Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Search for Exotic Particles with the CMS detector at the CERN LHC

Tesis presentada al

Posgrado en Física Aplicada

como requisito parcial para la obtención del grado de

Maestro en Ciencias

por

Nestor Raul Mancilla Xinto

Asesorado por

Dra. María Isabel Pedraza Morales

Puebla Pue.
6 de Noviembre de 2023

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Search for Exotic Particles with the CMS detector at the CERN LHC

Tesis presentada al

Posgrado en Física Aplicada

como requisito parcial para la obtención del grado de

Maestro en Ciencias

por

Nestor Raul Mancilla Xinto

Asesorado por

Dra. María Isabel Pedraza Morales

Puebla Pue.
6 de Noviembre de 2023

Título: Search for Exotic Particles with the CMS detector at the CERN LHC
Estudiante: Nestor Raul Mancilla Xinto

COMITÉ

Dr. Humberto Antonio Salazar Ibargúen
Presidente

Dra. Iraís Bautista Guzmán
Secretario

Dr. Justiniano Lorenzo Días Cruz
Vocal

Dr. Jhovanny Andres Mejia Guisao
Vocal

Dra. María Isabel Pedraza Morales
Asesor

Agradecimientos

A mi familia, especialmente a mis padres, quienes siempre me han brindado su apoyo en mis estudios y se alegran de poder sacar adelante a sus hijos de la mejor manera posible. A mis hermanos, que de una u otra forma siempre me han ayudado. Agradezco profundamente toda su ayuda, ya que gracias a ella he logrado presentar y concluir de la mejor forma este período y trabajo.

A mis amigos, en particular a ese grupo de personas del grupo de RPC, que me respaldaron en los momentos más difíciles. Siempre les estaré agradecido. Quiero expresar mi agradecimiento especial a Breno, por escucharme y por ofrecerme una perspectiva diferente sobre las cosas. Valoro mucho el tiempo que compartimos juntos, y estoy seguro de que nos encontraremos nuevamente en otro momento. A Pedro, es reconfortante saber que, a pesar de la distancia, siempre está ahí. A Diana, quien siempre ha estado dispuesta a apoyarme.

A mis amigos que se encuentran lejos, me llena de alegría saber que están haciendo lo que más les gusta, y nuestra amistad perdura a pesar de la distancia. Quiero agradecer a mis amigos de la facultad, incluyendo a aquellos que ya no continúan en ella. También, mi reconocimiento a mi grupo de amigos de la preparatoria. A todos les agradezco por su paciencia al ajustarse a mis horarios y actividades. Asimismo, quiero expresar mi gratitud al grupo de RPC, que me brindó enseñanzas valiosas y me permitió vivir nuevas experiencias y crear nuevas y grandiosas amistades. En definitiva, mi agradecimiento va dirigido a todos, sin importar la magnitud de su contribución, ya que cada uno ha sido parte importante en la obtención de estos resultados.

Contents

List of Figures

List of Tables

Summary

After discovering the Higgs boson in 2012, the necessity to find new physics beyond the Standard Model arose to solve the questions without answers. We are interested in the Mono-Higgs dark matter model, a simplified model of dark matter particles coupled to a new Z' boson, which can decay into a par of Standard Model particles, visible as par of b quarks. In this master thesis, we focus on the obtention of the scale factors to correct a double b tagger used for the decay of a dark Higgs boson, which needed to compare the background events appropriately with the obtained data from Run 2 at the LHC.

Introduction

The latest great discovery in particle physics was the Higgs boson, reported by the LHC at CERN in 2012 [74, 9]. Because of this, new gates were opened to search for new physics, as the Standard Model is only a partial theory. Questions such as neutrino mass, the hierarchy of mass, dark matter, and dark energy are still unsolved. Furthermore, the Higgs boson allows new annihilation channels, like dark matter (DM), involving new particles.

From astrophysical observations, we know that most of the matter in the universe is comprised of dark matter [102]. Probes of dark matter existence are, for instance, the rotational curves and gravitational lensing [103]. The first probe referred to the galaxy's speed when we expected that the speed would decrease far from the center in classical models. However, the observations show that the speed increases when the distance from the center increases. This effect is expected because of DM on the galaxy. The second one, predicted by Albert Einstein, says that if there is a massive foreground object from the observer to the source, two images form on both sides of the source as if a lens lay between the observer and the object. When a dark object in the Milky Way passes by a distant star, it is briefly illuminated as a smaller-scale gravitational lensing effect. [103] .

Three types of dark matter searches exist: direct, where proposed the interaction between DM particles and SM particles leading the recoil of atomic nuclei [122] ;indirect, searching for DM annihilation channels [62] and colliders [11]. One characteristic involving searches in colliders is missing transverse energy on final states, called mono-X searches [39]. This is because DM does not produce signals on the LHC, and one way to observe them is in association with SM particles, where X refers to g, q, γ, etc.

Considering the last sentence, the idea exists that DM particles interact with SM particles via the exchange of new mediators. Some models proposed the interaction with a new Z' boson [18], leading to new annihilation channels.

In this thesis, we focus on the scale factors measurement to correct the efficiency of a tagger used to identify a fat jet from a b quark pair as part of the signal event model. These scale factors are measured using data from Run 2 in Quantum Chromodynamics (QCD) proxy jets to apply in the signal jets.

Chapter 1

Dark Matter

The Standard Model (SM) is a theory that describes elementary particles and their interactions at the subatomic level. The last great success of the model was in 2012 with the Higgs boson discovery announcement [74, 9]. However, despite the results, astrophysical observations indicate a significant amount of another type of matter in the universe not described in the SM called Dark Matter (DM). These observations gave us some characteristics of the DM, and different models and types of detections are looking for valid answers.

1.1 Standard Model

The development of the Standard Model (SM) began a long time ago, with the idea of matter composed of atoms in the XIX century, continuing with a list of discoveries such as the discovery of the electron by Joseph John Thomson and finishing with the discovery of the Higgs boson.

The SM theory describes the fundamental forces and their interaction with elementary particles. Fermions and bosons characterize the elements defined by the SM. A fermion is a particle that obeys the Fermi-Dirac statistics, and its spin is fractional. Meanwhile, a boson follows the Bose-Einstein statistics and has an integer spin. Fermions (spin 1/2) are divided into two categories: Quarks with a color charge and participate in strong interactions, and Leptons without a color charge and participate in electromagnetic interactions. In the quarks category, it is the up, down, charm, strange, top, and bottom quarks. In the leptons category, it is the electron, the muon, the tau, and its corresponding neutrino (neutrino mass is not considered in the SM) Fig 1.1.

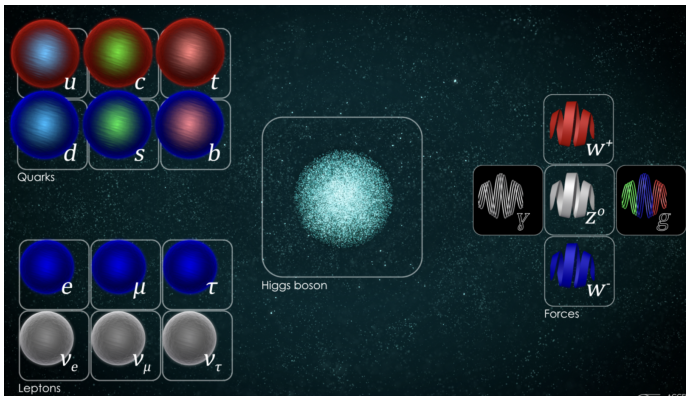

Figure 1.1: Particles of the Standard Model [61]

The SM framework is described by the quantum field theory by the use of the Lagrangian, which can be divided into its interactions. The forces described by the SM are electromagnetic, strong, and weak nuclear forces; the gravitational force is not included in the SM. A boson mediates each force: the γ, the Z and W for the electromagnetic force, and the gluons for the nuclear force. The last boson is the Higgs boson, produced by the quantum excitation of the Higgs Field. This field is responsible for giving the particles their mass. Correlated to these interactions are the symmetries, with the $SU(3)_C \times SU(2)_L \times U(1)_Y$ a non-Abelian gauge group describing the SM. The $SU(3)_C$ group is associated with eight gluonic fields G^{α}_{μ}, the $SU(2)_L$ group is associated with three electroweak fields W^{i}_{μ}, and the $U(1)_Y$ group is associated with the hypercharge field B_{μ}.

The SU(2)×U(1) called electroweak group comprises the special unitary matrix 2×2 of isospin L and the hypercharge group Y, unifying the electromagnetic interactions and the nuclear weak. The SU(3) color group comprises the special unitary matrix 3×3 made by the quantum chromodynamics characterized by strong interactions.

The SM is considered the gauge norm, and as an introduction, it is considered the quantum relativistic Lagrangian for a free Dirac particle:

$$\mathcal{L} = \overline{\psi}\left(i\gamma^{\mu}\delta_{\mu} - m\right)\psi; \tag{1.1}$$

if ψ describes a free electron, the considered Gauge transformation is

$$\psi \rightarrow \psi' = \exp[iq\alpha(x)]\psi, \tag{1.2}$$

where $\alpha(x)$ is a function, and q is the charge associated with the Gauge group U(1). Considering a vectorial field A_{μ} associated with the photon that transforms as:

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} - \partial_{\mu}\alpha(x), \tag{1.3}$$

to be invariant under this transformation, it is introduced the following term:

$$D_{\mu} = \partial_{m} + iqA_{\mu}. \tag{1.4}$$

To obtain the complete quantum electrodynamics (QED) Lagrangian it is introduced the invariant term that describes the free electromagnetic field is:

$$-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}, F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}. \tag{1.5}$$

Therefore, the quantum electrodynamics (QED) Lagrangian is written as

$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \overline{\psi}\left(i\gamma^{\mu}D_{\mu} - m\right)\psi. \tag{1.6}$$

The quarks and gluons dynamics are ruled by the quantum chromodynamics (QCD) Lagrangian

$$\mathcal{L}_{QCD} = \sum_{f}\overline{\psi}^{(f)}_{j}\left(i(\gamma^{\mu}D_{\mu})_{jk} - m_{f}\delta_{jk}\right)\psi^{(f)}_{k} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}, \tag{1.7}$$

where $\psi(x)^{(f)}$ are three spinor components and represent the different flavor quarks (f=u,d,c,s,t,b). The $G_{\mu}\nu^{a}$ represents the identity tensor of the gluon field given by [65]

$$G_{\mu}\nu^{a} = \partial_{\mu}A^{a}_{\nu} - \partial_{\nu}A^{a}_{\mu} + gf^{abc}A^{b}_{\mu}A\nu^{c}, \tag{1.8}$$

where $A^{a}_{\mu}(x)$ are the eight associated fields to each gluon and f^{abc} are the structure constant of the SU(3) group.

