stored in the *MLNotifications* table (Fig. 5.25)

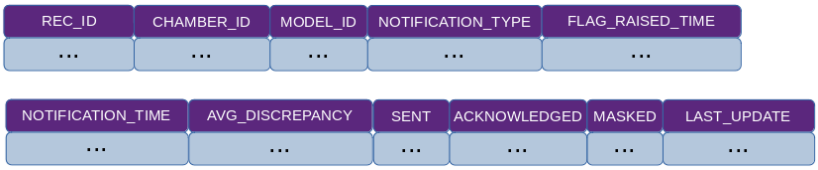


Figure 5.25: Structure of the *MLNotifications* table

5.3.3 Java implementation and deployment

The RPC automation software has been developed as a Java framework, utilizing the build automation tool Gradle. Communication with the Oracle databases is handled through a custom library called "Spartan", specifically created for the needs of this project. Each automaton within the framework is implemented using two primary Java classes, conventionally named <automaton-name>Automat and <automaton-name>Task. The Task class is responsible for breaking the job into discrete time steps, ensuring structured and efficient execution.

An exposition of the details of the Java implementation is given in Appendix A to this dissertation.

The automation framework has been successfully deployed in production and is currently operating on the CMS network. It runs via a Linux job scheduler, which triggers the framework every four hours.

5.4 Vision and prospects

The ideology behind the development of the RPC automation framework consists in a vision for detector operation where operator involvement would be minimized only to a supervision role and performance would tend to the optimal one.

Eliminating human errors and reaction delays, by means of an automatic correctional feedback mechanism, is projected as a novel approach to the construction and operation of modern high-energy physics experiments.

The remaining step in completing this feedback loop is the development of an automated interface between the RPC automation framework and the DCS.

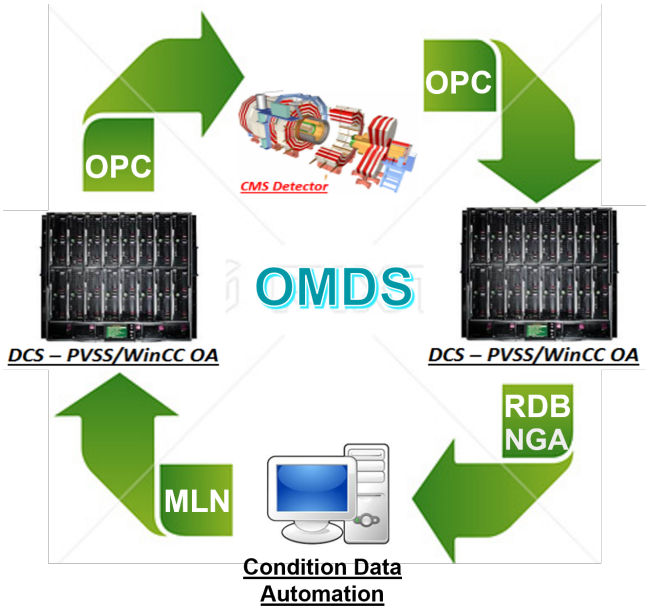


Figure 5.26: The RPC automation framework's end goal, to complete the feedback loop to the detector, by eventually reaching a level of sophistication that would allow for the automatic adjustment of detector parameters via self-monitoring and self-correction so as to ensure optimal operation

Conclusions

This work presented the precision measurement of CP violation in the $B_s^0 \to J/\psi\phi(1020)$ decay using data collected at the CMS experiment at $\sqrt{s} = 13$ TeV. The combined result, with the previous CMS Run-1 measurement at 8 TeV, yielded a value of -74 ± 23 mrad^{-1}. This value exhibits no deviation from the SM prediction, but with a difference from zero of 3.2 standard deviations, it constitues the first observation of CP violation in the interference between decay and mixing in the $B_s^0 \to J/\psi\phi(1020)$ decay.

The monitoring tool for RPC currents monitoring based on ML algorithms for anomaly detection, was shown to exhibit powerful predictive capabilities. Starting with the ongoing Run-III of the LHC (at the time of writing), it is being successfully used for recognizing chamber misbehavior tendencies during operation.

