BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

Facultad de Ciencias Físico Matemáticas
Maestría en Ciencias Física Aplicada

A search for anomalous H+X in association with two forward protons

Tesis
para obtener el título de
Maestra en Ciencias Física Aplicada

Presenta
Lic. Nadia Mariana Leal Reyes

Asesor: Dr. Humberto Salazar Ibargüen
Asesora: Dra. María Isabel Pedraza Morales

Puebla, Pue.

Diciembre 2024

Abstract

This Master's thesis analyzes the associated production of a Higgs boson (H) with additional, unspecified particles (X), specifically in events where two forward protons are tagged. The study is based on proton-proton (pp) collision data at 13 TeV, gathered in 2018 with the CMS detector and the Precision Proton Spectrometer (PPS). A focal point of this analysis is the contribution of Quantum Chromodynamics (QCD) background events, which is crucial for isolating potential signals of new physics. This contribution is estimated through a data-driven method, and the results are presented.

Contents

Listing of figures

To my mom, the light of my life.

Acknowledgments

To my loved ones, you have always been by my side in the good and bad times, and I wouldn't have made it this far without you. My mom, in particular, I have no words to thank you for everything you have done for me; I love you.

To the PPSX analysis team, who received me with open arms and guided me through this journey. Thank you for all your feedback, insights, jokes, advice, and good memories. Working with you and getting to know you all was a pleasure.

To my advisors Isabel Pedraza and Humberto Salazar. I have learned so much from you and appreciate the trust and opportunities you have given me. Thank you for your support and guidance; they are deeply appreciated.

Lastly, to the little me who was just curious about everything in life and dreamed of doing physics. You can be proud; this journey was your dream, and you've made it a reality.

0
Introduction

0.1 The Standard Model

The Standard Model of particle physics is a well-established theoretical framework that describes the fundamental particles and forces that constitute the universe, excluding gravity. It classifies all known subatomic particles into two categories: fermions, which make up matter, and bosons, which mediate forces. The fermions include quarks and leptons, while the bosons include the photon, W and Z bosons, gluons, and the Higgs boson.

Figure 1 illustrates the SM's particle contents classified by their corresponding roles. It consists of fermions and their anti-particles, gauge bosons and the Higgs boson. Fermions are spin 1/2 particles and are further divided into leptons and quarks, while gauge bosons carry a spin of 1 and represent the mediators of the fundamental forces. The

Higgs boson is the only particle in the SM that has a spin of 0. Both leptons and quarks can be grouped into three generations, where the particles of each generation are heavier copies of the particles from the preceding generation. However, all visible matter that surrounds us contains almost exclusively first-generation fermions.

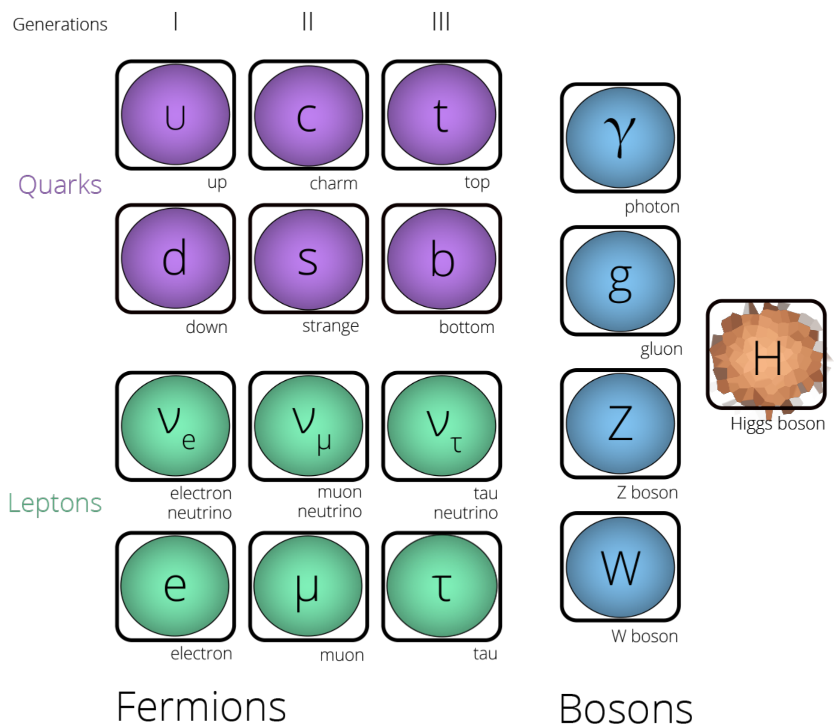

Figure 1: The Standard Model of Particle Physics. [1]

Based on the particles' properties, all particles are classified as fermions or bosons. Fermions have fractional spin values and follow the Fermi-Dirac statistics. Since they are separated into three families or generations based on their masses, the first family being the least massive and consisting of ordinary stuff, they are regarded as matter particles.

- Quarks: are the elementary fermions that interact strongly to make up hadrons and nuclear matter. They have color charges (red, green, or blue) in addition to electric charges. Up (u), down (d), charm (c), odd (s), top (t), and button (b) are the six varieties or flavors.

- Leptons: six elementary fermions that do not interact strongly, and three of them, the tau (τ), muon (μ), and electron (e), have electric charges. The other three, known as neutrinos, have neutral electric charge: tau (ν_τ), muon (ν_μ), and electron (ν_e).

On the other hand, bosons have integer spiral values and follow the Bose-Einstein statistics. They are referred to as norm (vector) bosons when they function as carriers of a fundamental interaction. These are the particles that make up the force fields. Gluons, W and Z bosons, and photons are the four norm bosons. They each correspond to three interactions: the strong interaction, the weak interaction, and the electromagnetic interaction, respectively. Lastly, a neutral scalar called the Higgs boson provides mass to all particles through the Higgs mechanism.

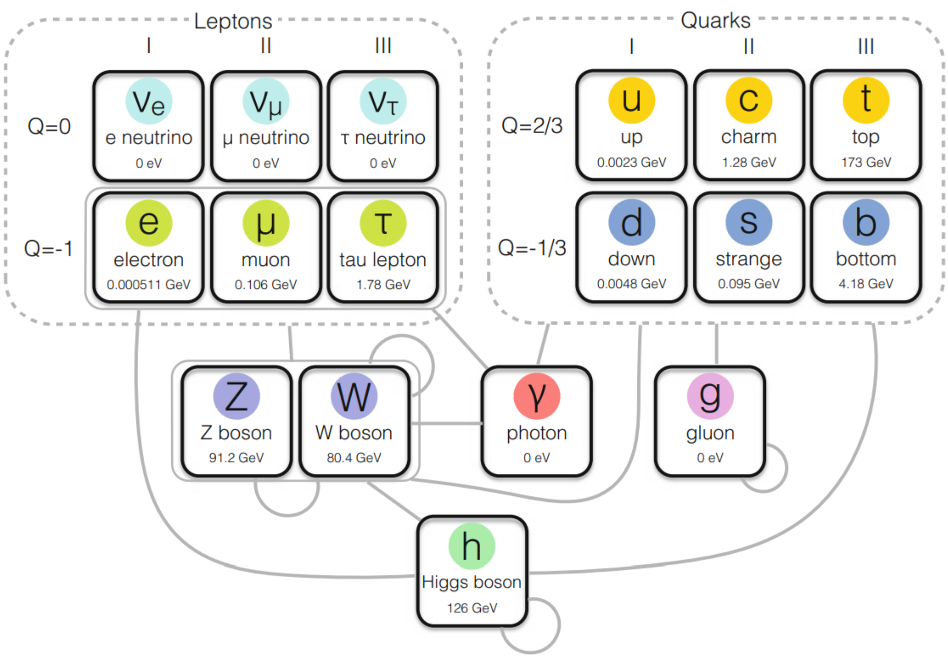

Figure 2: Observed Standard Model particles and their interactions. Each round square represents a mass eigenfield of the SM, with its name and mass indicated at the bottom. Particles and particle groups connected by a grey line interact directly. Particles connected to themselves exhibit self-interactions. [2]

The Standard Model is a relativistic quantum field theory (QFT) based on requiring local gauge invariance of the underlying Lagrange density $\mathcal{L}$, also referred to as Lagrangian. The gauge group of the SM is the non-abelian group:

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

where C represents the color charge, which is introduced in Quantum Chromodynamics, L stands for left-handed fields, and Y for the weak hypercharge.

0.1.1 Quantum Chromodynamics

The theory of quantum chromodynamics (QCD) describes the strong interaction between quarks and gluons, particles with color charge. In this case, the rotation of the quark color field is the local gauge transformation that requires the Lagrangian to be invariant. The quark color field can be expressed as

$$\psi^f = (\psi_R^f, \psi_G^f, \psi_B^f)^T, \tag{2}$$

where f represents the quark flavor, and the quark colors are represented by red (R), green (G), and blue (B). Moreover, for each color, there is an anti-color, and the RGB combination is color-neutral.

The symmetry group of QCD is the non-abelian SU(3) group and the resulting gauge transformation can then be written as

$$\psi^f \rightarrow e^{i\theta_a(x)\lambda_a}\psi^f, \tag{3}$$

where $a \in \{1, ...8\}$, λ_a corresponds to the eight Gell-Mann matrices that are the generators of $SU(3)$, and $\theta_a(x)$ is a real number for each component.

Gell-Mann matrices obey the commutation rule

$$[\lambda_\alpha, \lambda_\beta] = 2if_{\alpha\beta\gamma}\lambda_\gamma, \tag{4}$$

with $f_{\alpha\beta\gamma}$ being the completely antisymmetric structure constant.

Then, local gauge invariance is achieved using the covariant derivative

$$D_\mu = \partial_\mu + ig_s\lambda_a A_\mu^a, \tag{5}$$

where index a once more runs from 0 to 8, and A_μ^a describes the vector fields of the eight gluons.

On the other hand, g_s is a constant that quantifies the coupling strength and has a relation to α_s, the coupling constant, given by

$$g_s = \sqrt{4\pi\alpha_s} \tag{6}$$

Given that the field strength tensors $F^a_{\mu\nu}$ are defined as

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - 2g_s f_{abc} A^b_\mu A^c_\nu, \tag{7}$$

it is possible to evaluate the Lagrangian for QCD as

$$\mathcal{L}_{QCD} = i\bar{\psi}^f \gamma^\mu D_\mu \psi^f - m\bar{\psi}^f \psi^f - \frac{1}{4} F^{a\mu\nu} F^a_{\mu\nu} \tag{8}$$

$$\mathcal{L}_{QCD} = i\bar{\psi}^f \gamma^\mu \partial_\mu \psi^f - m\bar{\psi}^f \psi^f - \frac{1}{4} F^{a\mu\nu} F^a_{\mu\nu} - g_s \left(\bar{\psi}^f \gamma^\mu \lambda_a \psi^f\right) A^a_\mu. \tag{9}$$

This Lagrangian contains interaction terms between gluons because of the $F^a_{\mu\nu}$ structure, indicating that gluons and quarks have color charges. This can be attributed to the non-abelian nature of SU(3).

Moreover, the gluon self-interaction has a significant effect in that the running strong coupling constant α_s behaves in the following way with respect to Q^2

$$\alpha_s\left(Q^2\right) = \frac{12\pi}{(33 - 2n_f)\, ln\left(\frac{Q^2}{\Lambda^2_{QCD}}\right)}, \left(Q^2 \gg \Lambda_{QCD}\right), \tag{10}$$

where n_f is the number of active quark flavors.

It is noticeable that as energy transfer increases, α_s decreases. The term ”asymptotic freedom” refers to the limit at which Q^2 becomes extremely large. The strong force approaches zero in this limit, allowing for a perturbative analysis of QCD. On the other hand, as Q^2 decreases, the coupling constant increases, and the perturbative regime is passed at an energy often represented by Λ_{QCD}. This energy regime causes quarks and gluons to organize into color-neutral clusters referred to as hadrons, and this effect is known as ”color confinement.”

0.2 Physics motivation

So far, the Standard Model successfully unifies the electromagnetic, weak, and strong nuclear forces under a common theoretical structure, explaining how particles interact via these forces. However, gravity, dark matter, and dark energy still need to be included, leaving some questions in particle physics unanswered. Despite these limitations, the Standard Model has been remarkably successful in predicting experimental results, and it remains a cornerstone of our understanding of the quantum world.

In particular, regarding dark matter, it is now known that dark matter particles should be stable, fit the measured density of DM, and decouple from plasma and radiation at least before the beginning of the matter-dominated stage [3] and not interact electromagnetically. Some Models predict DM particles to interact with SM states via exchanging new mediators [4, 5, 6, 7], which can carry spin one such as a new Z' gauge boson, or spin 0 such as an additional Higgs boson. Moreover, these new mediators can lead to signals with direct [8, 9] and indirect [10, 11] detection.

One of the most important discoveries in particle physics has been the Higgs boson discovery by the Compact Muon Solenoid (CMS), and the A Toroidal LHC ApparatuS (ATLAS) collaborations at the European Organization for Nuclear Research (CERN) at the Large Hadron Collider (LHC) [12, 13] in July 2012. This discovery completed the scientific description of the visible matter in the Universe in the Standard Model (SM). However, this leaves many questions requiring physics beyond the Standard Model, including the origin of this matter and the nature of the invisible Dark Matter (DM) responsible for the formation and continued existence of structures in the Universe, as previously discussed. Answering these questions is the primary motivation for the ongoing Large Hadron Collider (LHC) research program.

The discovery of the Higgs boson is an outstanding achievement. It is unique because it is the first fundamental scalar particle discovered; before, scalars were bound states, and the fundamental particles discovered thus far were either spin-1/2 fermions or spin-1 vector bosons. The Higgs boson is the anticipated visual manifestation of the Brout–Englert–Higgs (BEH) mechanism [14, 15, 16], enabling particle mass production in a gauge-invariant manner. Therefore, this finding ultimately established the gauge invariance paradigm as the sub-nuclear environment's guiding principle at the smallest distances and most enormous energies ever studied in a laboratory.

Now that its existence has been proven, the focus is on its properties and the possibility of new phenomena or dynamics.

This work intends to describe a generic search for the central exclusive production of a neutral boson and an additional massive particle without any assumptions about its decay properties, except that its decay width can be considered narrow enough to produce a resonant mass peak.

Models with two doublet Higgs sectors further expanded with SM gauge singlets and beyond the Standard Model scenarios with extended Higgs sectors, like the two-Higgs-doublet model (2HDM) [17], are the primary sources of motivation for this research. Given that they allow new (pseudo)scalar states at various mass scales and broad phenomenology, the latter classes of the model are intriguing for this search.

Since theoretical and electroweak precision tests suggest a degenerate spectrum for the extra Higgs bosons, explaining these novel (pseudo)scalar states in the traditional 2HDM would be more challenging. The new scalar field can be produced via gluon fusion and photon fusion. In both cases, the effective coupling of the scalars with gluon and photons is due to triangle loop diagrams involving SM fermions.