To explain the masses of the Z and W bosons, a single $SU(2)_L$-doublet scalar field is introduced and causes the spontaneous breaking of the $SU(2)_L \times U(1)_Y$ gauge symmetry by the Higgs mechanism. This scalar field can be written as

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.9}$$

The scalar field Φ is associated with the hypercharge Y=1/2, and the terms in the Lagrangian associated with ϕ are given by

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) + \mathcal{L}_{\text{Yuwaka}}. \tag{1.10}$$

The meaning of the terms are as follows; the first term contains the kinetic and gauge-interaction terms via the covariant derivative, the second term is a potential energy function, and the third term contains Yukawa couplings of the scalar field to pairs of fermions. The expression for the scalar potential is written as

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \tag{1.11}$$

This equation describes the scalar potential of the Higgs field. The first term is a quadratic term with a negative coefficient, which means that the potential energy is minimized when the Higgs field has a non-zero value. The second term is a quartic term with a positive coefficient, which helps to stabilize the potential. Considering the covariant derivative

$$D_\mu = \partial_\mu - i\frac{g'}{2} B_\mu - i\frac{g}{2} W_\mu^a \sigma^a, \tag{1.12}$$

where σ is the Pauli matrix, g is the SU(2) gauge coupling, g' is the U(1) gauge coupling, and v the Higgs vacuum expectation value. Using the Equation 1.11 written in terms of four real scalar degrees of freedom and minimizing the potential, through the Higgs mechanism, it is possible to obtain the mass terms for the W and Z bosons [98]:

$$M_W^2 = \frac{g^2 v^2}{4}, \tag{1.13}$$

$$M_Z^2 = \frac{(g^2 + g'^2) v^2}{4}. \tag{1.14}$$

In this process, the state identified as the photon does not couple to the Higgs field and thus does not acquire a mass through the Higgs mechanism. Knowing the W mass, using Fermi's constant, it is possible to obtain an estimate for the vacuum expectation values $v = 246$ GeV. The mass of the leptons and quarks is obtained using a gauge-invariant renormalizable Lagrangian with a similar process. Furthermore, using the equation for the Higgs potential 1.11, considering the self-interactions of the Higgs, it is possible to obtain a term for the Higgs mass

$$-\lambda v^2 = -\frac{m_H^2}{2}, \tag{1.15}$$

with a measurement of $m_H = 125.25 \pm 0.17$ [17]. Because of the parameters that appear in the Higgs couplings Feynman rules for this method, the prediction of the partial widths for all the decays (and hence the decay branching ratios) has been made, presented in Fig 1.2.

1.2 Dark Matter observations

As expressed in the previous section, DM is one of the topics not described in the SM. The "Dark Matter" term arose in 1933 by Fritz Zwicky [115] to describe the unseen matter that must dominate one feature of the universe—the Coma Galaxy Cluster. This observation showed the

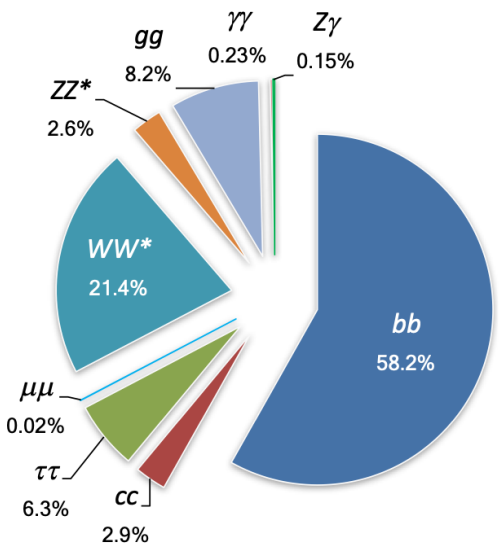

Figure 1.2: Branching fractions predicted for an SM Higgs boson of 125 GeV mass [86].

Coma Cluster has approximately two hundred more times mass than is visible from the galaxies. The DM's existence was confirmed in 1970 by Vera Rubin and W. Kent Ford, who found evidence for dark matter in their research on galaxy rotation. They found galaxy velocities larger than they expected because they believe that the galaxy velocity reduces when increasing the radius distance of the galaxy center. However, observations show velocity increases when increasing the radius distance and then is almost constant (See Fig 1.3).

Nowadays, we know that baryonic matter represents less than five percent of the total components of the universe, that it is the majority composed of dark energy (72 %), and the rest is dark matter (23 %) [102]. Among the observations for DM are the gravitational lensing, the speed of the Galaxies, CMB anisotropy, cluster kinematics, etc. For reference and better understanding, the explanation of the two first probes is as follows: As Einstein predicts, massive objects curve spacetime, and light passing will be curved by these objects. Using this consequence, it is possible to detect the effect of a massive galaxy cluster that theoretically contains DM, distorting the light of distant galaxies behind the cluster. This is called gravitational lensing Fig 1.4 [101].

Another proof is the galaxy velocity. One of the easiest models is the Kepler model, which assumes the galaxy's mass is at the center, and the stars have the same angular velocity. From the gravitation law, where G is the gravitational constant, and M is the galaxy's center of mass, the formula is

$$v(r) = \sqrt{\frac{GM}{r}}. \tag{1.16}$$

We can simplify the formula due to G, and M being constants, so instead of the formula 1.16, we can only use $1/\sqrt{r}$ and obtain a curve like the one presented in Fig 1.3. The observations show that there is more mass around the galaxy, explaining the reason when increasing the distance also, the velocity increases.

From these statements, we can conclude that DM must have these characteristics

- must be electrically neutral; otherwise, they would scatter light and thus not be dark,
- has to be non-relativistic (cold) at the epoch of structure formation. Relativistic (hot) dark matter has a larger free-streaming length that leads to inconsistencies with observations,
- Given the measurement of the total matter content; we must conclude that DM is non-baryonic,

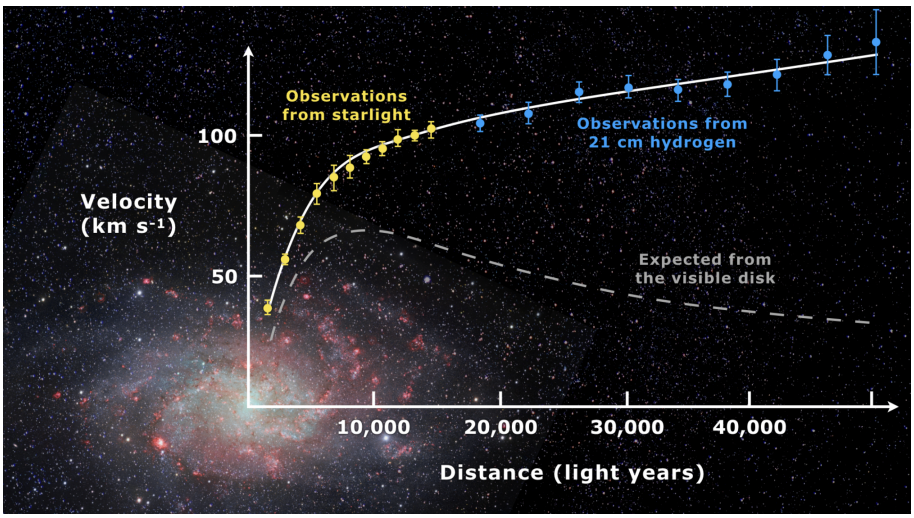

Figure 1.3: Rotation curve of spiral galaxy Messier 33. The discrepancy between the two curves can be accounted for by adding a dark matter halo surrounding the galaxy [97].

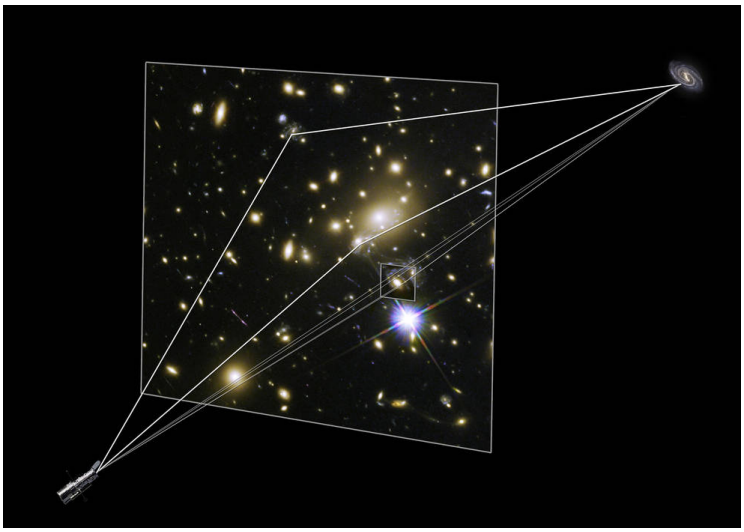

Figure 1.4: The sketch shows paths of light from a distant galaxy that is being gravitationally lensed by a foreground cluster [101].

- is a long-lived (if not stable) particle.

Type	Particle spin	Approximate Mass Scale
Axion	0	μeV-meV
Inert Higgs Doublet	0	50 GeV
Sterile Neutrino	1/2	keV
Neutralino	1/2	10GeV-10TeV
Kaluza-Klein UED	1	TeV

Table 1.1: Properties of various Dark Matter Candidates [28].