The RPC automation framework development has reached its final stages and is now running on production. The database populated by it is routinely used by RPC experts to monitor detector properties and to perform various analyses, especially regarding currents and rates. The full potential of the framework is yet to be exploited, as one final development step is necessitated to transform it into a correction agent for detector parameters during operation.

Throughout the presented work, the use of Machine Learning approaches was essential in overcoming challenges in a variety of tasks. The use of Deep Neural Networks and Boosted Decision Trees proved crucial in refining the physics analysis methods. On the other hand, Generalized Linear Models and Autoencoders where a central tool for the work done on detector monitoring and automation.

Contributions

- An algorithm for the estimation of the angular efficiency function in $B_s^0 \to J/\psi\phi(1020)$ decays with the use of the Kernel Density Estimate method, enhanced by Fast Fourier Transform, was implemented. The algorithm was applied to characterize the CMS detector's angular efficiency, a crucial step in the angular analysis of the $B_s^0 \to J/\psi\phi(1020)$ decay's final state.

- A novel procedure for the per-event reweighting of Monte-Carlo samples used in the study of $B_s^0 \to J/\psi\phi(1020)$ decays was developed. It was implemented via an Machine Learning approach based on Boosted Decision Trees and served to correct for the inherent mismodelling present in the MC samples.

- A procedure for the estimation of the background contribution in $B_s^0 \to J/\psi\phi(1020)$ decays, originating from $\Lambda_b \to J/\psi Kp$ decays, was implemented. The bootstrap method was used to account for bias introduced by the employment of Random Number Generators. The obtained result served to make a decision on whether this contribution is to be explicitly included in the fit model.

- The algorithms and methods listed above were successfully applied to enhance the precision measurement of CP violation performed in the $B_s^0 \to J/\psi\phi(1020)$ decay channel. The analysis yielded the first evidence for the studied channel (with a certainty of 3.2 standard deviations) of CP-violation in the interference between mixing and decay.

- A Machine Learning based tool for the monitoring of CMS RPC currents quality via anomaly detection, was developed. The tool was implemented in Python using the Tensorflow package and utilizes Generalized Linear Models and Autoencoders to model the behavior of RPC currents. Its decision-making logic produces notifications which starting with the ongoing Run-III are helping the operators to anticipate any chamber misbehavior before it leads to operational failure.

- The ML monitoring tool was integrated in the larger context of the Java-based CMS RPC automation framework. Furthermore, the framework was augmented by developing various task-processing units, called automata,

which broadened the range of studies that can be performed using its output.

Publications

- E. Shumka et al. Machine learning based tool for cms rpc currents quality monitoring. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1054:168449, 2023. doi: https://doi.org/10.1016/j.nima.2023.168449
- The CMS Collaboration. Measurement of time-dependent CP violation in $B_s^0 \to J/\psi\phi(1020)$ decays with the CMS detector. Technical report, CERN, Geneva, 2024. Expected to be published in *Physical Review Letters*. url: https://cds.cern.ch/record/2894821?ln=en
- A. Dimitrov, ..., E. Shumka et al. CMS RPC Non-Physics Event Data Automation Ideology. Submitted to *Nuclear Instruments and Methods Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*

Conference reports

- E. Shumka on behalf of the CMS muon group, "ML-based tool for RPC currents quality monitoring", *Poster*, RPC2022, CERN, Geneva, 2022
- E. Shumka on behalf of the CMS muon group, "ML-based tool for RPC currents quality monitoring", *Talk*, ACAT2022, Bari, Italy, 2022.
- E. Shumka et al., "Utilizing open-source toolkits for the simulation of avalanche formation and space-charge effects in Resistive Cylindrical Chambers", *Talk*, RPC2024, Santiago de Compostela, Spain, 2024.