Adding new vector-like fermions coupled with the Higgs sector can improve the effective couplings while suppressing the predicted cross-section for a TeV scale resonance. It may be possible to generate effective couplings of the new scalars with the gauge bosons and improve the photon fusion cross-section concerning the gluon fusion by adequately selecting the gauge quantum numbers of the new exotic states. Finally, there are intriguing links between this configuration and DM physics. [18].

In this framework, a new scalar particle could become the link to the Dark Matter sector (portal), coupled both with invisible particles and heavy SM particles (H, Z, top).

The Precision Proton Spectrometer (PPS) measures all the properties of the protons, leading to photon fusion and closing the kinematics beyond the event. Once the interaction vertex is identified and measured precisely, one obtains the relative mass produced by the collision. On the other hand, the central CMS detector reconstructs the visible particles emerging from the decay vertex, and the kinematics of a possible invisible particle can be resolved by balancing all the other measured variables.

On the other hand, the Branching ratio is the percentage likelihood that a particle will decay in a particular manner. The largest Higgs branching ratio is the decay into two b quarks, about 58%, which is why that is studied in this research. Even if the poor resolution of jet reconstruction reduces the capability of identifying the missing particle, the scalar particle kinematics is well reconstructed by PPS, thus providing an exceptional search tool for this phenomenon. Figure 3 shows the Higgs branching ratios.

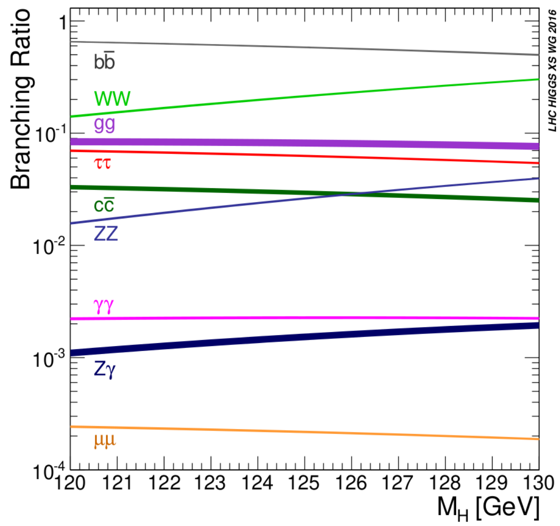

Figure 3: The Higgs boson branching ratio for various decay modes as a function of the Higgs mass. [19]

The main physics interest in PPS is the study of the central exclusive production (CEP) processes ($pp \rightarrow p \oplus X \oplus p$). In this scenario, the protons may remain intact, PPS can measure their kinematics, and the CMS detector can identify the X state in hard scattering events. Colorless exchanges, a photon-photon interaction, or double Pomeron can drive these processes. CEP processes are interesting as they provide very clean experimental signatures, given that they are less contaminated by background noise than other processes. This makes CEP particularly useful for studying rare and new physics phenomena. A diagram for these CEP processes is illustrated in figure 4. Moreover, relevant results in this study area using the Precision Proton Spectrometer have been published. In particular, a Z/γ + X Search. [20]

This analysis focuses on the central production of an H boson and an unknown particle X, as described in Figure 5. Two protons are present in the initial state, and at IP5, they interact electromagnetically by exchanging a photon. Photon-fusion results in the generation of a scalar resonance. Subsequently, the scalar particle decays into a Higgs boson and an unknown particle called X. The Beyond Standard Model (BSM) extends the Higgs sector, where this interaction finds theoretical roots. The CMS and PPS detectors can be used to calculate the mass of the scalar as $m_s = \sqrt{s\xi_1\xi_2}$

Compared to gluon fusion, the photon fusion mechanism is more likely. The coupling constant takes on smaller values at high masses, and there are two more vertices in gluon fusion than in photon fusion. Moreover, photon-

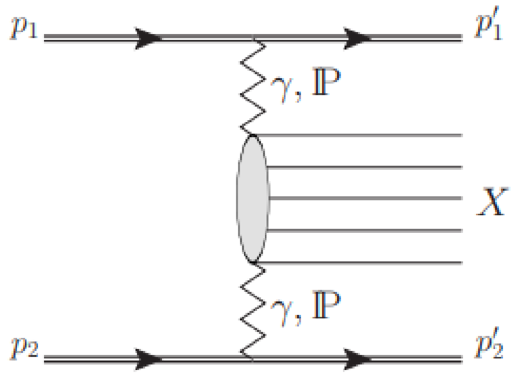

Figure 4: Diagram for the central exclusive production, $pp \to p \oplus X \oplus p$, which is characterized by large pseudorapidity gaps ($\oplus$). The figure is taken from Ref. [21]

fusion wins out over gluon-fusion in this particular scenario due to Sudakov suppression. Therefore, the photon fusion scenario is considered and investigated.

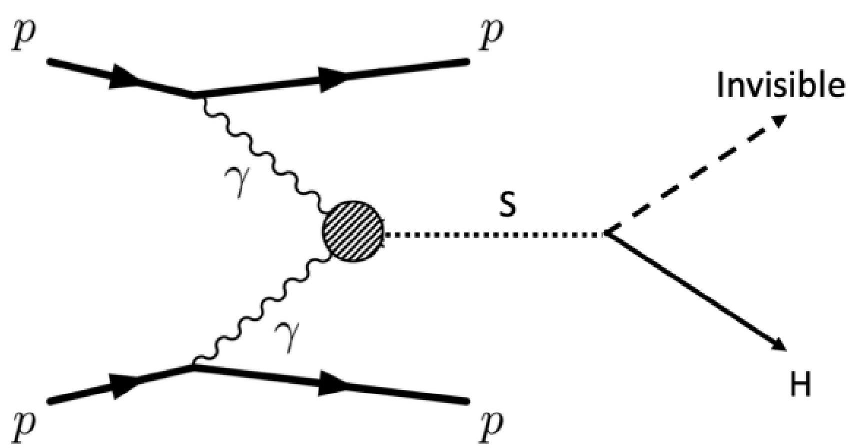

Figure 5: Feynman diagram of a two-photon production of an H boson with an additional, unknown particle giving rise to a missing mass $m_{miss} = m$. Note that the production mechanism does not have to proceed through photon exchange. Other colorless exchange mechanisms are also allowed. However, at such a high energy scale, EWK processes are enhanced, while QCD-based colorless exchanges are expected to be suppressed.

1

Experimental Setup

The European Council for Nuclear Research, known by its abbreviation CERN, is a renowned research facility founded in 1954. It uses enormous, intricate particle accelerator facilities to do ground-breaking research in high-energy particle physics. The accelerated phases of the proton bunches are described in figure 1.1, which depicts CERN's accelerator complex. The Large Hadron Collider (LHC) is where proton bunches are fed after passing via the Linear Accelerator (LINAC 2), Proton Synchrotron Booster (PSB), Proton Synchrotron (PS), and Super Proton Synchrotron (SPS), and are then accelerated to their highest possible energy and collide.

The primary accelerator at CERN and the most potent particle accelerator available today is the Large Hadron

Collider (LHC). The LHC is a ring of superconducting magnets that is 27 kilometers long. By providing opposing magnetic fields, the protons or heavy ions that make up the beams inside the LHC are accelerated in opposite directions in two different vacuum tubes. Four particle detectors—the Compact Muon Solenoid (CMS), the Large Ion Collider Experiment (ALICE), the Toroidal LHC Apparatus (ATLAS), and the Large Hadron Collider Beauty (LHCb)—are situated at the four points where the beams meet, sharing a beam pipe.

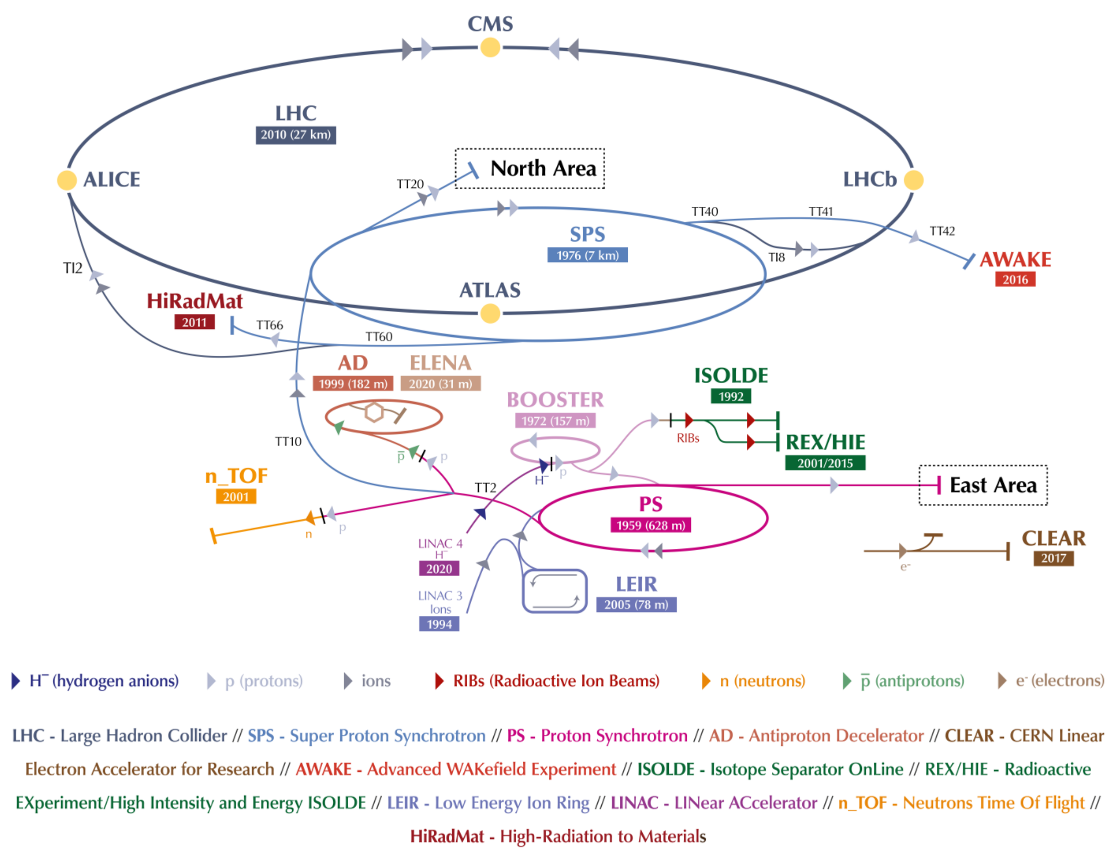

Figure 1.1: Scheme of CERN's accelerator facility. The figure is taken from Ref. [22]

From a physics point of view, high-energy and intense beams capable of producing a lot of collisions are ideal for studying processes with small cross sections or the SM with high precision. Moreover, the collision energy (also known as the center-of-mass energy) can be directly connected to the beam energy using the formula $\sqrt{s} = 2E_{beam}$ for symmetric beams.

On the other hand, Luminosity is a parameter that captures both the collision frequency and beam intensity. At the Interaction Point, two particle beams collide, and the luminosity can be computed by assuming that both beams have Gaussian distributions. The distributions' centers overlap to determine the luminosity, with negligible tail contributions.

Then, it is possible to write the luminosity for two colliding beams as

$$L = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{1.1}$$

Where σ_x and σ_y are the transverse beam size and the longitudinal beam size, respectively, while N_1 and N_2 are the number of particles in the two bunches, N_b is the number of bunches and f is the revolution frequency. [23]

If evaluating this luminosity over time is of interest, the integrated luminosity can be used.

$$L_{int} = \int_0^T L(t')\,dt' \tag{1.2}$$

Moreover, to get the number of events, it is possible to do so by doing the product of the cross-section and the integrated luminosity.

$$L_{int} \cdot \sigma = number\ of\ events \tag{1.3}$$

Another important notion is that to prevent undesired collisions outside of the designated collision locations, beams in colliders with several particle bunches must clash at a crossing angle θ_c between each other (see Figure 1.2).

If it is desired to take this angle into consideration for the calculation of the luminosity, then it is necessary to use a luminosity reduction factor S [23]. Defined as

$$S = \frac{1}{\sqrt{1+\left(\frac{\sigma_x}{\sigma_s}tan\frac{\phi}{2}\right)^2}}\frac{1}{\sqrt{1+\left(\frac{\sigma_s}{\sigma_x}tan\frac{\phi}{2}\right)^2}} \tag{1.4}$$

.

Then, the luminosity is given by

$$L = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \cdot S \tag{1.5}$$

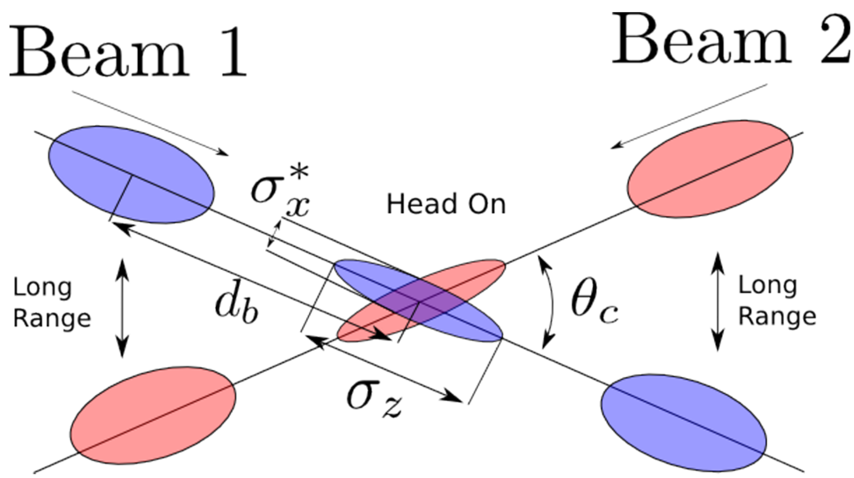

Figure 1.2: Diagram showing the collision of the beams at the LHC conditions.[24]

.

The LHC initiates what is referred to as a ”fill” each time beams are injected into the machine. That fill comes to an end when the beams are discarded. Every Fill is assigned a distinct number. The detectors gather the data necessary for a high-quality study when some of these fills are deemed stable and suitable for physics. CMS (and all the other experiments) determine whether to begin recording data as collisions occur in the LHC. A new ”run” is launched each time the start button is pressed, and a distinct number is assigned. Although there are several reasons why the Run can end, such as if the LHC dumps the fill, the acquisition stops typically due to a problem in one of the detector’s electrical channels. This explains why the data provided by the LHC and collected by the detectors are typically different. A fresh run number is assigned upon restarting the acquisition.

Specifically, Run 2 encompasses the years 2015–2018. Every Run is composed of many bunches, each with a different time, and each fill has multiple components. The bunch spacing was reduced during Run 2, resulting in a 40MHz collision rate. About 600 million interactions occur during each fill. Fills are made up of about 1014 protons. This work focuses on 2018 conditions. During that year, the maximum number of bunches reached was 2556, and the average time fill was 8.3 hours. Moreover, the integrated luminosity for this year was $55.7fb^{-1}$.