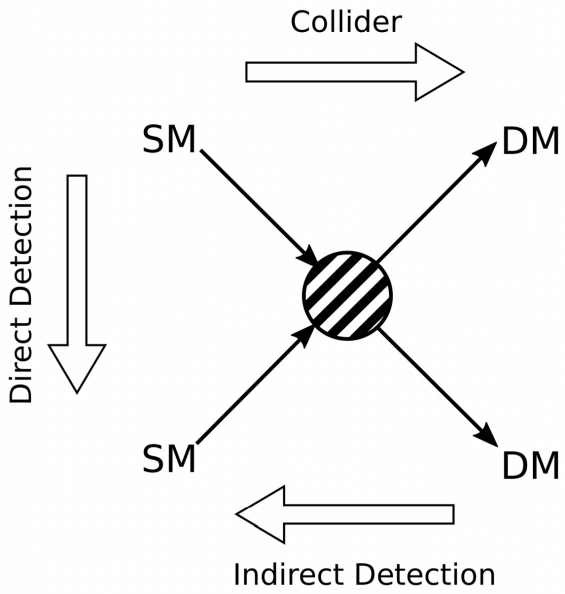

Figure 1.5: Schematic showing the possible dark matter detection channels [82].

Some models proposed that the late-time abundance of DM is controlled by its couplings to the SM; however, there exist exceptions. But if DM self-annihilates and was once in thermal equilibrium, the late-time relic abundance can be entirely determined by the cross-section for the annihilation process. This is typically called a "thermal relic" scenario [121] that measures the annihilation cross-section used for its detection. Of the candidates for makeup DM, some of them are described in Table 1.1.

1.3 Detection

Searches of DM are not trivial task. Several experiments use different techniques for this purpose, and we can summarize the types of DM detection in three categories, and each approach strictly correlates with the couplings of interactions. These methods are described as DM production at particle accelerators, indirectly by searching for signals from annihilation products or directly via scattering on target nuclei (Fig 1.5). This section aims to describe briefly these techniques.

1.3.1 Direct

This method is based on searching for the scattering between DM particles and nuclei in a suitable detection medium. The process is only observable through the recoiling nucleus, with energy E_R [40] given by

$$E_R = \frac{1}{2} m_\chi v^2 \frac{4 m_\chi m_N}{(m_\chi + m_N)^2} \frac{1 + \cos\theta}{2}. \tag{1.17}$$

This interaction implies the existence of the scattering of dark matter with ordinary matter. There are several detection techniques to measure the recoil of the nucleus from dark matter scattering. One of these techniques is the scintillator experiments, using crystals or liquids (Xe, Ar, Ne, etc.) to emit light collected by photomultipliers. Experiment examples of this technique are DAMA/LIBRA [29], Xmass [10], and MiniCLEAN [90]. Another method used for the experiments, CREST [56] and CUORICINO [70], focuses on observing photons. The last strategy focuses on ionization; examples of these experiments strategy are HDMS [25], MAJORANA [55], and IGEX [73].

1.3.2 Indirect

The first aim of indirect detection is to observe the radiation produced in dark matter self-annihilation. Direct DM detection often focuses on searching for SM particles produced by the decay or annihilation of DM or their secondary effects [121]. However, their interaction with SM particles is too weak, so the particle production rate is expected to be small. On the other hand, this detection type has the potential to answer the question if DM is stable as predicted. With its long lifetime, it would not decay to produce visible particles detected in collider experiments unless DM is a new particle decaying at some point to be observable.

Some indirect searches for DM use γ rays, X rays, charged cosmics rays, neutrinos, etc. Examples of this kind of search are the experiments located in space like AMS [116], EGRET [58], FERMI [57], etc., and Earth experiments are AMANDA [32], ANTARES [68], IceCube [87], etc.

1.3.3 Colliders

The purpose of collider searches is to detect signals coming from DM particles produced when collider SM particles in controller collisions. Besides, this method can probe the interaction between SM and DM particles by searching for the interested process's mediator [35]. However, the lifetime detection of the particles is limited by the time to exit the detector. Therefore, the main thrust of this method is to interpret the missing-energy signatures at colliders because DM does not produce signatures and capture the relevant kinematics properties with its model parameters. Comparing the results of searches for DM production at colliders with direct and indirect detection is possible due to the direct relation of the nucleon scattering cross section $\sigma_{SD,SI}$ and the mass of the dark matter candidate m_{DM} involving vector, axial-vector, scalar and pseudoscalar mediators [76].

Examples of this method are the Large Hadron Collider at CERN [12] and the Tevatron at FERMILAB [22].

Chapter 2

CMS detector

The Large Hadron Collider (LHC) at CERN was built to accelerate protons or ions and collide at a speed close to light speed at four points around the machine. At one of these points is the Compact Muon Solenoid (CMS), a multipurpose experiment to detect the particles generated in a proton-proton collision at the LHC. It is designed cylindrically 15 meters in diameter and 21.6 in length. One of its main features is a superconductor solenoid with a 4T magnetic field, which helps to identify the detected particles.

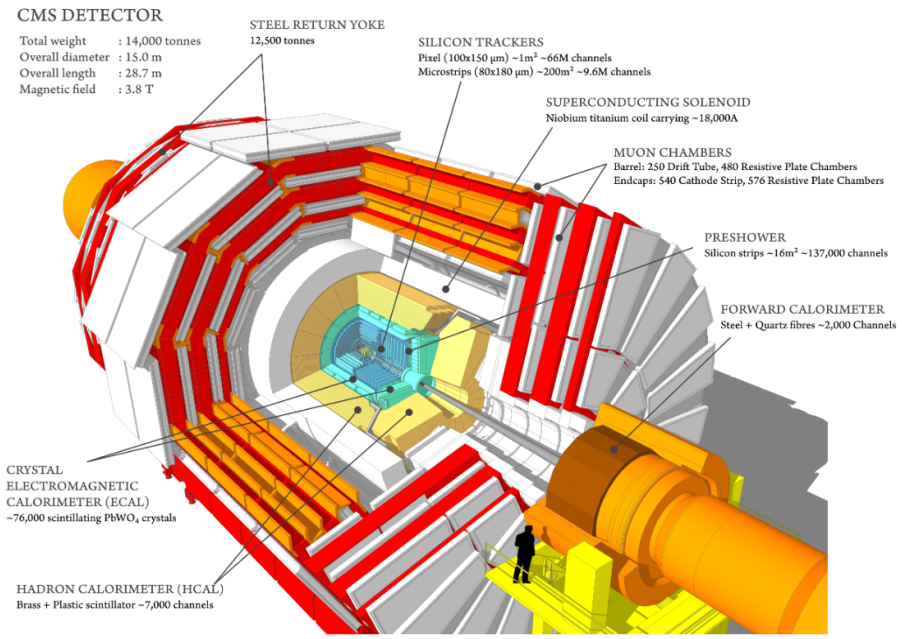

Figure 2.1: CMS components [124].

The CMS comprises sub-detectors, as shown in Fig 2.1: The inner tracking system, electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL), superconducting magnet, and muon detectors. The purpose of this chapter is to detail the structure of the CMS detector; each subdetector will be described to give an idea of the detector's aim.

2.1 The Large Hadron Collider

The Large Hadron Collider, constructed by the European Organization for Nuclear Research (CERN), is a circular accelerator that started on 10 September 2008 and consisted of a 27-kilometer two-ring-superconducting-hadron accelerator [77] located on the border between France and Switzerland. In addition to hadron collisions (proton-proton), the LHC is also capable of nucleus-nucleus collisions. The protons are obtained from the ionization of hydrogen molecules, which are accelerated in a chain of different linear and circular accelerates to an energy of 450 GeV before being injected into the LHC [27].

The LHC is divided into 8 points (See Fig 2.2), and the collisions occur at only 4 points. At this points, a main detector is located:

- Point 1: ATLAS (A Toroidal LHC Apparatus). A general-purpose experiment. It researches the SM, extra dimensions, and possible candidates of dark matter particles.
- Point 5: CMS (Compact Muon Solenoid). The second general-purpose experiment. Similar to ATLAS, it studies the Standard Model, extra dimensions, and dark matter particle candidates.
- Point 2: ALICE (A Large Ion Collider Experiment). Studies quark-gluon plasma correspond to strongly interacting matter at extreme energy densities.
- Point 8: LHCb (LHC-beauty). It investigates the slight differences between matter and antimatter through b quarks.

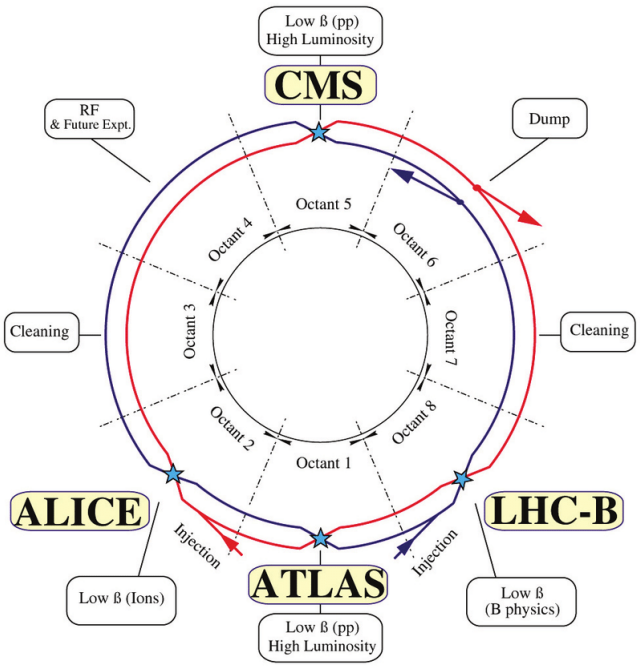

Figure 2.2: Schematic layout of the LHC [123] (Beam 1- clockwise, Beam 2 — anticlockwise).

The 2010–2012 LHC running period is referred to as Run 1. In 2010 and 2011, the LHC operated at a center-of-mass energy of 7 TeV, increasing to 8 TeV in 2012. After Run 1, there was a long shutdown 1 (LS1) during which the accelerator and the experiments underwent several improvements allowing collisions at a center-of-mass energy of 13 TeV during Run 2 from its start in 2015 until the end of 2018 [49]. After the second shutdown (LS2) in July 2023, Run 3 started a new data-taking period with a collision energy of 13.6 TeV, waiting for new physics discoveries.