Acknowledgments

Appendix A

Java implementation of the Automation Framework

Bibliography

[1] L Arnaudon, P Baudrenghien, M Baylac, G Bellodi, Y Body, J Borburgh, P Bourquin, J Broere, O Brunner, L Bruno, C Carli, Friedhelm Caspers, S M Cousineau, Y Cuvet, C De Almeida Martins, T Dobers, T Fowler, R Garoby, F Gerigk, B Goddard, K Hanke, M Hori, M Jones, K Kahle, Willi Kalbreier, T Kroyer, D Küchler, A M Lombardi, L A López-Hernandez, M Magistris, M Martini, S Maury, E Page, M Paoluzzi, M Pasini, U Raich, C Rossi, J P Royer, E Sargsyan, J Serrano, R Scrivens, M Silari, M Timmins, W Venturini-Delsolaro, M Vretenar, R Wegner, W Weterings, and T Zickler. Linac4 Technical Design Report. Technical report, CERN, Geneva, 2006. revised version submitted on 2006-12-14 09:00:40.

[2] H Damerau, A Funken, R Garoby, S Gilardoni, B Goddard, K Hanke, A Lombardi, D Manglunki, M Meddahi, B Mikulec, G Rumolo, E Shaposhnikova, M Vretenar, and J Coupard. *LHC Injectors Upgrade, Technical Design Report*. 2014.

[3] *Proton synchrotron of 28 GeV*. Technical notebook. CERN, Geneva, 1970.

[4] *Le super synchrotron à protons de 400 GeV*. Cahier technique. CERN, Geneva, 1977.

[5] Esma Mobs. The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018. 2018. General Photo.

[6] AC Team. Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC. 1999.

[7] L. Rossi. The LHC Superconducting magnets. *Proceedings of the 2003 Particle Accelerator Conference*, 2003.

[8] CERN Photographic Serive. Animation of CERN's accelerator network, 2013.

[9] A. Airapetian and et al. *ATLAS detector and physics performance: Technical Design Report, 1.* Technical design report. ATLAS. CERN, Geneva, 1999.

[10] B Alessandro and et al. *ALICE: Physics Performance Report. ALICE physics performance : Technical Design Report*, volume 32 of *Technical design report. ALICE*. CERN, Geneva, 2005.

[11] A Augusto Alves and et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. Also published by CERN Geneva in 2010.

[12] Sourav Sarkar, Helmut Satz, and Bikash Sinha. *The Physics of the Quark-Gluon Plasma*. Springer, 2010.

[13] ALICE Collaboration. https://alice.cern/experiment. Schematic cutaway ALICE experiment.

[14] Lindner, Rolf. LHCb layout_2. LHCb schema_2. LHCb Collection., 2008.

[15] ATLAS collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.

[16] CMS collaboration. A New Boson with a Mass of 125 GeV Observed with the CMS Experiment at the Large Hadron Collider. *Science*, 338(6114):1569–1575, 2012.

[17] The ATLAS experiment. Online; accessed September 2024.

[18] Public CMS Luminosity Information . Online; accessed October 2024.

[19] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, Aug 2008.

[20] I. Neutelings. How to draw diagrams in latex with tikz., Sept 2019.

[21] Klyukhin V. Design and Description of the CMS Magnetic System Model. *Symmetry*, 13, 2021.

[22] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 2010.

[23] V Karimäki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudenreich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[24] CMS Tracker Collaboration. Stand-alone cosmic muon reconstruction before installation of the CMS silicon strip tracker. *Journal of Instrumentation*, 2009.

[25] Lars O. S. Noehte on behalf of the Tracker Group of the CMS collaboration. CMS Phase-1 pixel detector refurbishment during LS2 and readiness towards the LHC Run 3. *Journal of Instrumentation*, 2022.

[26] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.

[27] M. Folin, V. Sharyy, J.-P. Bard, M. Korzhik, and D. Yvon. Scintillating properties of today available lead tungstate crystals. *Journal of Instrumentation*, 16, 2021.

[28] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[29] Adriaan H. Heering. CMS HCAL Hybrid Photodiode Design and Quality Assurance Stations. 2002.

[30] D Abbaneo et al. Studies on the upgrade of the muon system in the forward region of the CMS experiment at LHC with GEMs. *JINST*, 9:C01053, 2014.

[31] Fabio Sauli. The gas electron multiplier (gem): Operating principles and applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 805:2–24, 2016. Special Issue in memory of Glenn F. Knoll.