On the other hand, pileup is a phenomenon that occurs when two bunches cross and there are many proton-proton collisions. The number of pileup interactions is denoted by μ, but it is also frequently provided as a time-averaged quantity $< \mu >$. Figure 1.3 displays the pileup interactions throughout Run 2. Moreover, in 2018, the average pileup was 37.

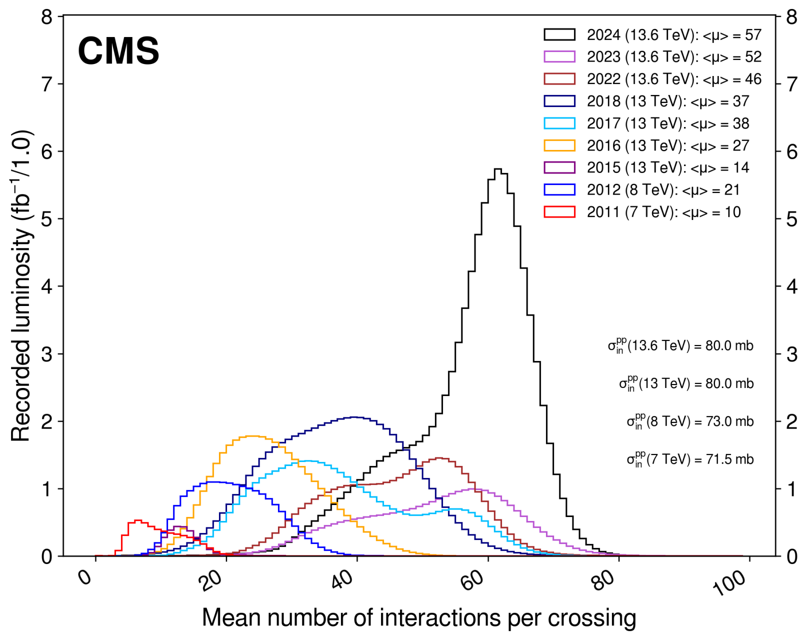

Figure 1.3: Ofifical CMS plot showing mean number of interaction per crossing for each year.

1.1 Compact Muon Solenoid

A general-purpose detector called the Compact Muon Solenoid (CMS)[25] was created to investigate physics at energies ranging up to many TeV. The multi-layer cylinder that makes up the detector's main body is 21 meters long and 15 meters in diameter. It weighs 14,000 tons in total. A silicon tracker, a superconducting solenoid, a hadron calorimeter (HCAL), an electromagnetic calorimeter (ECAL), and a muon spectrometer comprise the detector. A broad overview of the detector's design is shown in Figure 1.4.

- **Tracker** [27]. CMS' innermost subdetector functions under challenging conditions of high occupancy and high radiation levels. The tracker is made entirely of silicon-based sensors to ensure effective performance in this setting. This allows for the development of a very granular detector with quick reaction times. Two components compose the CMS tracker, pixel and strip detectors, that reconstruct the trajectories of charged particles.

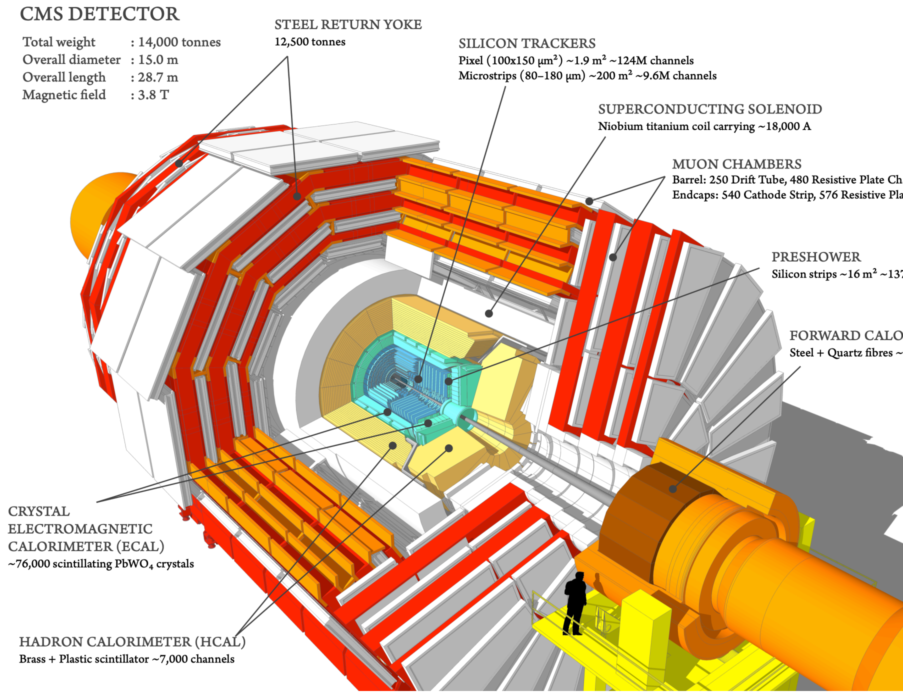

Figure 1.4: Overview of the Compact Muon Solenoid's detector structure. Its most relevant components are highlighted and a comparison of average people with respect to the detector for a better comprehension of the size of the experiment are shown. The figure is taken from Ref. [26].

- **ECAL** [28]. This subdetector monitors the energy and position of photons and electrons created in particle collisions. It encircles the tracking system and is made up of a single layer of lead tungstate ($PbWO_4$) crystals, which have a high energy resolution, quick response time, and can endure high doses of ionizing radiation.

- **HCAL** [29]. A set of calorimeters designed for absorbing and measuring the energy of massive amounts of hadrons. It is a sampling calorimeter with alternating layers of absorbing and sparkling materials, unlike the ECAL, which is a homogenous calorimeter. The layers at large $|\eta|$ are made of steel and quartz fiber, whereas the bulk of the HCAL volume comprises plastic scintillator tiles and brass absorber plates.

- **Superconducting Solenoid** [30]. It has an internal diameter of 6m and is 13 m long. Superconducting technology is used to create such a strong magnetic field. Liquid helium is used to cool the niobium-titanium (NbTi) alloy wires, making the solenoid magnet temperatures close to absolute zero. Because of the superconducting condition that these low temperatures provide, the magnet can conduct electricity without encountering any resistance. As a result, a powerful and steady magnetic field of 3.8 T is created.

- **Muon Spectrometer** [31]. Constituted by Drift Tubes (DTs), Resistive Plate Chambers (RPCs), Cathode Strip Chambers (CSCs), and the more recently added Gas Electron Multiplier (GEM) detectors that can measure muon tracks. The muon system detectors cover large areas and are located the furthest from the interaction point.

The CMS experiment's sections where the muons, electrons, charged hadrons, neutral hadrons, and photons are tagged and have their tracks recreated on the various CMS detector components are depicted in Figure 1.5.

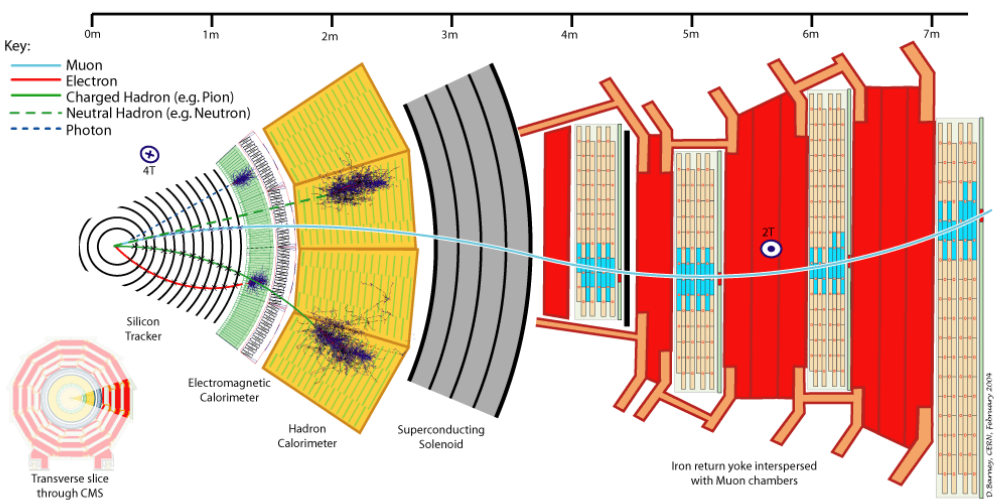

Figure 1.5: CMS cross-sectional distribution The figure is taken from Ref.

The right-handed coordinate system is used in the CMS experiment. The system's origin is at the interaction point in the middle of the detector. The y-axis extends upward, perpendicular to the LHC plane, while the x-axis points toward the center of the LHC. The azimuthal angle (ϕ) is defined in the transverse plane, which is spanned by the x and y axes. As seen above, the z-axis points along the anticlockwise beam's direction. The positive z-axis is used to measure the polar angle θ.

Rapidity is defined as

$$y = \frac{1}{2} ln \frac{E + p_z}{E - p_z}, \tag{1.6}$$

where E is the energy and p_z is the momentum along the beam axis.

Pseudorapidity is defined as follows

$$\eta = -ln\left(tan\frac{\theta}{2}\right). \tag{1.7}$$

This variable is frequently used in practice instead of the polar angle θ and is equivalent to the rapidity in the relativistic limit. Figure 1.6 shows the CMS pseudorapidity coverage, maximum $\eta = 2.4$, which is essential in this analysis.

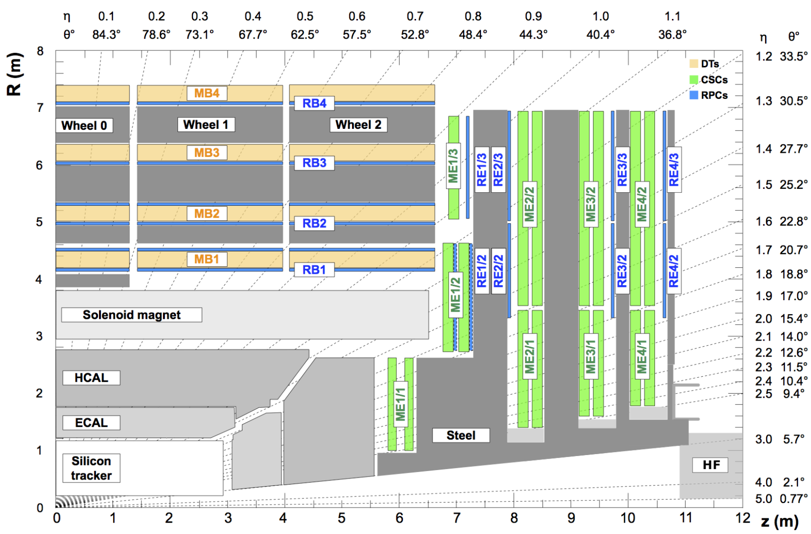

Figure 1.6: One CMS detector quadrant's longitudinal configuration. The illustration displays the RPC stations (RB1–RB4 and RE1–RE4, blue), four DT stations in the muon barrel, and four CSC stations in the muon endcap. [32]

Finally, the angular separation $\triangle R$ is relevant for describing the relative distance between particles and is given by

$$\triangle R = \sqrt{(\triangle\eta)^2 + (\triangle\phi)^2}. \tag{1.8}$$

1.1.1 Trigger

Nearly 40 million collisions per second (40 MHz) occur at the LHC. The quantity of data the various subdetectors produce for each collision once detected is roughly equivalent to what can be contained in a 1 MB file.

Fortunately, not every collision is recorded, as not all are interesting. Only the intriguing ones are preserved; above all, it is crucial to ensure that the ones of discovery quality are not missed. A trigger then is required to accomplish that.

Using detector information, an online trigger system is put in place to rapidly evaluate every event and determine whether it is worthwhile to preserve. It uses a two-level technique to lower event storage rates systematically.

- Level-1 trigger (L1T): hardware encodes algorithms to quickly reconstruct missing transverse energy, electrons, photons, muons, jets, and jet counters. Lowering the event rates from 40 MHz to 100 kHz is achieved by a global muon trigger (DT+CSC+RPC) and a global calorimeter trigger (ECAL+HCAL+HF).

  Field-Programmable Gate Arrays (FPGAs) are specialized hardware processing units used in constructing the L1T. It can function with a delay of roughly 3.8 μs because of these hardware components. To choose events, the L1T applies a trigger menu incorporating multiple algorithms implemented in the FPGAs. These algorithms are referred to as seeds. The seeds are instructed to accept events that fit specific criteria. Unless a prescale is used, the HLT accepts an event for further processing if it passes at least one seed in the trigger menu. A prescale determines the proportion of events selected from among all the events that a given seed accepts. When the prescale is set to N, one event is chosen for further processing for every N event that a seed accepts.

- High-level trigger (HLT): the HLT validates the events chosen by the level-1 trigger and processes them in computer farms. Its main goal is to significantly lower the event rates from 100 kHz to 100 Hz so they can be stored. Like the L1T, the HLTs select events from a trigger menu. This trigger menu is made up of HLT pathways, which are collections of processing methods carried out in a predetermined order. Multiple modules in each path replicate real-world objects and choose them based on preset standards. The tracker data is included, and the reconstruction methods and algorithms are comparable to those used for offline reconstruction.

  As a result, the HLT offers precise event information, enabling effective event selection. Online data quality monitoring (DQM) also uses the reconstruction results. The data acquisition system (DAQ) concentrates and stores the detector information on the disk after an event satisfies the HLT criteria.

1.2 Precision Proton Spectrometer

A subsystem of the Compact Muon Solenoid (CMS) experiment, is the Precision Proton Spectrometer (PPS)[33] that was created to measure scattered protons. PPS began as a collaborative effort between the TOTal Elastic and Diffractive

Cross Section Measurement (TOTEM)[34] and CMS[25] Collaborations. In addition to constructing new horizontal RP units to accommodate 3D pixel sensors and cylindrical Roma Pot units with diamond timing detectors, this new collaboration proposed a physics program that extended CMS acceptances for tagging scattered protons and measuring central exclusive processes. The sensitivity for tagging protons at high pile-up runs was enhanced by these improvements. Therefore, the second is employed for vertex identification and background suppression, whereas the first is more resilient to radiation.

PPS is then made up of a set of mechanical near-beam movable units, housing tracking[35] or timing[36] detectors that are sealed under a secondary vacuum with respect to the LHC primary vacuum. These kinds of detectors are referred to as Roman Pots (RPs) in the literature. As illustrated in Figure 1.7, PPS RP detectors are symmetrically positioned in stations along the LHC beam pipe at approximately $200\mathrm{m}$ from the CMS Interaction Point (IP5).

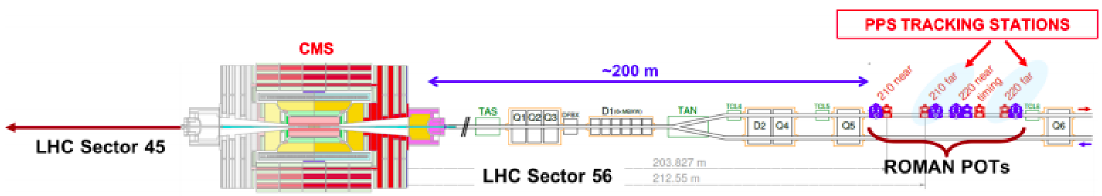

Figure 1.7: Schematic diagram of one of the LHC sectors where PPS is installed. The timing station is placed between two tracking stations. The figure is taken from Ref.[36].