2.2 Compact Muon Solenoid detector

The Compact Muon Solenoid detector (CMS) [52] is a general-purpose detector that analyses the SM and searches for new physics. It is made as an onion, then means that the subsystems are around the beam pipe (See Fig 2.1 and 2.3). Starting from the interaction point in the middle of the detector, the first subsystem is the tracker system used to measure the particle's momentum, track its path and determine its charge. The Electromagnetic calorimeter and the Hadron calorimeter cover the tracker. The first one measured the energy and stopped the photons and electrons; the second measured the energy of neutral and charged hadrons. After the subdetectors is the solenoid with a magnetic field of 3.8T that curves the charged particles and is used to identify its momentum. Finally, the last subdetector is the muon system. The signals collected by the detector allow us to reconstruct physical objects, for example, the jets' energy, reconstruction, and tagging and the muon identification by the muon system. More details about these physics objects are given in the next chapter.

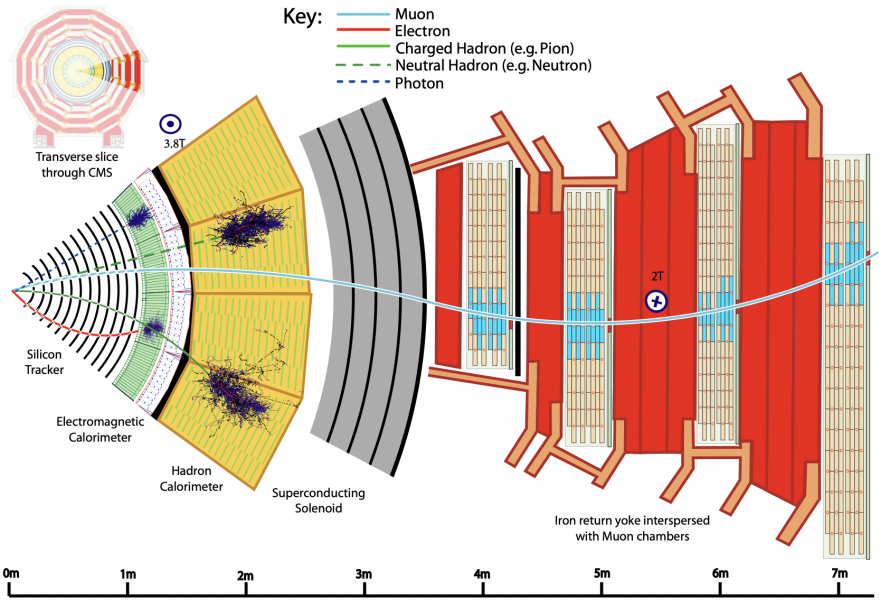

Figure 2.3: Asketch of the specific particle interactions in a transverse slice of the CMS detector [69].

The full length of the CMS detector is 21.6 m, a diameter of 14.6 m with an inner diameter of 5.9 m and a total weight of 12500 tons. The characteristics of the magnet and the CMS subdetectors are described in the next subsections.

2.2.1 Inner tracking system

The inner tracking system [92], the closest subdetector to the interaction point with an outer radius of 1.20 m and a length of 5.6 m, designed to measure the trajectories of charged particles and for second vertices reconstruction [52]. Due to the high number of particles the LHC produces, a fast response detector is required for correct trajectory identification. The tracker is based on silicon detector technology because of the radiation damage of the particle flux and is composed of a pixel detector with three barrel layers at a radius between 4.4cm and 10.2cm and a silicon strip tracker with 10 barrel detection layers extending outwards to a radius of 1.1m completing each system with endcaps with a total acceptance of the tracker up to a pseudorapidity of $|\eta| < 2.5$. The CMS tracker layout is presented in Fig 2.4 with its subsystems: Pixel, Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID), Tracker Outer Barrel (TOB), and Tracker Endcaps (TEC).

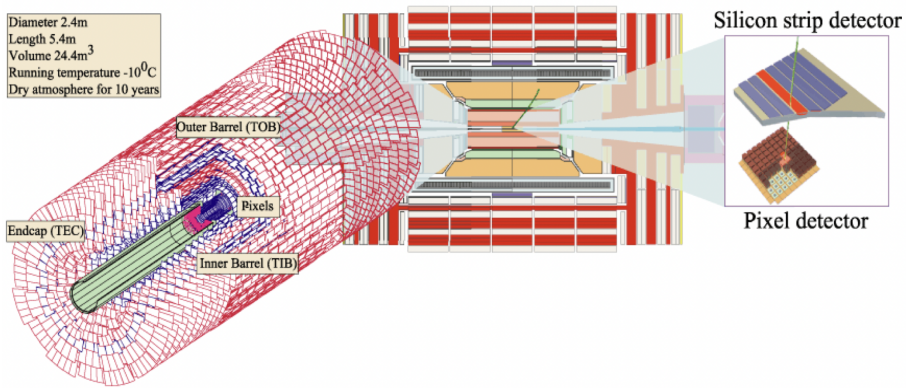

Figure 2.4: The layout of the CMS tracker [24].

The pixel detector is installed 20 cm from the beam pipe and designed to identify the trajectories. It covers a pseudorapidity range $-2.5 < \eta < 2.5$ and contributes tracking points in $r - \phi$ and z that help reconstruct secondary vertices. The active area of 1.1 m^2 is arranged in four cylindrical barrels and three endcap layers on each side of the barrel.

The silicon strip tracker contains a total active area of about 200 m^2 placed around the pixel detector covering a region up to 116 cm from the beam pipe and provides two coordinate measurements in at least approximately four hits for the pseudorapidity range $|\eta|$ <2.4. It is divided into four subsystems: The tracker inner barrel (TIB), the Tracker Outer Barrel (TOB), the Tracker Inner Disks (TID), and the Tracker EndCaps (TEC).

2.2.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL)[4], used for energy measurement of photons and electrons, is a hermetic homogeneous calorimeter made of tungsten crystals with good radiation resistivity and mounted in the central barrel part closed by the two endcaps. The barrel length is 23 cm, and the endcap length is 22 cm. The ECAL uses photomultipliers called avalanche photodiodes (APDs) for the barrel and vacuum phototriodes for the endcaps.

The ECAL is organized in Barrel ECAL (EB) and Endcap ECAL (EE) (See Fig 2.5). The EB covers the pseudorapidity range $|\eta| < 1.479$, and its granularity is 360-fold in ϕ and (2×85)-fold in η, with a total of 61200 crystals that form supermodules. Due to the high granularity, separating and identifying particles with a distance of around 5cm is possible [27]. The energy resolution of

the EB was measured with an ECAL supermodule directly exposed to an electron beam [69]. The EE rapidity range covers $1.479 < |\eta| < 3.0$, divided into two halves or Dees that hold 3662 crystals each. In addition to the Dee, in front of the endcaps, the Preshower is installed made of two layers, each comprising a lead radiator followed by a plane of silicon strip sensors and provides information on the electron identification and identification of neutral pions decaying into two photons.

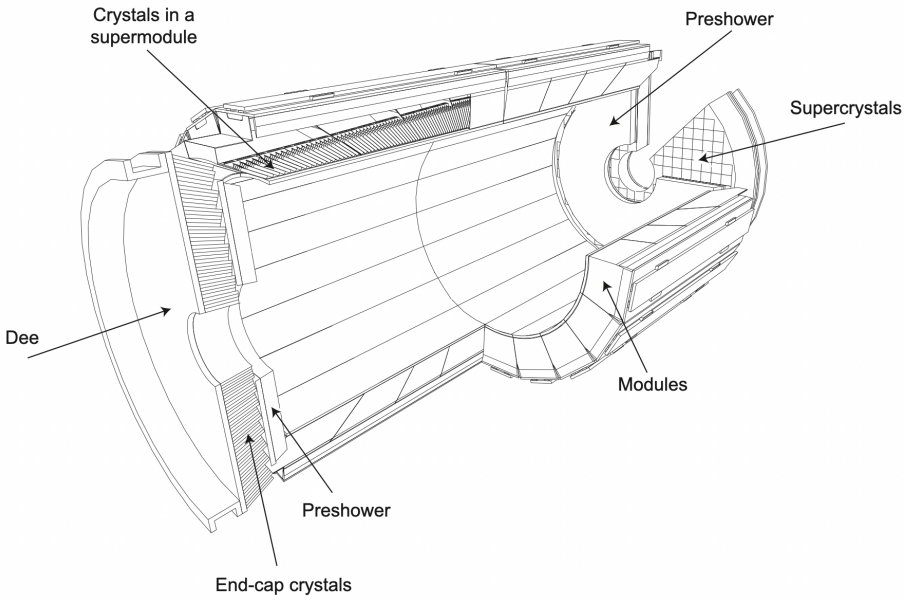

Figure 2.5: CMS electromagnetic calorimeter layout [52] showing the arrangement of crystal modules.

2.2.3 Hadronic calorimeter

The Hadronic Calorimeter (HCAL) [5] is a hermetic sampling consisting of several layers of the brass absorber and plastic scintillator tiles and is located inside the magnet coil and surrounds the ECAL system. The subdetector measured the missing transverse energy of the events, as well as the direction and energy of hadronic jets. The HCAL and ECAL form the calorimetry system of the CMS detector for measuring jets and missing transverse energy.

The subsystem comprises four parts: the barrel (HB), the endcap (HE), the outer (HO), and the forward calorimeter (HF), as shown in Fig 2.6. It comprises brass and scintillator layers arranged in a tower pattern (see Fig 2.7) in η, ψ space. It detects the tracks passing through the detector and converts them into an electrical signal proportional to the energy. The brass is chosen as the absorbing material, and the scintillators can carry the light to the readout system.