[32] Fabio Sauli. *Micro-pattern Gaseous Detectors: Principles Of Operation And Applications*. World Scientific, 2020.

[33] G. Apollinari, O. Bruening, T. Nakamoto, and L. Rossi. High Luminosity Large Hadron Collider HL-LHC. *arXiv*, 2017.

[34] P. Paolucci, R. Hadjiiska, L. Litov, B. Pavlov, P. Petkov, A. Dimitrov, K. Beernaert, A. Cimmino, S. Costantini, G. Guillaume, J. Lellouch, A. Marinov, A. Ocampo, N. Strobbe, F. Thyssen, M. Tytgat, P. Verwilligen, E. Yazgan, N. Zaganidis, A. Aleksandrov, V. Genchev, P. Iaydjiev, M. Rodozov, M. Shopova, G. Sultanov, Y. Ban, J. Cai, Z. Xue, Y. Ge, Q. Li, S. Qian, C. Avila, L.F. Chaparro, J.P. Gomez, B. Gomez Moreno, A.F. Osorio Oliveros, J.C. Sanabria, Y. Assran, A. Sharma, M. Abbrescia, A. Colaleo, G. Pugliese, F. Loddo, C. Calabria, M. Maggi, L. Benussi, S. Bianco, S. Colafranceschi, D. Piccolo, C. Carrillo, O. Iorio, S. Buontempo, P. Vitulo, U. Berzano, M. Gabusi, Minho Kang, Kyong Sei Lee, Sung Keun Park, Seungsu Shin, Min Suk Kim, Hyun Kwan Seo, Junghwan Goh, and Young-Il Choi. CMS Resistive Plate Chamber overview, from the present system to the upgrade phase I. RPC project overview, from the present system to the upgrade. *JINST*, 8:P04005, 2013. preliminary entry.

[35] R. Santonico and R. Cardarelli. Development of resistive plate counters. *Nuclear Instruments and Methods in Physics Research*, 187(2):377–380, 1981.

[36] M. Abbrescia, V. Peskov, and P. Fonte. *Resistive gaseous detectors: Designs, performance, and perspectives.* John Wiley & Sons., 2018.

[37] Wikipedia. Townsend discharge.

[38] A Gelmi, R Guida, and B Mandelli. Gas mixture quality studies for the CMS RPC detectors during LHC Run 2. *JINST*, 16(04):C04004, 2021.

[39] R Aly, J Eysermans, A Samalan, M Tytgat, M El Sawy, G A Alves, F Marujo, E A Coelho, E M Da Costa, H Nogima, A Santoro, S Fonseca De Souza, D De Jesus Damiao, M Thiel, K Mota Amarilo, M Barroso Ferreira Filho, A Aleksandrov, R Hadjiiska, P Iaydjiev, M Rodozov, M Shopova, G Soultanov, A Dimitrov, L Litov, B Pavlov, P Petkov, A Petrov, E Shumka, S J Qian, H Kou, Z A Liu, J Zhao, J Song, Q Hou, W Diao, P Cao, C Avila, D Barbosa, A Cabrera, A Florez, J Fraga, J Reyes, Y Assran, M A Mahmoud, Y Mohammed, I Crotty, I Laktineh, G Grenier, M Gouzevitch, L Mirabito, K Shchablo, I Bagaturia, I Lomidze, Z Tsamalaidze, V Amoozegar, B Boghrati, M Ebraimi, M Mohammadi Najafabadi, E Zareian, M Abbrescia, G Iaselli, G Pugliese, F Loddo, N De Filippis, D Ramos, W Elmetenawee, S Leszki, I Margjeka, D Paesani, L Benussi, S Bianco, D Piccolo, S Meola, S Buontempo, F Carnevali, L Lista, P Paolucci, F Fienga, A Braghieri, P Salvini, P Montagna, C Riccardi, P Vitulo, E Asilar, J Choi, T J Kim, S Y Choi, B Hong, K S Lee, H Y Oh, J Goh, I Yu, C Uribe Estrada, I Pedraza, H Castilla-Valdez, A Sanchez-Hernandez, R L Fernandez, M Ramirez-Garcia, E Vazquez, M A Shah, N Zaganidis, A Radi, H Hoorani, S Muhammad, A Ahmad, I Asghar, W A Khan, and F Torres Da Silva De Araujo. Latest results of Longevity studies on the present CMS RPC system for HL-LHC phase. *Nucl. Instrum. Methods Phys. Res., A*, 1055:168452, 2023.