In the figure, there is also a distinction depending on their position. The terms "near" and "far" identify the closest and furthest from the interaction point. Stations 0 near sit 203m away from IP5, stations 0 far sit at 214m, stations 2 near sit at 215m and stations 2 far at 220.

PPS comprises Roman Pots (RPs), a set of mechanical near-beam movable units sealed under a secondary vacuum to avoid contamination in the primary vacuum. Two units, near and far, make up a Roman Pot station. A Roman Pot unit is formed by three RPs: top vertical, bottom vertical, and horizontal. A photograph of the unit with its elements signaled is shown in Fig. 1.8

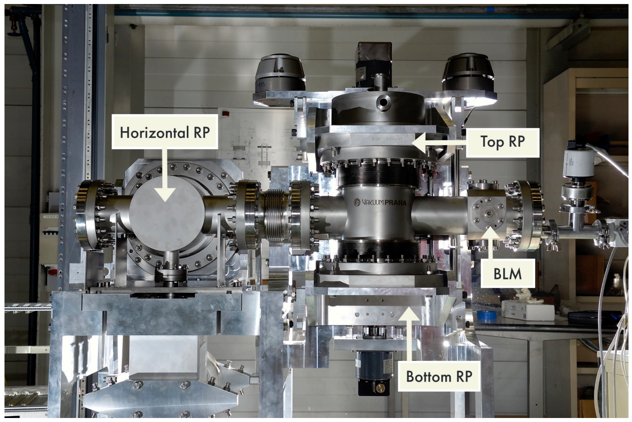

Figure 1.8: Photograph of the Roman Pot unit [37] before installation, the three RPs are signaled as well as the Beam Loss Monitor (BLM).

1.2.1 Tracking system

The PPS tracking system measures the interested protons' position and direction. TOTEM's Si-Strips detectors could not endure the radiation they were exposed to and could not identify multiple tracks, which is particularly important when there are high pile-up conditions. Nonetheless, as previously mentioned, this technology is still used in vertical RPs that only operate during the alignment or special low luminosity runs.

▷ 3D Pixels

3D pixel sensors have gradually replaced the strips on the horizontal detectors. The system modifications from the first installation to the most recent upgrade are depicted in Fig. 1.9, and the detectors are shown more in detail in Figure 1.10.

The tracking reconstruction is performed using the hits identified in at least three of the six planes that comprise the detector package. For Run 2, the 3D pixel sensors were built into a $230\mu m$ thick wafer, columns with $200\mu m$ depth and a diameter of $10\mu m$ [40]. They had a pixel size of 100 x 150 μm^2, and there were two configurations, one and two readout columns.

The hits found in at least three of the six planes that make up the detector package are used to reconstruct the tracking. For Run 2, the 3D pixel sensors were integrated into columns that had a diameter of $10\mu m$, a depth of $200\mu m$, and a wafer that was $230\mu m$ thick [40]. There were two configurations, one and two readout column formats, and each pixel measured 100 x 150 μm^2.

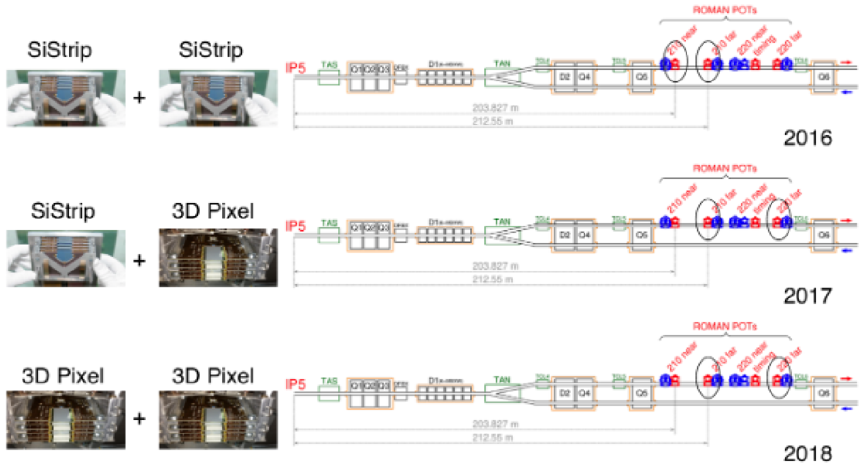

Figure 1.9: Scheme showing the upgrade to the horizontal tracking detectors from 2016 to 2018 and the different technologies used (3D Pixels and Silicon Strips).[38].

Figure 1.10: Silicon 3D Pixel Roman Pot. A picture taken during the tests on these devices is on the right, while a close-up photo is on the left. [39].

Because of their radiation hardness and the possibility of reducing the insensitive edge area by using slender edges, 3D pixel sensors were selected as the primary tracking technology [41]. The thin margins of the current detector are $200\mu m$, however, they could be lowered to $\sim 50\mu m$. The 3D Pixel detectors are described in further depth in references. [42, 43].

The Precision Proton Spectrometer (PPS) together with the CMS detector, obtained high-quality data during Run 2 in 2018 $\sim 55fb^{-1}$. In 2018, pixel detectors were installed in the two PPS arm tracking stations, offering proton kinematics measurement and track reconstruction of previously unheard-of quality.

The intricate alignment process entails lining up the RPs about the beam pipe and the sensor planes inside a single Roman pot. It is feasible to place the RPs near the beam when stable circumstances are present inside the LHC (during low luminosity runs). In figure 1.11, it can be seen that the proton, portrayed via the red line, impacts the devices previously discussed. The green rectangles indicate horizontal RPs (equipped with pixels), while the purple figures represent vertical RPs (equipped with strips).

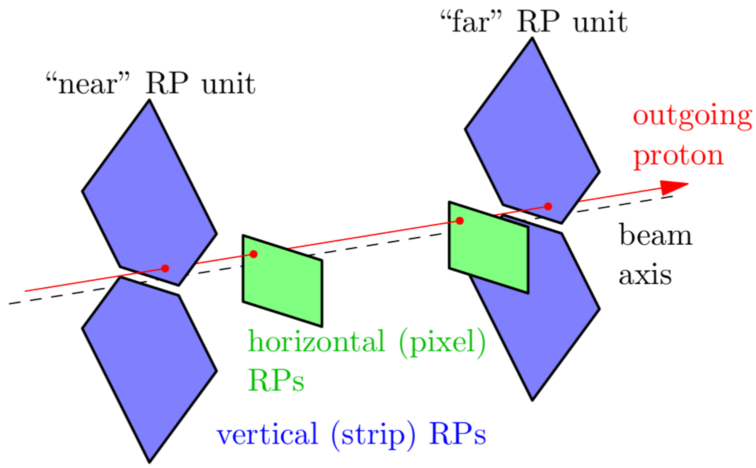

Figure 1.11: Scheme showing a proton going through the horizontal (green) RPs that are the ones used for tracking and are made up of pixels and vertical (blue) RPs that are made of strips and are only used for alignment purposes.[44].

PPS has used two reconstruction strategies: singleRP and multi-RP. However, the analysis described is focused on using multi-RP only as it provides the best resolution in transverse mass reconstruction. This configuration refers to cases where only events in which information from both stations of the same arm have been recorded.

1.2.2 Kinematics and Proton reconstruction

Given the required momentum balance between the central CMS system and PPS-detected protons, the kinematics of the interaction are closed, making it a great environment to study new physics.

The fractional momentum loss is given by

$$\xi = \frac{p_{nom} - p}{p_{nom}} \tag{1.9}$$

Where p_{nom} and p are the nominal momentum of the beam and the scattered proton momentum, respectively. The mass produced in the central region (in our case, the scalar (m_s)), given that we have small angles inside PPS, can be reconstructed in terms of the fractional momentum as:

$$m_s = \sqrt{s\xi_1\xi_2} \tag{1.10}$$

where ξ_1 and ξ_2 are the fractional momentum loss of the first and the second proton, respectively, and $\sqrt{s}$ is the center of mass energy.

On the other hand, its rapidity can also be written regarding the fractional momentum loss.

$$y = \frac{1}{2} ln \frac{\xi_1}{\xi_2} \tag{1.11}$$

Moreover, the mass of the scalar can also be calculated due to momentum conservation.

$$m^2_{miss} = \left[\left(p^{in}_{p1} + p^{in}_{p2} \right) - \left(p_H + p^{out}_{p1} + p^{out}_{p2} \right) \right]^2 \tag{1.12}$$

PPS utilizes the LHC accelerator magnets between the interaction point (IP) and the RPs. In the RPs, dispersed protons are detected after transiting through a region of the LHC lattice with 29 main and corrector magnets. Given that the protons that get to the PPS detectors travel more than 200 meters inside the LHC's vacuum pipe and are rather near the LHC beams, a standard method for model beams in an accelerator is used.[44]

The proton kinematics at a distance l from the interaction point (**d**) and the kinematics at the actual interaction point (**d***), are connected via the transport equation:

$$\mathbf{d}(l) = T(l, \xi) \cdot \mathbf{d}^*, \tag{1.13}$$

where the superscript $*$ is utilized to denote the given value at z= 0. The kinematics of the proton are explained by $\mathbf{d}(l) = (z, \theta_x, y, \theta_y, \xi)^T$, with (x,y) and (θ_x, y, θ_y) representing the transverse position and angles respectively. Meanwhile, ξ denotes the fractional momentum loss ξ that was previously described.

The transport matrix is:

$$T(s,\xi) = \begin{pmatrix} \nu_x & L_x & m_{13} & m_{14} & D_x \\ \frac{d\nu_x}{dl} & \frac{dL_x}{dl} & m_{23} & m_{24} & \frac{dD_x}{dl} \\ m_{31} & m_{32} & \nu_y & L_y & D_y \\ m_{41} & m_{42} & \frac{d\nu_y}{dl} & \frac{dL_y}{dl} & \frac{dD_y}{dl} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.14}$$

where D_x is the horizontal dispersion and the other matrix elements are the optical functions [45].

1.2.3 Proton acceptance

Pile-up events are a background source because they can effectively produce an event with two protons arriving at the PPS spectrometer from various interactions. Events produced by two distinct low-mass Single-Diffractive collisions (SD) dominate this pile-up background.

All the events having a single proton per arm in the PPS detectors have a proton mass spectrum. In this regard, the following describes the procedure to disentangle the shape of the random pileup background and the signal. As expected, if both protons from the signal are within acceptance, the spectrum has a typical Gaussian shape with the PPS spectrometer resolution width. But when one of the outgoing protons is not inside the PPS acceptance and is lost, the signal can generate a proton mass spectrum distributed over the full acceptance range. This case is also analyzed. For such an occurrence, there is also a nonnegligible probability that a random proton from the SD pileup traverses the RP pixels and generates a different mass spectrum that is not concentrated around the scalar particle mass peak. These are called ”Out of Acceptance events” and are described separately. This last class of events alters the background with a specific shape and will be treated as background.

There are two scenarios. The first one is illustrated in Figure 1.12, where one of the protons coming from the central event arrives below the RPs and is replaced by a pileup proton with a larger momentum. In comparison to the peak linked to events in the acceptance, the peak will occur at bigger masses. The second scenario is illustrated in Figure 1.13, where one of the protons coming from the central event arrives above the RPs and is replaced by a pileup proton with a smaller momentum. The peak in this case will occur at lower masses with respect to the one linked to the acceptance.

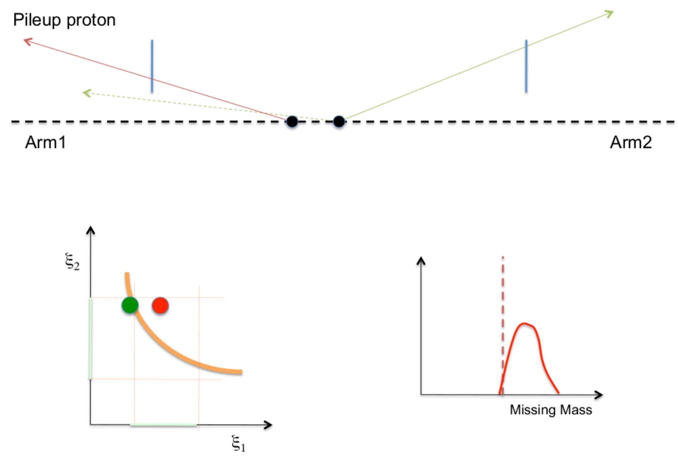

Figure 1.12: First out of acceptance scenario. The scheme shows the case in which one proton from the central event is registered in the RP, and one is missed but replaced by a pileup one. On the bottom, a graph of its acceptance and missing mass spectrum is also shown.[39]

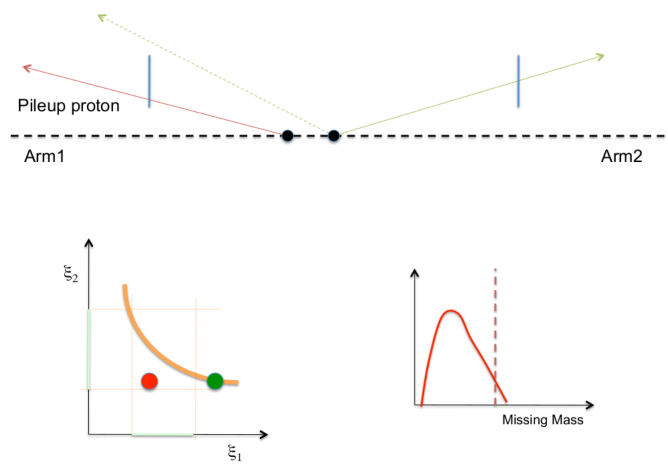

Figure 1.13: Second out of acceptance scenario. A proton from the central event hits the Roman Pot, while the other proton hits the collimator above the RP. This second one is replaced by a pileup proton with a smaller momentum and creates a blob in the proton mass spectrum at lower values.[39]

Figure 1.14 shows the ideal scenario where both protons are within the acceptance, and no pileup ones are present. When both protons from the central event are within the acceptance limits of the PPS there is a peak associated with the signal in the proton mass spectrum at $\sqrt{s\xi_1\xi_2}$.