The HB is an assembly of two half barrels made by flat brass alloy absorber plates parallel to the beam axis with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The HO contains scintillators with a thickness of 10mm and covers the region $-1.26 < \eta < 1.26$, and it improves the E_T^{miss} resolution of the calorimeter. The HO is located inside the barrel muon system and is divided into 5 "rings" along η, -2, -1, 0, 1, and 2, and has 12 sectors. The HE comprises 14 η towers covering the pseudorapidity region $1 - 3 < |\eta| < 3.0$. The HF is made of steel/quartz fiber and covers the pseudorapidity of 3.0 and 3.5.

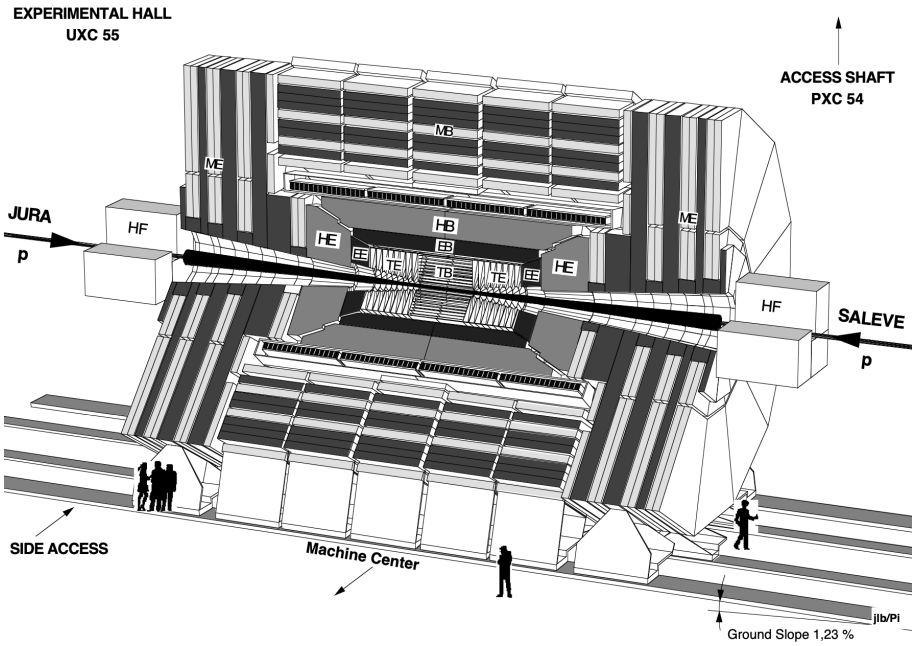

Figure 2.6: CMS components of the Hadronic Calorimeter [125]

The HF is designed to measure energetic forward jets to discriminate the narrow lateral shower profile and to increase the hermeticity of the missing transverse energy measurement. Central shower containment in the region $|\eta| < 1.26$ is improved with an array of scintillators located outside the magnet in the HO [26].

The subcomponents, HB, HE, and HO, receive the optical signals from the scintillators and then are converted to electrical signals using multichannel hybrid photodiodes (HPDs). However, in the HF, conventional photomultiplier tubes are used. The Secondary vertex reconstruction is explained using the Inner Tracking system and the calorimeters, detailed in the next chapter.

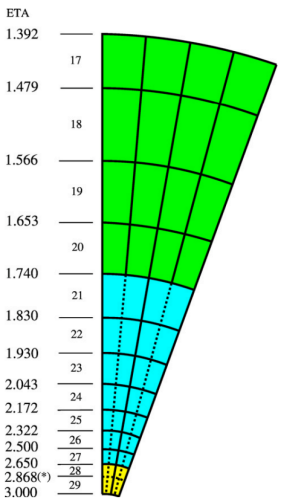

Figure 2.7: $\eta - \phi$ view of a 20-degree HE endcap sections showing the 5-degree regions and "split" 10-degree regions above $\eta = 1.740$ in detector pseudo-rapidity [26]

2.2.4 Magnet

The main feature of the CMS detector is a large superconducting solenoid magnet [93] with a homogeneous magnetic field of 3.8 T on the inside, with a 12.5 m length and a free-bore radius of 3.15 m Fig 2.8. Its design avoids increasing the material in front of the calorimeters not to harm the energy resolution because the solenoid is big enough to contain the tracker and the calorimeter system. Reinforced NbTi wire coils made the solenoid because they can produce a uniform magnetic field when the electricity flows and is at a low temperature (-268.5 degrees Celsius) [52]. The solenoid is closed for 12 500 tones steel return yokes, composed of 11 large elements, six endcap disks, and five barrel wheels, and the muon system is placed over this piece. The solenoid helps curve the particle trajectory detected in the tracker, and the muon system determines their momentum, which is an advantage for particle flow reconstruction.

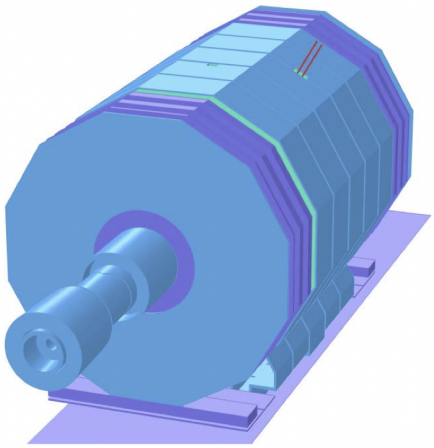

Figure 2.8: CMS magnet 3-D model [94] evaluated with the TOSCA program. The different colors correspond to the steels used in the magnet flux return yoke.

2.2.5 Muon System

The muon system [96] completes the CMS detector. As is written in the name, one of the purposes of the detector is muon identification. The muon system's general functions are transverse momentum measurement, muon triggering, and charge determination, where the tracker and the muon system makes the muon measurement.

The barrel and endcaps form the system, and it is composed of three types of gaseous detectors: Drift Tubes (DTs) in the barrel region ($|\eta| < 1.2$), Cathode Strip Chambers (CSCs) in the endcap region ($0.9 < |\eta| < 2.4$), and Resistive Plate Chambers (RPCs) both in the barrel and endcap regions ($1.2 < |\eta| < 1.6$). A schematic view of the system is shown in Fig 2.9.

The choice of the detectors depends on their properties; a characteristic of DT and CSC is that they can each trigger the transverse momentum of muons with good efficiency. Meanwhile, the RPCs can provide a fast, independent, and highly-segmented trigger with a sharp transverse momentum threshold over a large portion of the muon system's rapidity range ($|\eta| < 1.6$) [52].

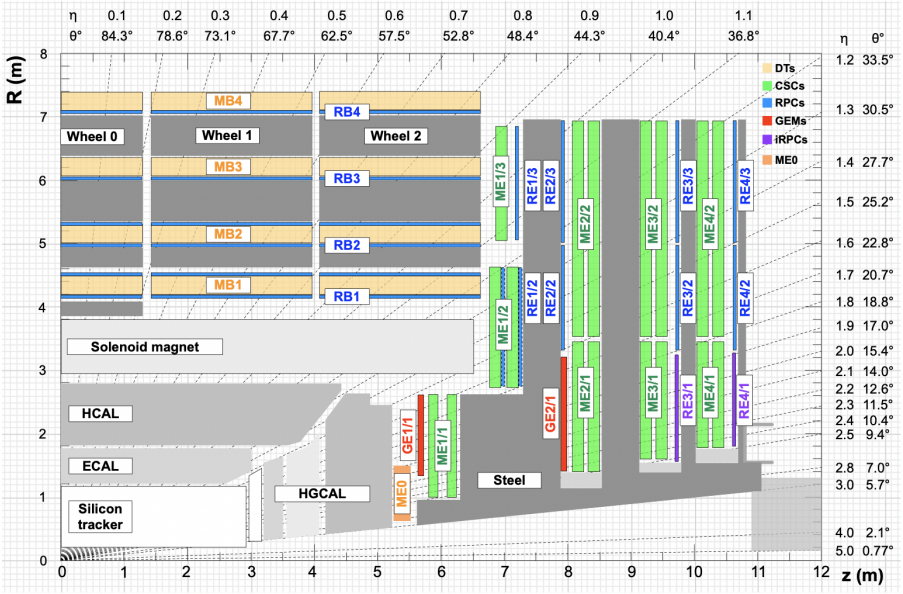

Figure 2.9: Layout of one quarter of the CMS muon system [105] including the Phase−2 upgrades (RE3/1, RE4/1, GE1/1, GE2/1, ME0). The interaction point is at the lower left corner. The dark gray areas represent the magnet yoke.

The CMS barrel muon system is divided into five mechanically independent wheels (W+2, W+1, W0, W-1, W-2) divided into 12 sectors. The endcaps have four disks called stations (ME1, ME2, ME3, ME4), divided into 36 sectors, closing each side of the barrel.

A short description of the gas detectors is described as follows:

- Drift Tubes: The primary detector in the barrel region covers a range of $|\eta| < 1.2$ and consists of 4 coaxial stations forming concentric wheels with a total of 250 DT chambers. DTs are gaseous detectors and produce ionization on them when passing charged particles. When the muons pass, they create electrons and ions, and the electrons with the strong electric field produce an avalanche and transform it into an electrical signal.

- CSC: gaseous particle detectors in four layers in each endcap covering the region $< 0.9|\eta| < 2.4$ with a total of 540 detectors. The CSC detectors are multi-wire proportional chambers that measure the azimuthal coordinate of passing muons for transverse momentum estimation and provide the radial coordinate of the passing muons. Like DTs, the passing muon through the gas detector causes an electron avalanche producing a signal on the anode wires and inducing a distributed charge on the cathode strips.

- RPC: gaseous detectors distributed into the barrel and endcaps, covering the range of $|\eta| < 1.9$ with a total of 1056 detectors, 480 in the barrel region distributed into the five wheels divided into 12 sectors (S01-S12), and 576 in the endcap, distributed into eight disks, where the chambers are trapezoidal indicated as a segment CH01 through CH36 whiting a ring. RPCs work with a gas mixture composed of 95.2% $C_2H_2F_4$, 4.5% iC_4H_{10}, and 0.3% SF_6 operating in avalanche mode to ensure good operation at high rates and have double gaps (see Fig 2.10) with a gas gap of 2 mm [26].