[40] Giovanni Polese. The detector control systems for the cms resistive plate chamber. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 3228–3232, 2008.

[41] Antonio Pich. *The Standard Model of Electroweak Interactions.* arXiv, 2012.

[42] Peter Skands. *Introduction to QCD.* CERN, 2017.

[43] Wikipedia. ALEPH experiment.

[44] Aron J. Bekman and Louk Rademaker and Jasper van Wezel. *An Introduction to Spontaneous Symmetry Breaking.* arXiv, 2019.

[45] Michael E. Peskin and Daniel V. Shroeder. *An Introduction To Quantum Field Theory.* CRC Press, 2019.

[46] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024.

[47] D0 Collaboration. Measurement of B_s mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$. *Phys. Rev. Lett.*, 101:241801, 2008.

[48] 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 8 fb^{-1} of $p\overline{p}$ collisions. *Phys. Rev. D*, 85:032006, 2012.

[49] CDF Collaboration. First flavor-tagged determination of bounds on mixing-induced CP violation in $B_s^0 \to J/\psi\phi$. *Phys. Rev. Lett.*, 100:161802, 2008.

[50] 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. *Phys. Rev. D*, 85:072002, 2012.

[51] CDF Collaboration. Measurement of the bottom-strange meson mixing phase in the full CDF data set. *Phys. Rev. Lett.*, 109:171802, 2012.

[52] 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. *JHEP*, 12:072, 2012.

[53] 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. *Phys. Rev. D*, 90:052007, 2014.

[54] 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. *JHEP*, 12:072, 2016.

[55] ATLAS Collaboration. Measurement of the CP-violating phase ϕ_s in $B_s^0 \to J/\psi\phi$ decays in ATLAS at 13 TeV. *Eur. Phys. J. C*, 81(4):342, 2021.

[56] 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. *Phys. Lett. B*, 757:97, 2016.

[57] CMS Collaboration. Measurement of the CP-violating phase ϕ_s in the $B_s^0 \to J/\psi\phi \to \mu^+\mu^- K^+K^-$ channel in proton-proton collisions at $\sqrt{s} =$ 13 TeV. *Physics Letters B*, 816:136188, 2021.

[58] LHCb Collaboration. Measurement of the CP-violating phase ϕ_s in $\bar{B}_s \to J/\psi\pi^+\pi^-$ decays. *Phys. Lett. B*, 713:378, 2012.

[59] LHCb Collaboration. Measurement of CP violation and the B_s meson decay width difference with $B_s^0 \to J/\psi K^+K^-$ and $B_s^0 \to J/\psi\pi^+\pi^-$ decays. *Phys. Rev. D*, 87:112010, 2013.

[60] LHCb Collaboration. Measurement of the CP-violating phase ϕ_s in $\bar{B}_s^0 \to J/\psi\pi^+\pi^-$ decays. *Phys. Lett. B*, 736:186, 2014.

[61] LHCb Collaboration. Precision measurement of CP violation in $B_s^0 \to J/\psi K^+K^-$ decays. *Phys. Rev. Lett.*, 114:041801, 2015.

[62] LHCb Collaboration. Updated measurement of time-dependent CP-violating observables in $B_s^0 \to J/\psi K^+K^-$ decays. *Eur. Phys. J. C*, 79(8):706, 2019.

[63] R. Aaij et al. Improved measurement of *cp* violation parameters in $B_s^0 \to j/\psi K^+K^-$ decays in the vicinity of the $\phi(1020)$ resonance. *Phys. Rev. Lett.*, 132:051802, Feb 2024.