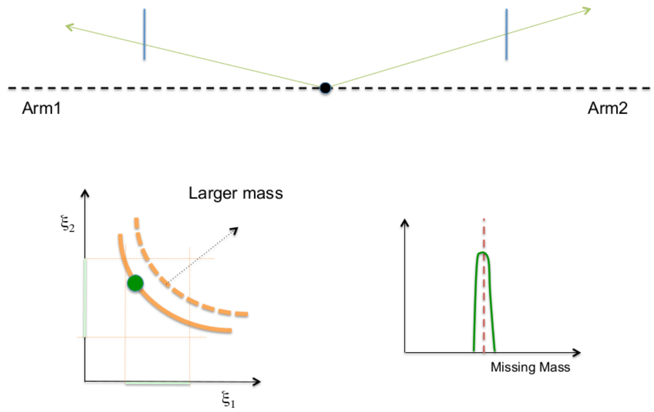

Figure 1.14: Scheme showing the case in which both protons from the central event hit their respective RPs, so they are within the acceptance, and there are no pileup protons.[39]

Moreover, it was determined that events with only one reconstructed proton per arm in the Roman Pots were to be chosen. Since the contribution from events with several proton tracks per arm would introduce biases that are hard to correct, the decision was made due to the difficulties in evaluating the background shapes. The loss in acceptance discarding events with more than one proton track per arm is sufficiently small to simplify the study. While random protons from the data are eventually incorporated into the events in the Monte-Carlo (MC) process, data without protons detected is used for background studies.

Additionally, by determining the frequency of a proton configuration compared to the others, the fractions of events with varying random proton content in the CMS Roman Pots were assessed for each arm. For this thesis, the critical highlight in this regard is the case with zero-zero pileup protons present. This corresponds to a factor of 27.6%, which is applied in the studies for background validation using Monte Carlo.

2
Analysis strategy

As previously explained, an SM Higgs as the CMS-triggered central object is considered, and then it is determined if a few protons identified by the Roman Pots are associated with it. Up to two leading protons can be formed by each narrow scalar particle created by photon fusion, and the PPS has excellent resolution for identifying any mass resonance. If both protons survive the collision, the forward PPS tracker will measure the scalar particle mass with a resolution of about 20 to 30 GeV. This should reveal a narrow structure on a background that is triangular.

PPS is highly sensitive to photon fusion-produced high-mass scalars. The mass window begins with limited acceptance at 600 GeV, reaches a value above 800 GeV, and extends to masses of 2 TeV. Additionally, photon fusion is the only source of CEP left when the Sudakov suppression factors decrease the formation of central exclusive scalars by gluon fusion.

While the scalar particle should appear as a mass bump in the spectrum calculated with the protons as measured by PPS, the Higgs can be detected efficiently by two isolated b-tagged jets. With these assumptions, the produced scalar decays into a Higgs and some invisible particle, scalar or pseudoscalar. The different decay modes that will complicate the kinematics of the event cannot be excluded.

However, the Higgs recoils mainly in the transverse plane with the invisible particle(s) since the CEP events have a very limited transverse momentum spread ($< 10\,GeV$). The event will be associated with missing transverse momentum, and the central kinematics will be driven exclusively by the mass difference between the scalar and the invisible part of the event.

The background is mainly dominated by events generated by two separated low mass Single Diffractive collisions from the pileup. Pile-up events can efficiently generate two protons from different interactions that reach the PPS spectrometer, as it was previously described.

2.0.1 Masses

The investigation discovered relationships between the masses and the event's kinematics. Specifically, the difference between the masses (which corresponds to the accessible phase space of the final state) seems more important than the masses themselves. Every generated dataset has been examined, and in every instance, the kinematic properties remain unchanged for varying scalar masses but for equal Δ. Instead, the primary distinctions are evident when the Δ fluctuates and the m_s is fixed.

Therefore, this analysis is focused on producing masses for the scalar from 800 to 1600 GeV. There are simulations for five $\Delta M = m_S - m_X - m_H$ of 40, 60, 80, 100, and 120 GeV and ten distinct S masses (MS) of 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, and 1600 GeV. Certain distributions of the Higgs kinematics, such as the S and X masses, change and are repeated for every combination of these parameters. More details on this part of the analysis can be found in [39].

It was decided to analyze the 2018 sample because the proton reconstruction is well-defined in the POG, and the Roman Pots have Pixels in both stations of each arm. As well as only taking into account the Multi-RP configuration since it has the best resolution and the POG rules are well understood.

The main idea of this search is that the two photons create a resonance with a width below the PPS resolution, which can be seen as a peak in the proton mass spectrum with the described PPS resolution. The Scalar decays into a Higgs and one invisible (X) particle. The strategy is to select a central object, namely a Higgs decaying into two b-Jets. Then match all the kinematics up to some level to later proceed with a bump search on top of the background from pileup (as in the Z+X[20] analysis)

Finally, a statistical analysis of the Proton Mass spectrum is done.

3

Event selection

The simulation, calibration alignment, and reconstruction modules that process event data for physicists to perform analysis are designed around an event data model (EDM) framework and services. Together, these components form the overall collection of software known as CMSSW. The primary goal of the Framework and EDM is to facilitate the creation and use of reconstruction and analysis tools. In particular, $CMSSW$_10_6_30 is used for this analysis.

The MiniAOD is one of the data tiers used by the CMS collaboration to manage the vast amounts of data generated from the detector. It stands for ”minimized Analysis Object Data. This format is designed to contain only essential information needed for a broad range of physics analyses, significantly reducing the size of the dataset compared to the more detailed AOD (Analysis Object Data) format.

A specialized software framework[46] called "Skimmer" has been built to handle the samples from various MC models and CMS data streams and generate the final analysis files. It acts on the official datasets by producing a root NTuple with pre-selected CMSSW collections (and physical-computer objects) on an event-by-event basis. This framework takes advantage of the collections of physics-objects candidates compiled in Table 3.1.

Table 3.1: Collections of the physics objects-candidates used.

Candidate	CMSSW Collection
Hadronic level objects (MC)	*prunedGenParticles*
$\not{E}_T(MC)$	*genMetTrue*
AK4-jets (MC)	*slimmedGenJets*
AK4-jets	*slimmedJets*
particle flow	*packedPFCandidates*
$\not{E}_T$	*slimmedMETs*
primary-vertex (PV)	*offlineSlimmedPrimaryVertices*
multi Roman Pot protons	*ctppsProtons, multiRP*

Since color-charged particles cannot exist freely in nature, quarks and gluons created within CMS hadronize and generate a jet, a collection of charged and neutral particles. The detector's jets are cone-shaped structures that hold the tracks of many particles produced during hadronization. For its reconstruction, particle flow objects are clustered by a jet algorithm such as the anti-kT which is used in this case. For CMS analysis, jets with a cone size R = 0.4, referred to as "AK4" jets, are used as standard jets.

This research reconstructs the jets by rejecting charged hadrons linked to a pileup primary vertex and applying the AK4 clustering algorithm to the particle flow candidate particles. Additionally, jets have a minimum p_T of $10\ GeV$ and corrections L1, L2, and L3. The leading and second leading jets' b-tagging discriminators were calculated as *pfDeepCSVJetTags:probb* + *pfDeepCSVJetTags:probbb*. When at least one of the triggers listed in the menu is passed, the leading and second-leading jets are chosen.

On the other hand, missing energy is a highly helpful quantity for processes that are not fully understood. Missing energy is energy not picked up by the detectors but anticipated to exist due to momentum and energy conservation. Therefore, it is possible to utilize this amount to infer the presence of particles that would otherwise be invisible.

The missing transverse energy ($\not{E}_T$) used for this analysis is the *type-1*, computed subtracting the momentum vectors of the corrected jets from the original jets with anti-k_T and charged hadron suppression algorithm with a minimum p_T of 15 GeV.

The Skimmer's output is then used as an input for the Analyzer, which produces a TTree after some basic analysis cuts and the final NTuples used for the study.

3.1 Reducible Background

A Higgs decaying into two b-quark jets that recoil towards an unseen item linked with protons tagged by the CMS Roman Pots is present in this research's topological category of interest. Nonetheless, the CMS Roman Pots independent arms estimate a non-negligible probability of two b-tagged jets related to some Missing Transverse Momentum and associated with two pileup protons (or one diffractive and one pileup proton). These topologies can, therefore, imitate the relevant search signal. The most likely origin of background comes from proton-related tt, WW, ZZ, and QCD events.

The more important objective is to maximize the signal selection and reduce the background. To achieve this goal, a series of cuts were chosen. These cuts are the same for all data sets, regardless of the available scalar mass or phase space; therefore they are referred to as "static" cuts.

- $m_X > 0$: A negative mass is not physically allowed; this value results from an incorrect reconstruction, so these cases are excluded. The cut in the software framework terms is missingMassDijetMulti>0.
- $|dijet| < 2.4$: This value is the limit within which the particle can be detected inside CMS. See Figure 1.6.
- 85 GeV < dijetMass < 150 GeV: This is a window for the Higgs mass. From the simulations, this distribution lies almost entirely within the 85 to 150 GeV range (in all data sets), as seen in figure 3.1
- $\pi - |\pi - |Dijet\phi - Met\phi|| > 2.3$: Describes the angle between the X particle and the Higgs, the range is between 0 and π, the interaction results in being back to back.
- $leadingJetBtag > 0.8$ and $secondJetBtag > 0.8$: Efficiencies are calculated with a cutoff of the b-tagging estimator greater than 0.8. With this cutoff, approximately 25% of the signal remains, while the remaining background is an order of magnitude lower, around 3%. The b-tagging estimator greater than 0.8 selects the signal, while the b-tagging estimator less than 0.2 selects the background.

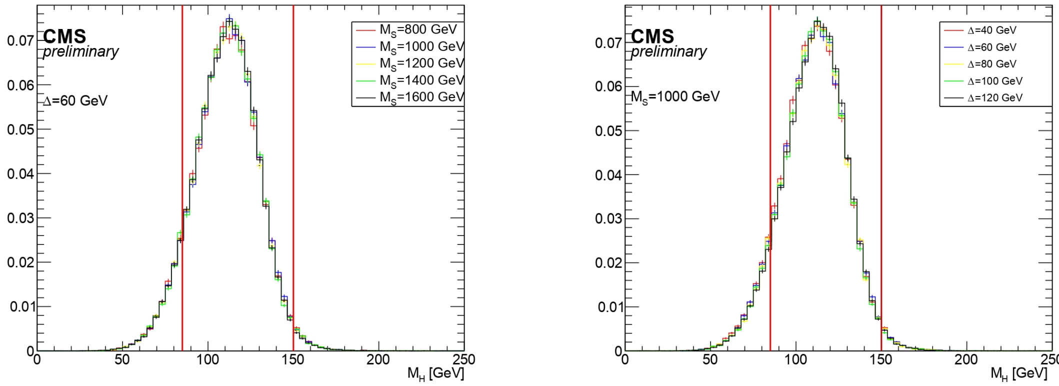

Figure 3.1: M_{jj} as M_S varies for $\Delta_M = 60GeV$ on the left plot, M_{jj} as Δ_M varies for $M_S = 1000GeV$ on the right one, the vertical lines show the cut.

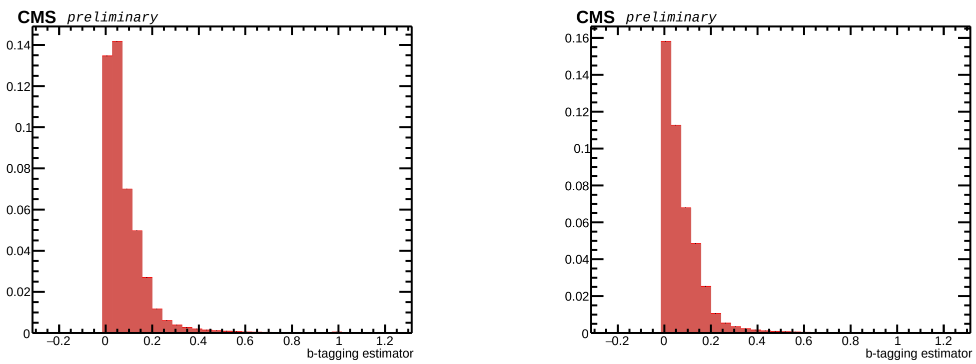

Figure 3.2: "Zerobias" sample B-Tag estimator for leading (a) and subleading jets (b). [39]

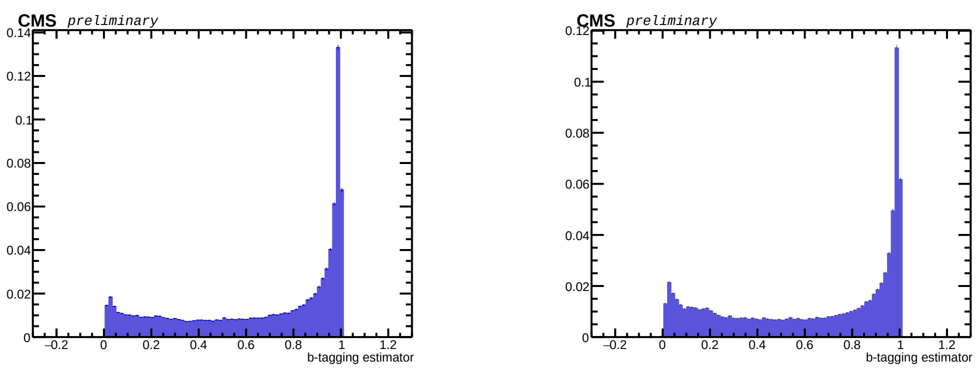

Figure 3.3: Signal simulation B-Tag estimator for leading (a) and subleading jets (b).[39]

One such method for choosing the signal over the noise is to apply a b-tagging estimator. The tagging of the quark bottom is a typical technique for the recognition of b-jet because mesons B have a long lifetime and high mass. This means that the decay of this quark is inside the detector, and it can be observed, so it's possible to detect the primary and the second vertex separately.

In addition, muons are present in the jets due to the semileptonic disintegration of B mesons[47]. The CMS muon chambers at the detector's extremities can detect muons. Muons are leptons with an extremely long and well-defined course that enables them to pass through the detector's end and be detected by particular external muon chambers. However, because there is dependence solely on the data from the calorimeter, it is more challenging to identify the electrons in a jet.

The figures in 3.3 and 3.2 show the background-associated b-tagging estimators in red, with the leading jet on the left and the second jet (distribution produced by ZeroBias) on the right. The same estimator for the signal is displayed in blue, with the leading jet on the right and the second on the left. This estimator can then be used to choose between the signal and the background given that the two distributions behave oppositely.

3.2 Trigger selection

The triggers studied and their efficiencies are shown in Table 3.2.

Table 3.2: Overall efficiency after the prescaling of the triggers for the signal MC sample generated with m_S=1000 GeV and $\Delta_M = 60$GeV.

Trigger Name	$\mathcal{P}\times$ Abs. Trig. Eff. [%]
HLT_BtagMu_AK4Dijet70_Mu5	0.01
HLT_BtagMu_AK4Dijet70_Mu5_noalgo	0.02
HLT_BtagMu_AK4Dijet40_Mu5	0.02
HLT_BtagMu_AK4Dijet40_Mu5_noalgo	0.04
HLT_BtagMu_AK4Dijet20_Mu5	0.01
HLT_BtagMu_AK4Dijet20_Mu5_noalgo	0.0001
HLT_Mu17_Mu8_Mass3p8	0.14
HLT_Mu17_Mu8_Mass8	0.14
HLT_IsoMu24_eta2p1	0.5
HLT_IsoMu24	0.5
HLT_IsoMu27	0.4

In the end, the triggers used for the study described here are only the HLT_Mu17_Mu8_Mass3p8 and HLT_Mu17_Mu8_Mass8. These were chosen based on the efficiency studies, where it was found that as the B Triggers have prescales applied, they are not as efficient as the muon triggers. From table 3.2 it also can be seen that the IsoMu triggers have a good efficiency but they are not in the same data stream as the other muon triggers so they are not used at the moment.