To extend the sensitivity for new physics searches, a major upgrade of the LHC is prepared, the High Luminosity LHC (HL-LHC), thus an upgrade on the CMS components is prepared, including the muon system, such as the addition of the GEM detectors and the installation of the improved RPCS (iRPCs) [6].

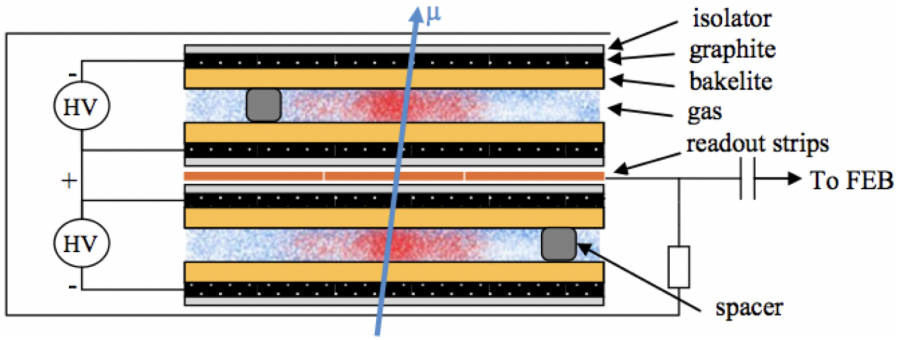

Figure 2.10: Working principle of the double gap RPCs in CMS [6].

2.2.6 Trigger and data acquisition

Due to the luminosity of the LHC, it is expecting a data rate of ∼40 TB/s that needs to be analyzed and collected by the detectors; However, it reaches the accepted rate of ∼ 100MB/s, and the data need to be reduced. For this reason, the CMS uses a two-step trigger system: Level 1 Trigger (L1) and High-Level Trigger (HLT). The first reduces the events rate to less than 100 kHz, and the second gives the final output data rate of about 100 Hz [118]. A schematic view of the CMS trigger system is presented in Fig 2.11.

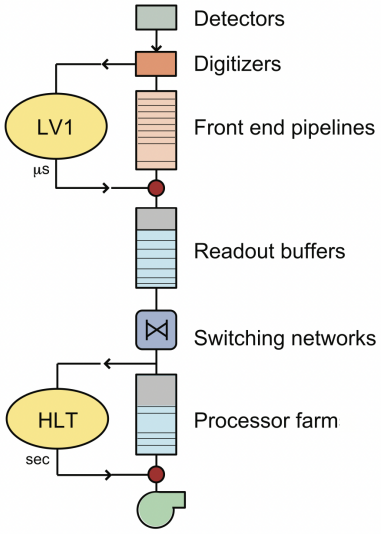

Figure 2.11: Schematic view of the CMS trigger system.[43].

The first data selection step is the L1 trigger that comprises custom hardware processors and runs synchronously with the LHC bunch crossing frequency of 40 MHz, using the information from the calorimeters and detectors to select the most interesting events deciding whether or not to save the experimental data. This selection is based on the local objects like photons, electrons, muons, jets, and missing transverse energy. Next, after a fixed time interval of 3.2 μs, these events

pass through the HLT, which consists of a single processor farm called the "Event Filter Farm" and reduces the event rate to 100Hz and are written to a temporary disk buffer before being transferred to the computing center (Tier 0) at CERN for offline processing [100].

The Data Acquisition System (DAQ) combines the LT1 and the HLT, reading the output of the L1 trigger to merge all the events and then passes these events to the Event Filter Farm [52], where the HLT reduces the storage events. The DAQ is also the first source to see the detector's response where the collisions can be monitored, making the DAQ relevant for event selection and monitoring the CMS detector. The elements that compose the DAQ are the detector Front-Ends that store the data, the Readout Systems that read the data from the Detector Front-End System (FES), the Builder Network that provides interconnections between the components, the Filter Systems provided by the Readout, the Event Manager to control the data flow, the Computing Services that receive the filtered events, and the Controls for the interface, monitoring, and configuration of the DAQ [42].

Chapter 3

Object reconstruction

The electronic signals the detector gives have to be interpreted into particles after the collisions using the subsystems to identify each particle decay. Another important particle feature is momentum and energy, essential for several analysis signals and backgrounds. Some particles that leave a distinctive signature for the CMS detector are electrons, muons, taus, and photons. In the quark case, they hadronize and form spay hadrons visible as jets in the detector.

Using the signals (hits) that charged particles leave on the tracker system and their trajectory (track), the momentum can be determined by comparing it with the calorimeter measurement [27]. The link of this process, or the other subsystems, is the goal of the Particle Flow Algorithm [69]. The objective is the track and energy cluster reconstruction in the subsystems, the link between the subsystems, and identify the particles. The algorithm can reconstruct physical objects like photons, electrons, muons, and jets, muons missing transverse momentum, electrons, photons, and charged and neutral hadrons. Below is a summary of all physics objects and the relevant selection requirements for this analysis.

3.1 Primary vertex

The particle reconstruction position is essential to understand their origin; however, in a proton-proton collision, several processes are generated, and not all of them are relevant to the analysis. For this reason, it is important to define their vertices to identify the particle origin. The tracker hits are clustered and fitted using a Kalman-Filter [14] to reconstruct the primary vertices to obtain the track origin, momentum, and direction. The reconstruction occurred in three stages [69]: initial seed generation with a few hits compatible with the particle trajectory, trajectory building (or pattern recognition), and final fitting to determine the particle properties: origin, transverse momentum, and direction. All these tracks were required to originate from a few mm radii centered around the beam axis and to have p_T larger than 0.9 GeV.

Once the primary vertices are reconstructed, the one with the largest summed squared p_T of the tracks associated with that vertex corresponds to the event's hard scattering process called the leading vertex (LV). The CMS experiment at the LHC aims to study the rarest interactions of proton collisions. To increase the probability of these rare events, the LHC collides not just individual protons but rather large bunches of protons. This results in multiple proton interactions happening simultaneously, leading to a mixture of particles from the desired interaction and particles from other interactions, known as pileup interactions [8]. Those tracks do not originate from the primary vertex. One of the criteria parameters is the impact parameter, which ensures that tracks from a secondary vertex do not interfere. The LV is relevant for jets clustered with the anti-kt algorithm [37], which uses the tracks assigned to the vertex and the negative vector sum of these tracks.

3.2 Muons

The muon reconstruction is made by the muon system and the tracker system using the offline reconstruction and the HLT. They are identified using tracks in the tracker that are compatible with hits in the muon chambers. Also, compatibility with calorimeter deposits is required. A global fit of the tracks and muon path curvature is performed to estimate their charge and momentum. The muon system covers $|\eta| < 2.4$ restricting the muon reconstruction to this pseudorapidity range.

In the momentum range of 10–30 GeV/c, the dominant muon source is from b and c decays produced in jets, and the muon isolation tool helps to identify these jets and those coming from the decays of heavy objects [26]. The algorithm compares the total energy deposited (can be transverse energy in a calorimeter or the sum of transverse momenta of reconstructed charged-particle track) in a cone around the muon with a certain threshold. The geometrical definition of the cone is given by the condition $\Delta R \leq \Delta R_{\text{max}}$, where

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.1}$$

The muon contribution to the energy measurement inside the cone is called the veto value and can be subtracted to improve the discriminating power of the isolation algorithm.

There are three stages for muon reconstruction: local, standalone, and global reconstruction [26]. Local reconstruction begins from a "seed"; the compatible chambers are identified, and a reconstruction is performed in these chambers. Standalone reconstruction, also called Level-2 muon reconstruction, uses only information from the muon system; the tracker is not used. For this reconstruction, the muon system (CSC, DT and RPCs) only participates. In contrast, global reconstruction, also called Level-3 muon reconstruction, includes tracker hits.

In addition to the PF algorithm, there are other available algorithms [69]:

- Soft muon identification; it aims for muons from decays of quarkonia states. The tracker muon is required with a tighter matching to the muon segment. Also, the inner track is included for inner-tracker layer reconstruction.

- Tight muon identification; it focuses on muons from Z and W decays. It requires a global muon track, and the candidate should be a tracker muon with at least two matched muon segments in different muon stations and an inner track reconstructed from at least five inner-tracking layers.

Isolated global muons are selected by considering the inner tracks and calorimeter energy deposits with a distance ΔR to the muon direction in the (η, ϕ) plane smaller than 0.3 [69]. These criteria permit reject hadrons misidentified as muons. However, for muons inside jets, an additional criterion is required.

To nonisolated global muons, the tight-muon selection is applied [53] where it is also required at least three matching track segments be found in the muon detectors or deposits in the calorimeter associated with the tracks. More muon identification algorithms are described in ref [53], such as soft, tight, and particle flow muon selection. However, only tight and loose algorithms are used for this analysis, and their selection is described in the following sections.

3.3 Electrons

Electron reconstruction combines information from the inner tracker and the calorimeters due to their energy deposit in several crystals in the ECAL. The properties and technical problems for the tracking and the energy deposition patterns for electrons and photons are similar because electrons often emit bremsstrahlung photons, and photons often convert to electron pairs in the tracker

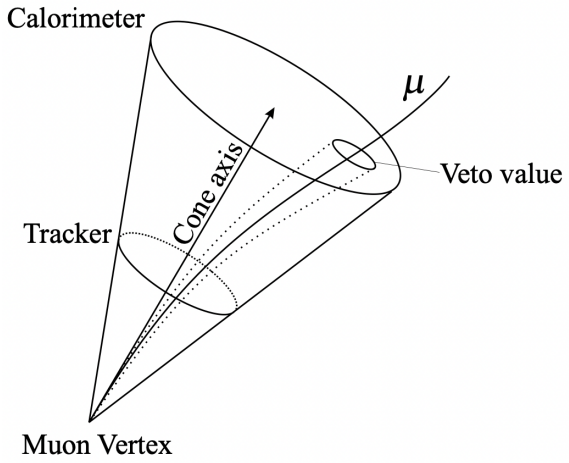

Figure 3.1: Schematic illustration of the isolation cone [26].

material [69], making the isolated photon reconstruction together with electron reconstruction.