[64] The CKMfitter Group. Predictions of selected flavour observables within the Standard Model. *Phys. Rev. D*, 84:033005, 2011.

[65] UTfit Collaboration. New utfit analysis of the unitarity triangle in the cabibbo-kobayashi-maskawa scheme, 2022.

[66] Alexander Lenz and Gilberto Tetlalmatzi-Xolocotzi. Model-independent bounds on new physics effects in non-leptonic tree-level decays of B-mesons. *JHEP*, 07:177, 2020.

[67] Y. Amhis, Sw. Banerjee, E. Ben-Haim, E. Bertholet, F. U. Bernlochner, M. Bona, A. Bozek, C. Bozzi, J. Brodzicka, V. Chobanova, M. Chrzaszcz, S. Duell, U. Egede, M. Gersabeck, T. Gershon, P. Goldenzweig, K. Hayasaka, D. Johnson, M. Kenzie, T. Kuhr, O. Leroy, A. Lusiani, H.-L. Ma, M. Margoni, K. Miyabayashi, R. Mizuk, P. Naik, T. Nanut Petrič, A. Oyanguren, A. Pompili, M. T. Prim, M. Roney, M. Rotondo, O. Schneider, C. Schwanda, A. J. Schwartz, J. Serrano, A. Soffer, D. Tonelli, P. Urquijo, and J. Yelton. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021. *Phys. Rev. D*, 107:052008, 3 2023.

[68] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. *arXiv*, 2017.

[69] B. W. Silverman. Kernel density estimation using the fast fourier transform. *Journal of the Royal Statistical Society*, Series C (Applied statistics), 1982.

[70] S. J. Sheather and M. C. Jones. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991.

[71] Artur Gramacki and Jaroslaw Gramacki. Fft-based fast computation of multivariate kernel density estimators with unconstrained bandwidth matrices. *arXiv*, 2016.

[72] CERN ROOT framework. TKDE class reference.

[73] Ulrich Oberst. The Fast Fourier Transform. *SIAM Journal on Control and Optimization*, 2007.

[74] Stanislaw Weglarczyk. Kernel density estimation and its application. *ITM Web of Conferences*, 2018.

[75] University of Padova HEPH group and University of Sofia HEPH group. Cms-bsjpsiphi-fullrun2.

[76] Ilka Antcheva, Maarten Ballintijn, Bertrand Bellenot, Marek Tomasz Biskup, Rene Brun, Nenad Buncic, Philippe Canal, Diego Casadei, Olivier Couet, Valery Fine, Leandro Franco, Gerardo Ganis, Andrei Gheata, David González Maline, Masaharu Goto, Jan Iwaszkiewicz, Anna Kreshuk, Diego Marcos Segura, Richard Maunder, and Lorenzo Moneta. Root - a c++ framework for petabyte data storage, statistical analysis and visualization. *Comput. Phys. Commun.*, 180:2499–2512, 2009.

[77] CERN ROOT. RooFit library.

[78] Kesar Singh and Minge Xie. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 1979.

[79] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr. Advanced event reweighting using multivariate analysis. *Journal of Physics, Conference Series*, 368, 2012.

[80] Yann Coadu. Boosted Decision Trees. *arXiv*, 2022.

[81] CERN ROOT framework. Machine learning with ROOT.

[82] Muriel Pivk and Francois R. le Diberder. sPlot: a statistical tool to unfold data distributions. *arXiv*, 2005.

[83] Zachary Baldwin. A multidimensional, event-by-event, statistical weighting procedure for signal to background separation. *CHEP*, 2023.

[84] Wikipedia. Johnson's S_U distribution.

[85] Anastasios Bikos, Panagiotis E. Nastou, Georgios Petroudis, and Yannis C. Stamatiou. Random Number Generators: Principles and Applications. *Cryptography*, 2023.

[86] Wikipedia. Crystal Ball function.

[87] CERN ROOT framework. RooSimultaneous class reference.

[88] Richard Nisius. Blue: Combining correlated estimates of physics observables within root using the best linear unbiased estimate method. *SoftwareX*, 11:100468, 2020.