3.3 Monte-Carlo

Monte Carlo methods are employed for validation and to show good comprehension of the effectiveness of the primary CMS rebuilt objects and triggers, even though the background templates in this study are entirely modeled from data. A Higgs decaying into two b-quark jets and recoiling towards an invisible entity linked to protons tagged by the CMS Roman Pots is present in this investigation's topological category of interest. Nonetheless, the CMS Roman Pots independent arms assess a non-negligible probability of having two b-tagged jets connected to Missing Transverse Momentum ($\not{E}_T$) and associated with two pileup protons (or with a single diffractive and a pileup proton). Such topologies can imitate the interested search signal.

The standard model (SM) is used to compute particle decay widths and cross-sections of particle collisions. However, computing these values requires solving multidimensional integrals in the relativistic phase space, which can be challenging. Monte Carlo methods are frequently utilized to solve these integrals.

Monte Carlo is quite helpful in theoretical models since it allows one to explore the parameter space, test various theories and scenarios, change parameters, and see the simulated results. This helps validate or disprove suggested models. MadGraph5 aMC@NLO [48] (Matrix Elements for Multi-Process Generation at Next-to-Leading Order) is an event generator for particle physics that generates Feynman diagrams. It is used to calculate matrix amplitudes, which are mathematical expressions describing the probability of occurrence of a given particle scattering process. Matrix amplitudes are calculated at the tree and next-order (NLO) level. After calculating the matrix amplitudes, it generates simulated events that describe the outcome of the particle collisions in the experiment. This is done using Monte Carlo methods to sample the phase space and generate events statistically efficiently. MadGraph5 amc@nlo can be interfaced with PYTHIA8 for hadronization, a program used to create high-energy physics collision events or to describe collisions between heavy nuclei, protons, electrons, and photons at high energies. Hard and soft interactions, parton distributions, initial and final-state parton showers, multiparton interactions, fragmentation, and decay are just a few of the physics-related topics it covers through theory and models. Although it draws heavily from unique research, it also uses several formulas and information from other works of literature. As a result, It falls within the general-purpose Monte Carlo event generator.

Although the background templates in this analysis are fully modeled from data, specific Monte Carlo (MC) models are employed for validation purposes and to show our level of understanding of the efficiencies of the central (CMS) reconstructed objects and triggers.

In this work, the MiniAOD MC samples used are taken from the Data Aggregation System (DAS), as they are generated events processed through full CMS reconstruction and have already been validated.

The process of interest in the analysis can suffer from a large quantum chromodynamics (QCD) background, so it is crucial to understand this contribution. Table 4.1 contains all the QCD MC samples.

The MC samples are run over the skimmer individually, and later, the final Ntuple is generated with the Analyzer. A critical remark is that the distributions of the number of pileup interactions must match the conditions during each data-taking period; the final Ntuples consider this. The PileUp distribution is sensitive to certain aspects of the primary vertex reconstruction and other potential effects from differences between events in data and MC. Therefore, to compare the yields from the MC samples with the dataset, a pileup re-weighting was needed, and it was done through the Analyzer, which takes as input the data luminosity distribution that is estimated by running the pileupCalc tool (a utility provided by the lumi POG) with the corresponding pileup JSON file, and the MC pileup distributions for each sample.

The probability of zero pileup protons on the two arms (0-0) is 27.6%. If protons are on the RPs, zero pileup protons indicate that they originate from the signal.

The following plots (3.4, and 3.5) show the distribution for era A data sample and QCD Monte Carlo samples with the pileUp weight, Corresponding cross-section normalization factor, and percentage proton efficiency (27.6%) applied for E_T^{miss} which is the main variable used for validation purposes. However, as seen on the plots on the right side, QCD MC does not have many statistics when the trigger selection and the static cuts are applied.

The data sample used is: $DoubleMuonRun2018AUL2018\_MiniAODv2v1MINIAOD$.

Table 3.3: Official miniAOD MC datasets with 2018 data conditions and the respective cross-sections ($\sigma_{\rm MC}$).

Dataset	$\sigma_{\rm MC}$ [pb]
/QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	23700000
/QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	1547000
/QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	322600
/QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	29980
/QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	6334
/QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	1088
/QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	99.11
/QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	20.23

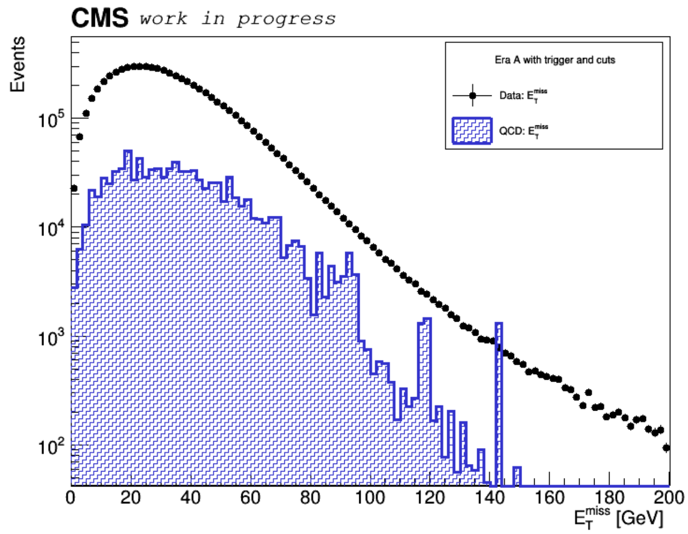

Figure 3.4: QCD and data era A E_T^{miss} distributions with only trigger selection applied. On the right: QCD and data era A E_T^{miss} distributions with trigger selection and static cuts applied. The pileup weight, Corresponding cross-section normalization factor, and percentage proton efficiency (27.6%) are applied to QCD MC.

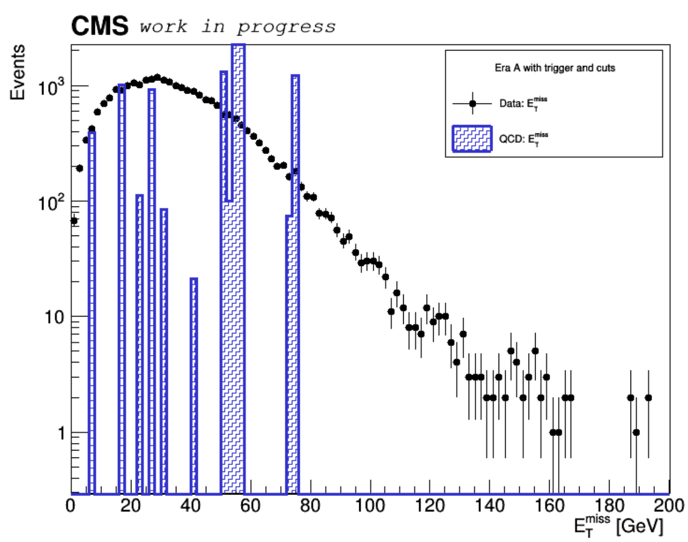

Figure 3.5: QCD and data era A E_T^{miss} distributions with trigger selection and static cuts applied. The pileup weight, corresponding cross-section normalization factor, and percentage proton efficiency (27.6%) are applied to QCD MC.

4

Data-driven estimation of QCD

An essential job for any study is estimating the event yields from various background processes and the shape of these backgrounds as a function of multiple variables. When a background estimation is said to be ”data-driven,” it implies that it is mainly based on actual collision occurrences that have been seen. Estimates based on data help verify the predictions made by simulations. Furthermore, there are situations where the simulations simply will not yield a trustworthy background estimate; in these circumstances, a data-driven background estimate is essential.

Given that the results obtained after applying the trigger selection and static cuts on the QCD MC have low statistics, it was decided to complement the study by following a data-driven method.

A variety of data-driven estimating techniques could lower the background estimation uncertainty. In data-driven approaches, background contributions in the ”signal” region (SR), where interesting events could be detected, are estimated using data from the ”background” dominated control region (CR).

A method commonly used in High-Energy physics is known as the ”ABCD” method. In this method, two variables (f and g) with uncorrelated background distributions are chosen. If the 2D-plane (f−g) is divided into four regions (A, B, C, and D), then the signal region, A, is where both the f and g selection requirements are applied. This method has been used in several analyses, and the results have been validated. So, it has been decided to follow this approach. An illustration of the regions is depicted in the figure 4.1.

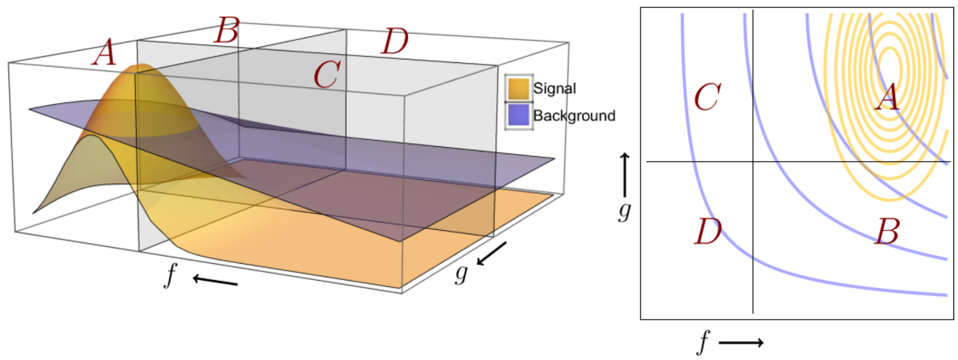

Figure 4.1: ABCD visual representation.[49]

The ABCD method then allows the data-driven estimation of a background rate when events are selected by a pair of cuts in a plane of two uncorrelated variables such that both cuts enhance the signal-to-background ratio. The main idea is that there are sufficient background statistics in control regions B, C, and D to estimate the small background rate in the signal region. Other assumptions done are that there is no signal leakage to regions B, C, and D, and that there is no correlation between the variables for the background in question.

Moreover, this approach is based on the idea that the following is accurate:

$$\frac{N_C}{N_D} = \frac{N_A}{N_B} \tag{4.1}$$

Three defined regions, B, C, and D, are background-dominated. The fourth, region A, is the signal region. If the observables are independent (no correlation), then the background in the signal region can be predicted from the other three areas through the following relation

$$N_A = \frac{N_C N_B}{N_D} \tag{4.2}$$

The ABCD method's observables, f and g, are typically selected to be straightforward, physically well-motivated characteristics. They are manually guaranteed to be independent, for example, by choosing features that are physically proven to have a low association or by trial and error.

Independence suggests that the scatter plot's points should be evenly distributed. Dependency may be indicated by any grouping or empty spots in specific plot areas.

The variables that were chosen for this study are the dijet mass (defined in the software framework as dijetMass) and the angle between the X particle and the Higgs ($\pi - |\pi - |Dijet\phi - Met\phi||$). The function get correlation factor in ROOT was used to verify the independence of the variables. The value obtained was 0.0219868, which is close to zero; therefore, they are considered independent, and these variables can be used for the method.

The regions are then defined as follows:

- Region A: ($\pi - |\pi - |Dijet\phi - Met\phi||$) > 2.3 and dijetMass > 85
- Region B: ($\pi - |\pi - |Dijet\phi - Met\phi||$) < 2.3 and dijetMass > 85
- Region C: ($\pi - |\pi - |Dijet\phi - Met\phi||$) > 2.3 and dijetMass < 85
- Region D: ($\pi - |\pi - |Dijet\phi - Met\phi||$) < 2.3 and dijetMass < 85

where the signal is concentrated in region A.

A counter was used to get the number of events in each of the defined regions and the results were as follows for B, C, and D:

- Region B: 41182
- Region C: 11946

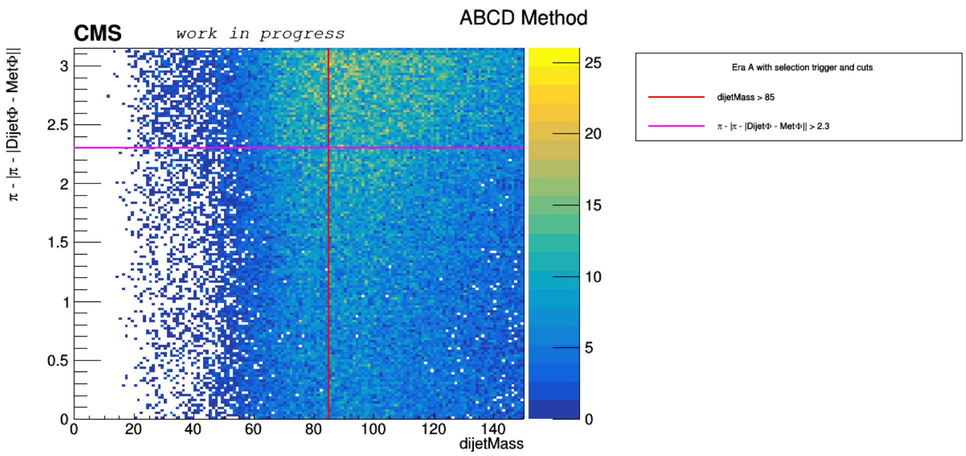

Figure 4.2: 2-D Map of $(\pi - |\pi - |Dijet\phi - Met\phi||)$ and dijetMass for data sample $DoubleMuonRun2018AUL2018\_MiniAODv2$-$v1MINIAOD$

- Region D: 20056

However, in this case, in the background, there are significant contributions of TTjets and Wjets in the data; the expected values of those are subtracted from the data (normalizing them to the corresponding cross-section of each sample and the luminosity for era A data, 12101.9613 pb^{-1}) and are to be left only with QCD in the background as a result. The samples used are described in 4.1.

Then, the expected QCD background events in each region is

- Region B: 40511.68963
- Region C: 11500.49896
- Region D: 19958.52072

Using these values in equation $N_A = \frac{N_C N_B}{N_D}$, the expected QCD background events in the signal region A is 23343.64611. Finally, the expected number of events is obtained and can be used to show the proper contribution of the QCD.

On the other hand, what is often referred to as the ”template method” is used to obtain the missing energy shape, the main variable of interest for validation purposes, and that lacked statistics in the Monte Carlo. To do so, 2 regions are defined (following the notation used for the ABCD method). CD region corresponds to the region with dijetMass

Table 4.1: miniAOD MC official datasets with 2018 data conditions and the respective cross-sections (σ_{MC}).

Dataset	σ_{MC} [pb]
/TTJets_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	833.9
/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM	1498.72
/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	1618.2
/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM	473.164
/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM	66.6768
/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM	14.9292
/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM	6.22456
/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	1.24584
/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM	0.00928116

< 85 cut applied and AB corresponds to the region with 150> dijetMass > 85. In figure 4.3, both distributions in the two separate regions are shown, here it is visible that the shape does not change, but to make it more evident, they are normalized to 1 and plotted in the same canvas in 4.4.