In a given PF block, the super crystals (SCs, see Fig 2.5) match track seeds constructed from pairwise or triplet hits in the tracking system, and electron candidates are seeded from the Gaussian sum filter (GSF)[13] that estimates the electron track parameters. A photon candidate is seeded from an ECAL supercluster with E_T larger than 10 GeV, with no link to a GSF track.

At Level-1, the triggers used are A single isolated trigger, a double-isolated trigger, and a double trigger corresponding to 50% efficiency for electrons and photons with E_T of 23, 12, and 19 GeV, respectively [26]. For HLT, the selection is made in three steps: The first step, Level-2, uses only calorimeter information. The second step, Level-2.5, uses the information from the pixel detectors, and in the final step, Level-3, the full tracker information is used.

For single showers, the energy sums of fixed arrays of crystals give good energy resolution performance according to a cluster or "bump" finding algorithm. The Hybrid algorithm uses the $\eta - \phi$ geometry of the barrel crystals. It is designed to reconstruct high energy electrons in the barrel and reconstruction of electron showers down to $p_T = 5$ GeV/c. Meanwhile, the Island algorithm searches for seeds defined as crystals with an energy above a certain threshold [26].

The photon energy is reconstructed by summing the energy deposited in a cluster (by crystals or by the hybrid algorithm or island algorithm, depending on the spread of the deposited energy). On the other hand, electron reconstruction is initiated by the presence of electromagnetic superclusters. Their composition with a single track emerging from the interaction vertex forms the primary electrons.

Electron candidates must satisfy identification criteria such as the amount of energy radiated off the GSF track, the distance between the GSF track extrapolation to the ECAL entrance and the position of the ECAL seeding cluster, and the ratio between the energies gathered in HCAL and ECAL by the track-cluster association process. These criteria are combined for isolated and nonisolated electrons.

Photons are considered if they are isolated from other tracks and calorimeter clusters in the event and if the ECAL cell energy distribution is compatible with the ratio of the calorimeter energies (HCAL and ECAL) as expected in parton shower.

The distinction between electrons and photons in the PF global event description differs for some analyses. Thus, the electron and photon reconstruction recording is tracked and saved for different interpretations.

3.4 Jets

After muons, electrons, and isolated photons reconstruction, the remaining particles to be identified are hadrons from jet fragmentation and hadronization, which are identified as charged or neutral hadrons or nonisolated photons [69]. Within the tracker acceptance ($|\eta| < 2.5$), the ECAL clusters not linked to any track are turned into photons, and similar to the HCAL clusters not linked to any track, they are turned into neutral hadrons. However, beyond the tracker acceptance, they can not be distinguished, and ECAL clusters linked to a given HCAL cluster are assumed to arise from the same hadron shower. In contrast, the ECAL clusters without such a link are classified as photons.

These reconstructed hadrons are used to build the jets that, at LHC energies, the physics at high p_T is dominated by these physic objects, which can be used not only to understand the detector but also for searches beyond the Standard Model in models like SUSY, Higgs boson production, and other processes which require jet coming from high p_T quarks of gluons [26]. The jet energy resolution is essential to separate signal events from backgrounds. Also, missing transverse energy resolution is related to the calorimeter jet energy response.

In the HCAL towers, the associated energy to them is calculated as the sum of all contributing readout cells that pass the threshold and additional offline software thresholds [26]. The HCAL towers are used as inputs to several jet clustering algorithms associating the energy with the sum of all contributing readout cells. Some of the reconstruction algorithms are the iterative cone, the midpoint cone, and the inclusive k_T algorithm. The jets used in this analysis are reconstructed using the anti-k_T algorithm [37]. The anti$-k_T$ algorithm idea is to loop over the PF candidates and cluster them if they are within a certain distance, where the distance between two particles i and j is defined as

$$d_{ij} = min\{p_{T,i}^{2n}, p_{T,j}^{2n}\}\frac{\Delta R_{ij}^2}{R^2}, \tag{3.2}$$

where $p_{T,i/j}$ is the particle transverse momentum, n $\in$ -1, 0, 1, ΔR_{ij} is the particle distance in the $\eta - \phi$ plane defined as $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is a dimensionless parameter. R is the distance parameter that indicates the maximum jet radius. Thus, the second distance considering the particle p_T itself is defined as

$$d_{iB} = p_{T,i}^{2n}. \tag{3.3}$$

Using an iterative procedure, combining these two distances, the jet clustered is formed until all particles are within a jet [27]. The n parameter defines the order for clustering. The anti-k_T algorithm clusters all particles reconstructed by the PF algorithm (PF jets), the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo jers), or all stable particles produced by the event generator, excluding neutrinos (Ref jets)[69], as shown in Fig 3.2

After the clusterization, the momentum is defined as the sum of the PF clustered candidates.

The content of jets in terms of particle type and energy distribution depends on the flavor of the parton that initiated the jet, described by the fragmentation functions described in Ref [99]. However, the same energy corrections are applied to all jets because the parton that initiates the jet cannot be determined most of the time with sufficient confidence. Only charged hadrons from the primary vertex are considered to mitigate pile-up effects.

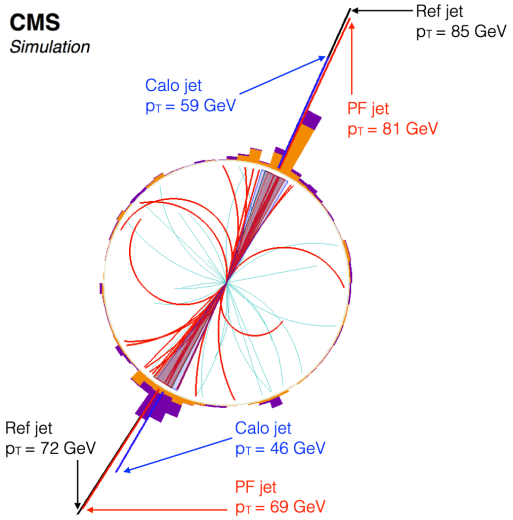

Figure 3.2: Jet reconstruction in simulated dijet event [69].

3.5 Missing Transverse Energy and Hadronic Recoil

The CMS can detect almost all stable or long-lived particles produced in pp collisions. Nonetheless, some particles exist that do not interact with the detector material, such as neutrinos or particles produced in the direction of the beam pipe. These particles are often called missing transverse energy [54] because they do not leave a signal in the detector. However, their presence can be inferred from the momentum imbalance in the plane perpendicular to the beam direction, denoted as $\not\!\!E_T$ or E_T^{miss}, to provide a picture as possible of the event. The measurement of the missing transverse energy vector ($\vec{E}_T$) is not an easy task due to pile-up mitigation and is dominated by calorimeter resolution in the CMS detector [26].

The raw $\vec{E}_T$ is defined as a vectorial sum of the transverse momenta of all particles [69]

$$\vec{p}_{T,PF}^{\,\text{miss}}(\text{raw}) = -\sum_{i=1}^{N\text{particles}} \vec{p}_{T,i}, \tag{3.4}$$

where i runs over each PF particle in the event. The measurement of the $\vec{E}_T$ does not contain only the particles without signal detection; it also includes the mismeasurement of the jet's momentum. Thus, a correction is applied to this possibility by propagating the changes due to jet calibration of $\vec{E}_T$ [27]. The jet-energy-corrected $\vec{E}_T$ is written as follows

$$\vec{p}_{T,PF}^{\,\text{miss}} = \sum_{i=1}^{N\text{particles}} \vec{p}_{T,i} - \sum_{j=1}^{N\text{PF jets}} \left(\vec{p}_{T,j}^{\,\text{corr}} - \vec{p}_{T,j}\right). \tag{3.5}$$

The E_T^{miss} used in this analysis is computed by taking the negative vector sum of the transverse momentum of all particle flow candidates reconstructed of a given event. Corrections to the momenta for the used jets (See following sections for event selection) are used (Type-1 corrections).

3.6 Hadronic Tau decays

The tau (τ) particle is unstable, and its decay produces either a charged lepton and two neutrinos or hadrons and one neutrino. To see the τ decay modes and branching ratio fractions, see Ref

[104]. The τ decays hadronically 65% of the time, producing a τ jet, and in 77% of hadronic τ decays, the jet consist of only one charged hadron and several neutral pions (π^0). To distinguish the hadronic τ decay (τ_h) from quark and gluons jets, the decay products' multiplicity, collimation, and isolation are used [69].

Identifying τ jets requires matching the jet axis determined by the calorimeter and the jet axis determined by the charged particle tracking [26]. Hadronic τ decays produce localized energy deposits in the electromagnetic calorimeter, using different variables to quantify it and reject QCD jets, using the ECAL isolation criteria that use the energy deposits in the electromagnetic calorimeter and the distance parameter. Another tool for τ jet identification is tracker isolation (see Fig 3.3), where the calorimeter jet's axis defines the jet's direction. These two isolation requirements are used in the HLT for jet τ identification.

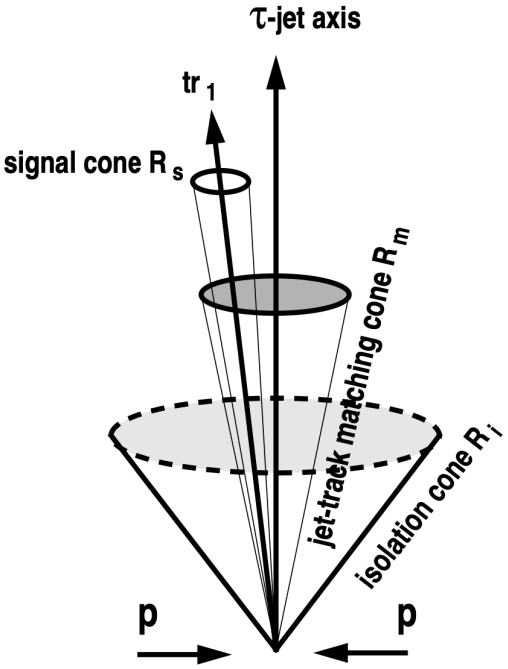

Figure 3.3: Sketch of τ tracker isolation [26].