[89] Richard Nisius. On the combination of correlated estimates of a physics observable. *The European Physical Journal C*, 74(8):3004, Aug 2014.

[90] The CMS Collaboration. Measurement of time-dependent CP violation in $B_s^0 \to J/\psi\,\phi(1020)$ decays with the CMS detector. Technical report, CERN, Geneva, 2024.

[91] Annette J. Dobson and Adrian G. Barnett. *An Introduction to Generalized Linear Models*. Chapman & Hall, 2018.

[92] Umberto Michelucci. An introduction to autoencoders. *arXiv*, 2022.

[93] OKD. OKD: Deploy at scale on any Infrastructure.

[94] E. Shumka et al. Machine Learning based tool for CMS RPC currents quality monitoring. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1054:168449, 2023.

[95] S. Di Guida, G. Govi, M. Ojeda, A. Pfeiffer, R Sipos, and on behalf of the ATLAS Collaboration. The cms condition database system. *Journal of Physics: Conference Series*, 664(4):042024, dec 2015.

[96] W. Karimeh. The CMS Detector Control Systems Archiving Upgrade. In *Proc. 19th Int. Conf. Accel. Large Exp. Phys. Control Syst. (ICALEPCS'23)*, number 19 in International Conference on Accelerator and Large Experimental Physics Control Systems, pages 1022–1024. JACoW Publishing, Geneva, Switzerland, 02 2024.

[97] André, Jean-Marc, Behrens, Ulf, Branson, James, Brummer, Philipp, Cittolin, Sergio, Diego Da, Silva Gomes, Georgiana-Lavinia, Darlea, Deldicque, Christian, Demiragli, Zeynep, Dobson, Marc, Doualot, Nicolas, Erhan, Samim, Fulcher, Jonathan Richard, Gigi, Dominique, Gladki, Maciej, Glege, Frank, Gomez-Ceballos, Guillelmo, Hegeman, Jeroen, Holzner, André, Janulis, Mindaugas, Lettrich, Michael, Mečionis, Audrius, Meijers, Frans, Mommsen, Remigius K, Morovic, Srecko, O'Dell, Vivian, Orsini, Luciano, Papakrivopoulos, Ioannis, Paus, Christoph, Petrova, Petia, Petrucci, Andrea, Pieri, Marco, Rabady, Dinyar, Rácz, Attila, Rapševičius, Valdas, Reis, Thomas, Sakulin, Hannes, Schwick, Christoph, Šimelevičius, Dainius, Stankevičius, Mantas, Cristina, Vazquez Velez, Wernet, Christian, and Zejdl, Petr. Presentation layer of cms online monitoring system. *EPJ Web Conf.*, 214:01044, 2019.

[98] Marcin Konecki. The rpc based trigger for the cms experiment at the lhc. *Journal of Instrumentation*, 9:C07002 – C07002, 2014.

[99] Tomasz Bawej, Ulf Behrens, James Branson, Olivier Chaze, Sergio Cittolin, G. L. Darlea, Christian Deldicque, Marc Dobson, Aymeric Dupont, S. Erhan, Andrew Forrest, Dominique Gigi, Frank Glege, Guillelmo Gomez-Ceballos, Robert Gomez-Reino, Jeroen Hegeman, A. Holzner, L. Masetti, Frans Meijers, Emilio Meschi, Remigius K. Mommsen, S. Morovic, Carlos Nunez-Barranco-Fernandez, Vivian O'dell, Luciano Orsini,

Christoph Paus, Andrea Petrucci, Marco Pieri, Attila Rácz, Hannes Sakulin, Christoph Schwick, Benjamin Stieger, Konstanty Sumorok, Jan Veverka, and Petr Zejdl. The new cms daq system for run-2 of the lhc. *IEEE Transactions on Nuclear Science*, 62:1099–1103, 2014.

[100] E.F.V. Valencia O. Miguel Colin*, S.C. Moreno and C.O. Barrera. Study of the CMS RPC detector performance in high radiation background conditions. In *Proceedings of Science*, 2018.

[101] R. Guida. GIF++: The new CERN Irradiation Facility to test large-area detectors for HL-LHC. In *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4, 2015.