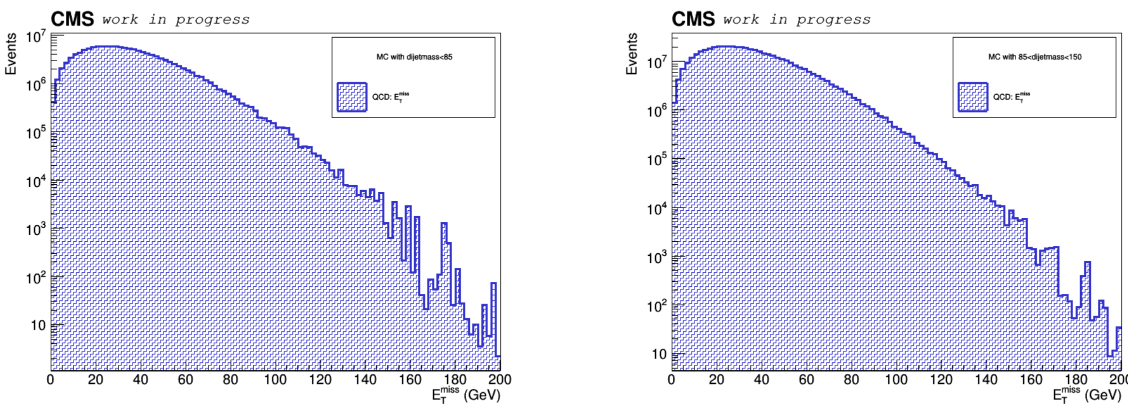

Figure 4.3: The plot shows the distribution of missing energy in the two defined regions, AB (right) and CD (left). AB refers to the region with 150> dijetMass > 85 and CD refers to the region with dijetMass < 85 .

Then, it has been shown that the shape of missing energy is conserved. However, if all the cuts and triggers were applied on MC then there would not be enough statistics as previously discussed. From the data then the shape is taken. The histogram used for this has all the trigger selection and the static cuts, except for dijetMass, for this variable dijetMass < 85 is applied. Given that the data for era A is used and the interest lies in only having the QCD contribution, here it also is necessary to subtract the TTjets and Wjets contribution (normalizing them to the corresponding cross-section of each sample and the luminosity for era A data). Finally, the shape is scaled to the number of expected QCD events obtained from the ABCD method.

Figure 4.5 shows then the result for the contribution of QCD compared to the data sample for era A with the trigger selection and all the static cuts applied. This corresponds to what is expected as a contribution for the QCD and it is clear that the estimation has improved significantly and statistics are no longer a problem.

Figures 4.7 and 4.8 show the significant improvement obtained with this data-driven method.

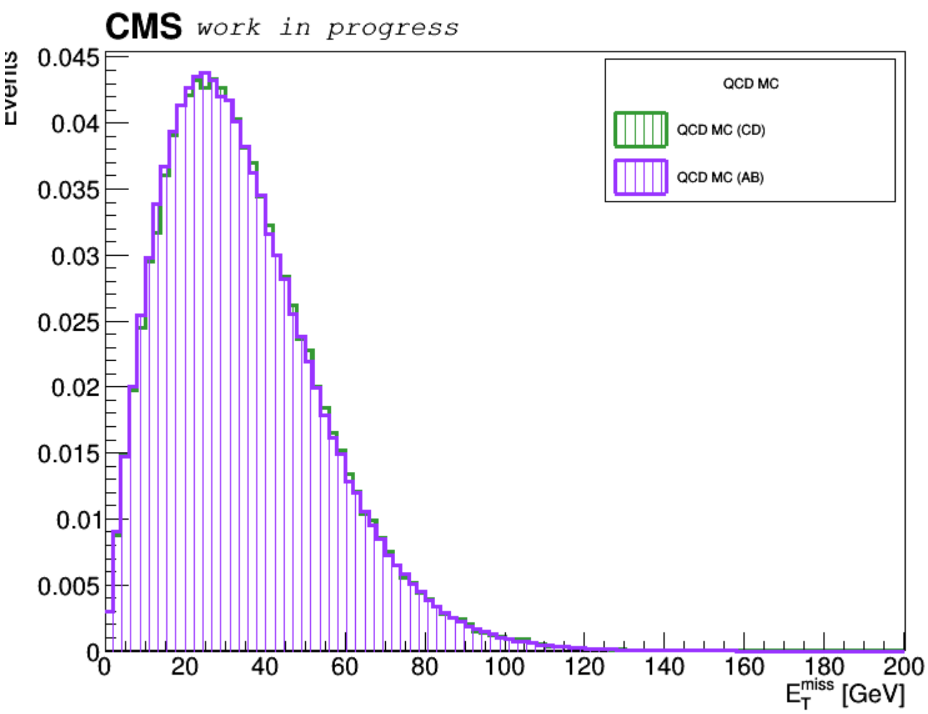

Figure 4.4: The plot shows the distribution of missing energy in the two quadrants of interest for the shape study normalized to 1. AB refers to the region with 150> dijetMass > 85 and CD refers to the region with dijetMass < 85 .

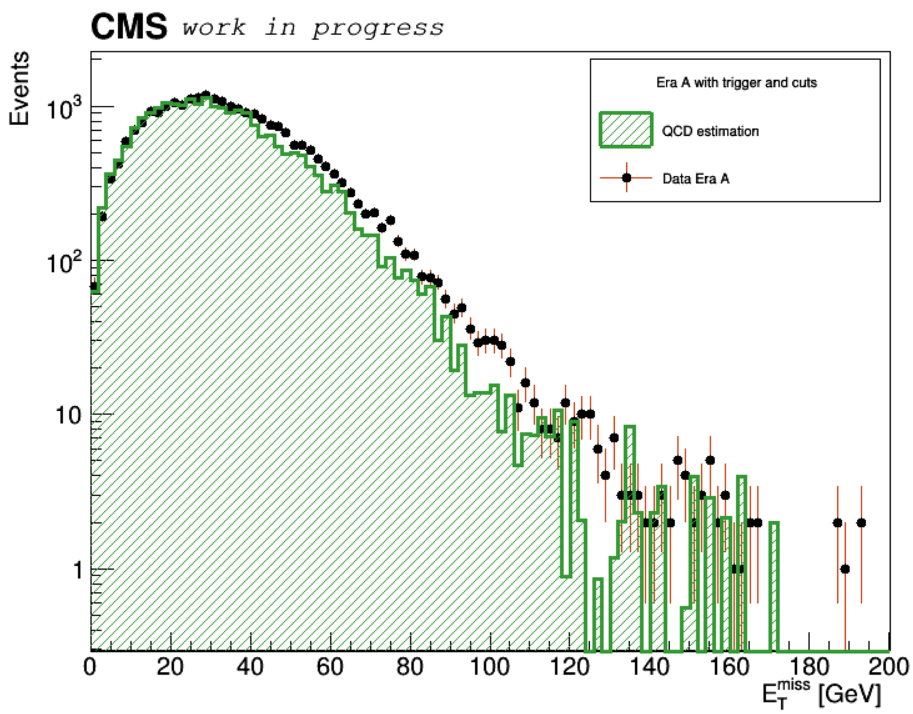

Figure 4.5: The plot shows, in black, the Data (DoubleMuon) for era A with all the trigger selection applied and static cuts. The QCD estimation, thanks to the ABCD and template methods, is shown in blue.

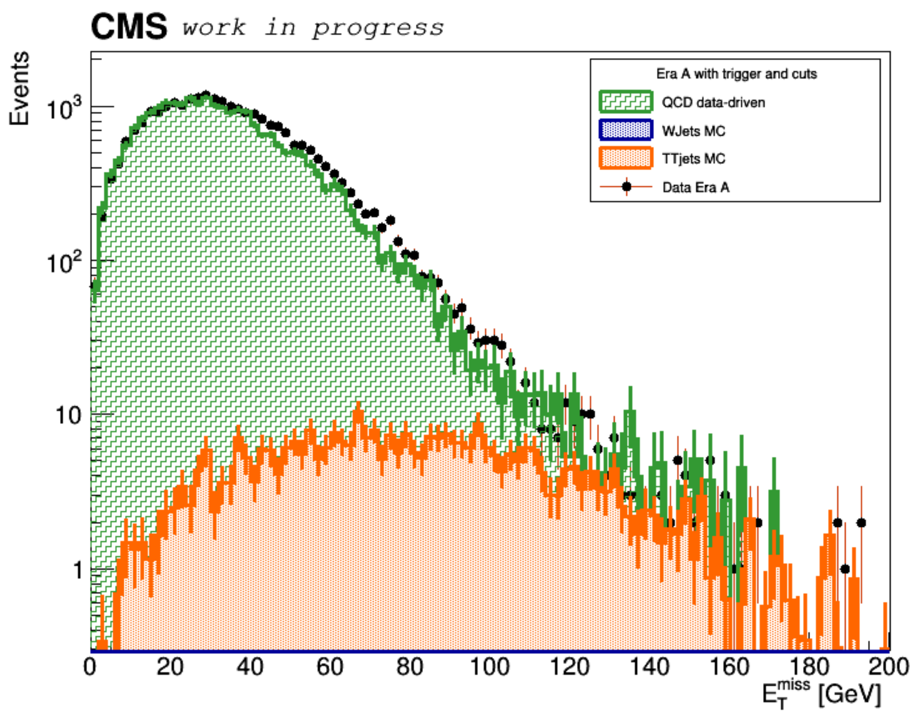

Figure 4.6: The plot shows the data for era A with the trigger selection and all static cuts applied. The background contributions are shown in a stack with their statistical errors.

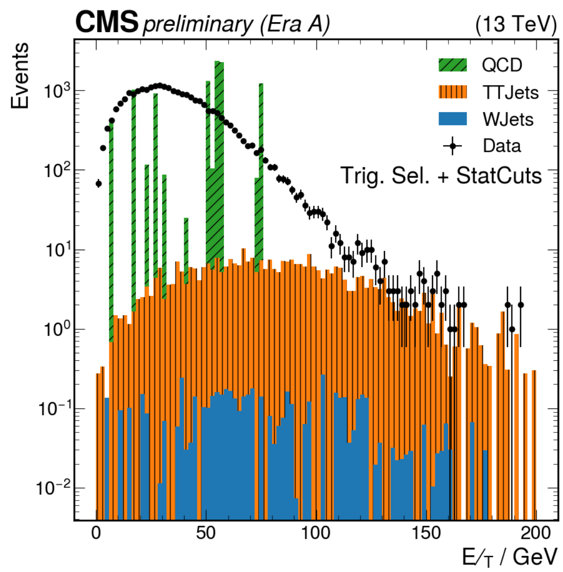

Figure 4.7: The plot shows the data (DoubleMuon) for era A with all the trigger selection applied and static cuts, and all the background contributions from Monte Carlo are shown in the stack.

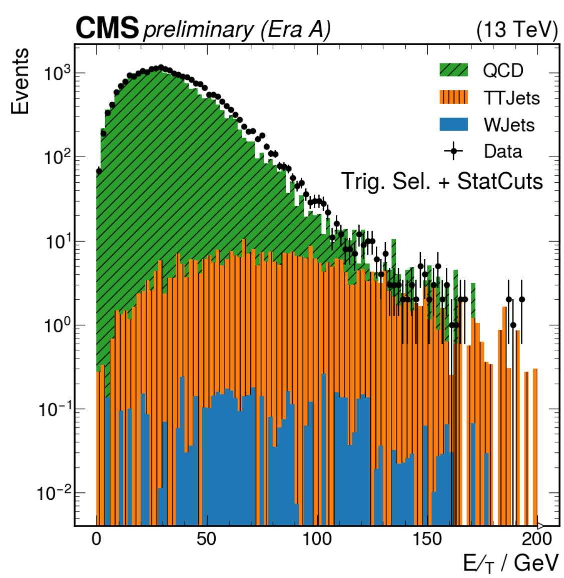

Figure 4.8: The plot shows the data (DoubleMuon) for era A with all the trigger selections applied and static cuts. The background contributions from MC for the TTjets and Wjets are shown, along with the new data-driven estimation for QCD, in the stack.

5

Conclusions

This thesis focused on studying the central production of an H boson and an unknown particle X. Two protons are present in the initial state, and at IP5, they interact electromagnetically by exchanging a photon. Photon-fusion results in the generation of a scalar resonance. Subsequently, the scalar particle decays into a Higgs boson (which also decays into two b quarks) and an unknown particle called X. The study is based on proton-proton (pp) collision data at 13 TeV gathered in 2018 with the CMS detector and the Precision Proton Spectrometer (PPS).

A general description of the ongoing analysis was presented. Moreover, specific Monte Carlo (MC) models are used to validate and demonstrate the level of comprehension of the efficiencies of the central (CMS) reconstructed objects and triggers, even if the background templates in this study are fully modeled from data.

The contribution of Quantum Chromodynamics (QCD) background events, which is crucial for isolating potential signals of new physics, was discussed. Given the selected so-called static cuts, QCD from Monte Carlo does not have enough statistics, and another method had to be implemented to solve this problem and have the correct QCD contribution. Through a data-driven method known as the ”ABCD Method,” the expected number of QCD background events in the signal region was obtained. On the other hand, through a template method, the shape for the main variable used for validation was obtained and scaled to the QCD expected background events. Results show the expected contribution for QCD with a smooth distribution, given that statistics is no longer a problem with this data-driven approach. This result also agrees with the rest of the contributions for the background, as seen from figure 4.8, where all the contributions are stacked and match the data. Given this result, a good background estimation for QCD via a data-driven method was obtained.

A
Outlook

As part of this master's, technical training at the National Institute of Advanced Industrial Science and Technology (AIST) in Tokyo, Japan, was received. The purpose of this technical training was to acquire complete training in data analysis and machine learning methods. The training focused on studying the main characteristics of machine learning, the components of learning algorithms, and the different tasks they can perform. Particular focus on the classification task was taken, given that in this case, the computer program can specify which of k categories some input belongs to, and therefore, it is very useful in various fields.

It was discussed how to use a deep learning method to improve the analysis further, but it was decided not to do so and in the end, was left as something to consider again in the near future. However, this appendix presents an overview of the different applications of these kinds of innovative algorithms.

With applications in classification and regression utilizing traditional methods like boosted decision trees, support vector machines, basic multi-layer perceptrons, etc., machine learning has long been important in particle physics.

Moreover, particle physics has started investigating deep learning as a special tool for solving challenging scientific issues beyond simple classification, organizing and interpreting large amounts of data, drawing conclusions about unknown causative factors, and even identifying the physical principles underlying intricate phenomena.

On the other hand, complex non-euclidean data is becoming increasingly prevalent in a wide range of applications. Therefore, deep neural networks for graphs have emerged as a potent tool for learning such data. However, even though the machine learning community has long acknowledged graph learning's potential, its application to downstream tasks has remained primarily unexploited.

In this context, self-attention is regarded as a graph-based architecture in the class of architectures known as Graph Neural Networks (GNN), which may be trained from data to learn functions on graphs. Graphs are a natural way to express data represented as unordered groups of items with rich relationships and interactions. Moreover, in high energy physics (HEP), GNNs are developing quickly. The different applications and models are reviewed in [50, 51].

In particle physics, jet tagging is a crucial yet difficult categorization assignment for the previously described analysis. Machine learning methods are usually used to solve this difficult task. Based on the extensive dataset JETCLASS, Particle Transformer (ParT), a novel Transformer-based jet tagging architecture, is proposed in [52]. ParT outperforms the previous state-of-the-art ParticleNet and performs better tagging than a basic Transformer by integrating paired particle interactions into the attention process.

A similar study to the one presented in this thesis is often done at Hadron colliders to look for dark matter; this is mono-X. Mono-X is a signal of a single particle created alongside missing (transverse) energy, X often signifies a (light) jet, $W\pm$, or Z or photon.

Some models are also applied to complement data-driven ones, as in the case of "ABCDisCo: Automating the ABCD Method with Machine Learning" [49]. Where instead of the independent classifiers intuitively being selected "by hand," it is investigated whether machine learning can be used to automate the creation of one or both of these classifiers. It also shows how to create strong yet independent discriminators using cutting-edge decorrelation techniques.

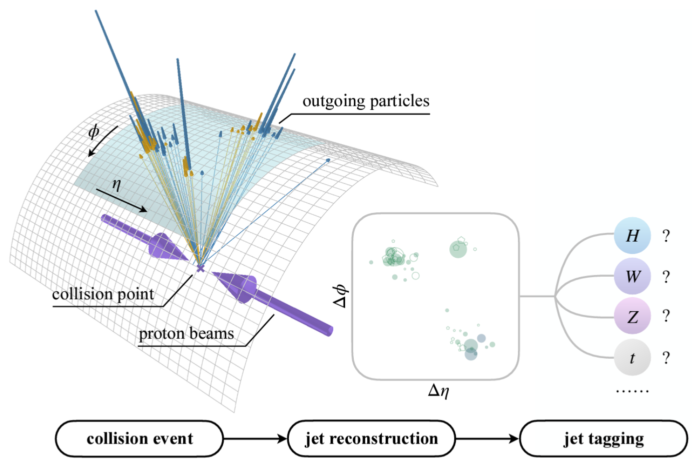

Figure A.1: The figure shows an example of jet tagging at the LHC at CERN. New unstable particles created by high-energy proton-proton collisions at the LHC have the potential to disintegrate and release a collimated spray of outgoing particles. Complex particle detector systems measure these outgoing particles, and the observed particles can produce jets. Jet tagging aims to categorize the jets and determine whether ones originate from high-interest particles, such as the top quark, the W or Z boson, or the Higgs boson. [52].

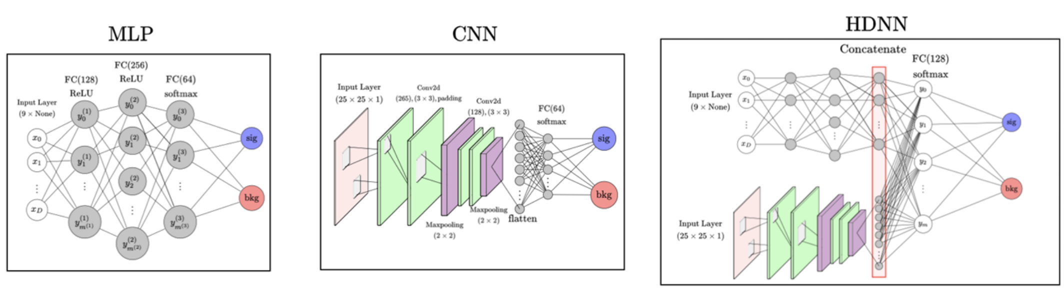

Figure A.2: An outline of the DNNs used to search for Mono-Higgs signals in $b\bar{b}$ final states using deep neural networks. MLP (left), CNN (center), and HDNN (right). HDNN's unlabeled layers resemble those of CNN and MLP. [53].

Nowadays, the list of innovative methods grows every day, and more tools are available to solve the challenges that come up with complex scientific analysis. Only some were described in this appendix and were not used this time for the thesis analysis, but they are an interesting approach to keep in mind.

Bibliography

[1] E. M. Le Boulicaut. “The Standard Model - ATLAS Physics Cheat Sheet”. General Photo. 2023. url: https://cds.cern.ch/record/2851985.

[2] Y. Iiyama. “The Standard Model and Its Supersymmetric Extension”. In: *Search for Supersymmetry in pp Collisions at* $\sqrt{s} = 8$ *TeV with a Photon, Lepton, and Missing Transverse Energy*. Cham: Springer International Publishing, 2017, pp. 5–45. isbn: 978-3-319-58661-8. doi: 10.1007/978-3-319-58661-8_2. url: https://doi.org/10.1007/978-3-319-58661-8_2.

[3] K. M. Belotsky et al. “”Signatures of primordial black hole dark matter””. In: *Modern Physics Letters A* 29.37 (2014), p. 1440005. doi: 10.1142/S0217732314400057. eprint: https://doi.org/10.1142/S0217732314400057. url: https://doi.org/10.1142/S0217732314400057.

[4] M. T. Frandsen et al. “LHC and Tevatron bounds on the dark matter direct detection cross-section for vector mediators”. In: *Journal of High Energy Physics* 2012.7 (July 2012). issn: 1029-8479. doi: 10.1007/jhep07(2012)123. url: http://dx.doi.org/10.1007/JHEP07(2012)123.

[5] A. Alves, S. Profumo, and F. S. Queiroz. “The Dark Z’ Portal: Direct, Indirect and Collider Searches”. In: *JHEP* 063 (2014), p. 029. doi: 10.1007/JHEP04(2014)063. arXiv: 1312.5281 [hep-ph].

[6] T. Jacques et al. “Complementarity of DM Searches in a Consistent Simplified Model: the Case of Z’”. In: *JHEP* 71 (2016), p. 038. doi: 10.1007/JHEP10(2016)071. arXiv: 1605.06513 [hep-ph].

[7] N. F Bell, Y. Cai, and Leane R. K. “Dark Forces in the Sky: Signals from Z’ and the Dark Higgs”. In: *JCAP* 08 (2016), p. 017. doi: 10.1088/1475-7516/2016/08/001. arXiv: 1605.09382 [hep-ph].

[8] A. Tan et al. “Dark Matter Results from First 98.7 Days of Data from the PandaX-II Experiment”. In: *Phys. Rev. Lett.* 117 (12 2016), p. 121303. doi: 10.1103/PhysRevLett.117.121303. url: https://link.aps.org/doi/10.1103/PhysRevLett.117.121303.

[9] D. S. Akerib et al. “Results from a Search for Dark Matter in the Complete LUX Exposure”. In: *Phys. Rev. Lett.* 118 (2 2017), p. 021303. doi: 10.1103/PhysRevLett.118.021303. url: https://link.aps.org/doi/10.1103/PhysRevLett.118.021303.

[10] M. Ackermann et al. “Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data”. In: *Phys. Rev. Lett.* 115 (23 2015), p. 231301. doi: 10.1103/PhysRevLett.115.231301. url: https://link.aps.org/doi/10.1103/PhysRevLett.115.231301.

[11] M. G. et al Aartsen. “Improved limits on dark matter annihilation in the Sun with the 79-string IceCube detector and implications for supersymmetry”. In: *JCAP* 04 (2016), p. 024. doi: 10.1088/1475-7516/2016/04/022. arXiv: 1601.00653 [hep-ph].

[12] G. Aad and et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. issn: 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.020. url: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[13] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. issn: 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.021. url: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[14] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. doi: 10.1103/PhysRevLett.13.321. url: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[15] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. doi: 10.1103/PhysRevLett.13.508. url: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. doi: 10.1103/PhysRevLett.13.585. url: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[17] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), pp. 1–102. doi: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].

[18] A. Angelescu and G. Arcadi. "Dark Matter Phenomenology of SM and Enlarged Higgs Sectors Extended with Vector Like Leptons". In: *Eur. Phys. J. C* 77.7 (2017), p. 456. doi: 10.1140/epjc/s10052-017-5015-2. arXiv: 1611.06186 [hep-ph].

[19] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG. Geneva: CERN, 2017. doi: 10.23731/CYRM-2017-002. url: https://cds.cern.ch/record/2227475.

[20] "A search for new physics in central exclusive production using the missing mass technique with the CMS detector and the CMS-TOTEM precision proton spectrometer". In: (Mar. 2023). arXiv: 2303.04596 [hep-ex].

[21] The CMS Collaboration. *The CMS Precision Proton Spectrometer at the HL-LHC – Expression of Interest*. Tech. rep. 88 pages. Geneva: CERN, 2020. arXiv: 2103.02752. url: https://cds.cern.ch/record/2750358.

[22] E. Mobs. "The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019". In: (2019). General Photo. url: https://cds.cern.ch/record/2684277.

[23] Werner Herr and B Muratori. "Concept of luminosity". In: (2006). doi: 10.5170/CERN-2006-002.361. url: https://cds.cern.ch/record/941318.

[24] R. De Maria. "LHC Interaction region upgrade". Presented on 14 Apr 2008. Ecole Polytechnique, Lausanne, 2008. url: https://cds.cern.ch/record/1127611.

[25] The CMS Collaboration et al. "The CMS experiment at the CERN LHC". In: *JINST* 3 S08004 (2008). doi: 10.1088/1748-0221/3/08/S08004. url: https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004.

[26] T. Sakuma. "Cutaway diagrams of CMS detector". In: (2019). url: https://cds.cern.ch/record/2665537.

[27] V. Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. url: https://cds.cern.ch/record/368412.

[28] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. url: https://cds.cern.ch/record/349375.

[29] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. url: https://cds.cern.ch/record/357153.

[30] *The CMS magnet project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. doi: 10.17181/CERN.6ZU0.V4T9. url: https://cds.cern.ch/record/331056.

[31] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. url: https://cds.cern.ch/record/343814.

[32] Aram Hayrapetyan et al. "Performance of CMS muon reconstruction from proton-proton to heavy ion collisions". In: *JINST* 19.09 (2024). All the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/MUO-21-001 (CMS Public Pages), P09012. doi: 10.1088/1748-0221/19/09/P09012. arXiv: 2404.17377. url: https://cds.cern.ch/record/2896442.

[33] M. Albrow et al. *CMS-TOTEM Precision Proton Spectrometer*. Tech. rep. 2014. url: https://cds.cern.ch/record/1753795.

[34] V. Berardi et al. *Total cross-section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN: TOTEM Technical Design Report*. Technical design report. TOTEM. Geneva: CERN, 2004. url: http://cds.cern.ch/record/704349.

[35] M. M. Obertino. "The PPS tracking system: performance in LHC Run2 and prospects for LHC Run3". In: *JINST* 15 (2020), C05049. 11 p. doi: 10.1088/1748-0221/15/05/C05049. url: https://cds.cern.ch/record/2747949.

[36] E. Bossini. "The CMS Precision Proton Spectrometer timing system: performance in Run 2, future upgrades and sensor radiation hardness studies". In: *JINST* 15 (2020).

[37] M. Brice. "The Roman Pot for the TOTEM experiment". 2006. url: http://cds.cern.ch/record/1001240.

[38] *CMS Data Analysis School LPC 2022: Forward protons*. url: https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideCMSDataAnalysisSchoolLPC2022TaggedProtonsShortExercise.

[39] A. Feliziani. "BSM (Beyond Standard Model) searches with escaping invisible particles associated to a Higgs boson and protons tagged with the PPS (Precision Proton Spectrometer)". 2024. url: https://etd.adm.unipi.it/t/etd-02092024-140933/.

[40] M. M. Obertino. *The PPS tracking system: performance in LHC Run2 and prospects for LHC Run3. The CT-PPS tracking system performance in LHC-Run2 and prospects for LHC-Run3*. Tech. rep. Geneva: CERN, 2020. doi: 10.1088/1748-0221/15/05/C05049. url: http://cds.cern.ch/record/2797458.

[41] M. Povoli et al. "Slim edges in double-sided silicon 3D detectors". In: *Journal of Instrumentation* 7.01 (2012), pp. C01015–C01015. doi: 10.1088/1748-0221/7/01/c01015. url: https://doi.org/10.1088/1748-0221/7/01/c01015.

[42] F. Ravera. "The CT-PPS tracking system with 3D pixel detectors". In: *Journal of Instrumentation* 11.11 (2016), pp. C11027–C11027. doi: 10.1088/1748-0221/11/11/c11027. url: https://doi.org/10.1088/1748-0221/11/11/c11027.

[43] F. Ravera. "3D silicon pixel detectors for the CT-PPS tracking system". Presented 16 Jun 2017. 2017. url: https://cds.cern.ch/record/2682003.

[44] A. Tumasyan et al. "Proton reconstruction with the CMS-TOTEM Precision Proton Spectrometer". In: *Journal of Instrumentation* 18.09 (2023), P09009. doi: 10.1088/1748-0221/18/09/P09009. url: https://dx.doi.org/10.1088/1748-0221/18/09/P09009.

[45] F. Willeke and G. Ripken. "Methods of Beam Optics". In: *AIP Conf. Proc.* 184 (2008), pp. 758–819. doi: 10.1063/1.38050.

[46] A. Bellora et al. *PPS Framework Software*. https://github.com/InvisibleHunters/PPSFramework. 2024.

[47] C. Ferro. *B-tagging in CMS*. 2012. arXiv: 1201.5292 [hep-ex]. url: https://arxiv.org/abs/1201.5292.

[48] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (July 2014). issn: 1029-8479. doi: 10.1007/jhep07(2014)079. url: http://dx.doi.org/10.1007/JHEP07(2014)079.

[49] G. Kasieczka et al. “Automating the ABCD method with machine learning”. In: *Physical Review D* 103.3 (Feb. 2021). issn: 2470-0029. doi: 10.1103/physrevd.103.035021. url: http://dx.doi.org/10.1103/PhysRevD.103.035021.

[50] S. Thais et al. “Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges”. In: (2022). arXiv: 2203.12852 [hep-ex]. url: https://arxiv.org/abs/2203.12852.

[51] J. Shlomi, P. Battaglia, and J. Vlimant. “Graph neural networks in particle physics”. In: *Machine Learning: Science and Technology* 2.2 (Jan. 2021), p. 021001. issn: 2632-2153. doi: 10.1088/2632-2153/abbf9a. url: http://dx.doi.org/10.1088/2632-2153/abbf9a.

[52] Huilin Q., Congqiao L., and Sitian Q. “Particle Transformer for Jet Tagging”. In: (2024). arXiv: 2202.03772 [hep-ph]. url: https://arxiv.org/abs/2202.03772.

[53] A. Hammad, S. Khalil, and S. Moretti. “Search for Mono-Higgs Signals in $b\bar{b}$ Final States Using Deep Neural Networks”. In: *Physical Review D* 107.7 (Apr. 2023). issn: 2470-0029. doi: 10.1103/physrevd.107.075027. url: http://dx.doi.org/10.1103/PhysRevD.107.075027.