The τ_h reconstruction began with the Calo τ_h jet candidates [51] that were reconstructed as collimated and isolated calorimetric jets. Nowadays, the PF algorithm is used to identify the particle from the τ decay, which also reconstructs the surrounding particles to determine their isolation for τ identification. The particles are used as input to the hadrons-plus-strips (HPS) algorithm [46] for the reconstruction and identification of τ candidates.

The HPS and Calo τ_h algorithms are compared in terms of their performance for identification efficiency, jet misidentification rate, and momentum reconstruction. The τ_h identification efficiency is defined as the probability of reconstructing and identifying a τ_h matching a generated τ_h within ΔR=0.3 [69]. For this, the τ_h are required to have $p_T > 20$GeV and $|\eta| < 2.3$. The jet misidentification rate is defined using the same selection as the probability to reconstruct and identify a quark or gluon jet from the multijet sample as a τ_h. The τ tagging efficiency can be evaluated with Monte Carlo simulation, assuming that some events are known and considering systematic uncertainties by the event selection related to the calorimeter information [26].

3.7 b jets

Many physics channels produce b jets in the final state, such as the top quark that decays into a W boson and a b quark, visible as a jet. The tagging of these b jets relies upon their distinctive properties of b hadrons, like their large lifetime, large mass, decays to final states with high charged track multiplicities, large semileptonic branching ratios, and hard fragmentation function [26]. The tagging algorithms rely upon the reconstruction of physics objects. Most of the b hadron properties used for b tagging use charged particle tracks because they can offer a spatial resolution to detect, for example, the decay length.

An example of common track selection cuts for b tagging algorithms is described as follows:

- Al least eight reconstructed hits in total
- At least two reconstructed hits in the pixel detectors.
- Transverse momentum $p_T > 1$ GeV/c.
- χ^2/ndf of the track fit <10.
- Transverse impact parameter concerning the reconstructed primary vertex < 2mm.

Nowadays, the CMS b-tagging algorithms relay into three main variables: the impact parameter significance of the tracks, the position of the secondary vertex, and the transverse momentum of the muon relative to the jet direction [79] (due to the semileptonic decay). This semileptonic decay of b hadrons is explained by the b jets that contain a muon (11% branching ratio) or when the b quark decay into a c quark that decays into a lepton (20% branching ratio).

The impact parameter (IP) is defined as the distance between the track and the primary vertex (PV) at the point of closest approach, as shown in Fig 3.4

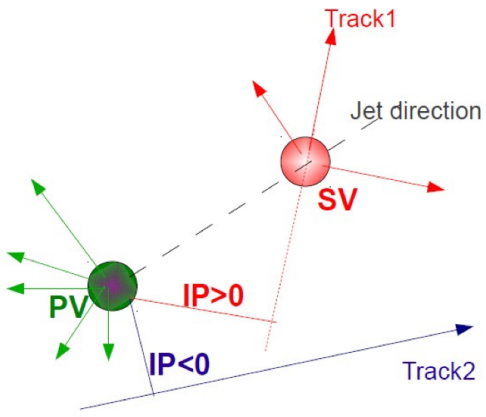

Figure 3.4: Geometric meaning of IP[79].

Some of the algorithms used for b-tagging are the Jet probability algorithm, the Soft-Lepton tagging algorithm, and the secondary vertex algorithm. An example is the combined b-tagging algorithm based on the secondary vertex of a b-hadron reconstruction. To distinguish b jets, topological and kinematic vertex variables are combined with track impact parameters' significances [26]. Secondary vertices are reconstructed using the Trimmed Kalman Vertex Finder algorithm [67] (the conventional robust version of the Kalman vertex fitter), which searches for vertex candidates among a set of tracks in an iterative way.

Depending on the cuts for discrimination, the efficiency of the taggers is different. The efficiency ϵ_q to tag a certain flavor of the jet as b jet [26] is defined as

$$\epsilon_q = \frac{\text{Number of jets of flavour q tagged as b}}{\text{Number of jets of flavour q}} \tag{3.6}$$

Some methods to measure the b-tagging efficiency include using a jet sample with enriched b-jet content, for example, a $t\bar{t}$ sample, in data or Monte Carlo.

For b tagging efficiencies measured in data, which is different than those predicted by the simulation, scale factors are applied to simulated events that depend on the jet flavor (b, c, light jets), jet pT, and jet η [78]. There are established standard operating points as loose (L), medium (M), and tight (T), being the value at which the tagging of udsg jets is estimated from MC to be 10%, 1%, or 0.1%, respectively, for jet transverse momentum of about 80 GeV [79].

Chapter 4

Dark Higgs analysis

This chapter presents an analysis searching for evidence of dark matter at the LHC. DM has not yet been observed in particle physics experiments; however, if the interaction exists, the DM particles could be produced at the LHC. Due to the lack of detectable signals at the LHC, one possible way to detect them is in association with a visible SM particle. These processes are observed at colliders as particles or jets recoiling against an invisible state from the dark matter and are called "mono-X" or MET+X reactions [30, 34] where the missing transverse momentum is the visible observable in the detector as DM signal. Thus, at the ATLAS-CMS Dark Matter Forum [11], the CMS and ATLAS groups of experimenters discussed and proposed a prioritized, compact set of benchmark simplified models for Run-2 searches. This forum proposed simplified models for Dirac fermion DM comprising different spin mediators, including an additional U(1) gauge symmetry. However, new mediators (spin 0 or 1) can produce observable signals in a wide range of DM searches for missing transverse momentum at the LHC. Furthermore, their presence can be responsible for establishing thermal equilibrium between the visible and the dark sector in the early Universe, including being related to the processes for DM relic abundance via thermal freeze-out [41, 88].

Typically, the most sensitive analysis channel to DM production is the mono-jet channel [89], which has strongly constrained the parameter space in which DM particles can obtain their relic abundance from direct annihilation into SM final states. The tension is relaxed if DM particles are not in the lightest state in the dark sector, and if the DM mass is generated via a Higgs mechanism in the dark sector and the resulting dark Higgs boson is lighter than DM, a new annihilation channel where DM particles annihilate into a pair of dark Higgs bosons (h_s), with subsequent decay into SM states, would be possible. If the dark sector includes an additional spin-1 Z' mediator, the probability of the Z' producing and radiating a dark Higgs (dark-Higgsstrahlung) can be large. Also, assuming a small mixing with the SM Higgs boson, the dark Higgs boson decays predominantly into a pair of b quarks, leading to a distinctive signature in the detector of large missing transverse momentum, arising from the decay of the Z' mediator into DM, and a double-b-tagged, highly-boosted large-radius jet coming from the dark Higgs decay. The following sections present an analysis searching for evidence of DM in the mono-h_s final state, focusing on the b quark channel. Results are presented for the 137.2 fb^{-1} dataset collected by CMS at a center-of-mass energy of 13 TeV during Run 2 (2016 to 2018).

4.1 Dark Higgs model

The dark Higgs model [63] proposes that dark matter particles, which are Majorana particles, are not in the lightest state, leading to new annihilation channels and the mechanism to acquire mass, including the existence of a Higgs boson in the dark sector. This sector [33] suggests the existence of low-mass particles with weak interactions with visible matter.

The model assumes the DM particle is a Majorana particle χ that obtains its mass from the vacuum expectation value (vev) w of a new complex Higgs S field, a singlet under the SM gauge group [64, 91]. The scalar field S carries a charge q_S under a novel U(1)$'$ gauge group. Thus w breaks the gauge symmetry spontaneously, generating the mass of a corresponding Z' boson, which is the mediator between DM particles and SM states, and also arising into a new Higgs boson defined by

$$S = \frac{1}{\sqrt{2}(s+w)}, \tag{4.1}$$

that is referred to as the dark Higgs boson (h_s). Within this framework, the DM particle obtains an axial coupling to the Z' boson, resulting in a mutual coupling between all three particles within the dark sector. The expression gives the renormalizable terms in the Lagrangian

$$\mathcal{L}_\chi = -\frac{1}{2} g' q_\chi Z'^{\mu} \overline{\chi} \gamma^5 \gamma_\mu \chi - \frac{y_\chi}{2\sqrt{2}} s \overline{\chi} \chi + \frac{1}{2} g'^2 q_S^2 Z'^{\mu} Z'_\mu (s^2 + 2sw), \tag{4.2}$$

where g' refers to the $U(1)'$ gauge coupling, and gauge invariance requires that the charge of the DM particle satisfy $q_\chi = q_S/2$. The masses of the DM particles and the Z' boson are parameters given by $m_\chi = y_\chi w/\sqrt{2}$ and $m_{Z'} = 2g' q_\chi w$. The m_s mass is also a parameter. Therefore, the model parameters are $m_\chi, m_{Z'}, m_s$ and $g_\chi \equiv g' q_\chi$, and using these parameters, we can write the equation 4.2 as follows:

$$\mathcal{L}_\chi = -\frac{1}{2} g_\chi Z'^{\mu} \overline{\chi} \gamma^5 \gamma_\mu \chi - g_\chi \frac{m\chi}{m_{Z'}} s \overline{\chi} \chi + 2 g_\chi Z'^{\mu} Z'_\mu (g_\chi s^2 + m_{Z'} s). \tag{4.3}$$

We used the same coupling choice to compare with existing LHC searches: $g_q = 0.25$ and $g_\chi = 1$ [11].

In this scenario, the h_s boson is the lightest state in the dark sector, and DM particles can annihilate into other dark sector states, which subsequently decay into SM states. The non-zero mixing between the h_s boson and the SM Higgs boson [91, 80] ensures that the h_s boson is unstable even if it is the lightest state in the dark sector and decays into SM states, making the relic density primarily set by the $\chi\chi \to ss$ process.

For this analysis, only the vector couplings of the Z' boson to SM quarks will be considered:

$$\mathcal{L}_\chi = -g_q Z'^{\mu} \bar{q} \gamma_\mu q \tag{4.4}$$

Axial-vector coupling of the Z' boson to SM quarks will not be considered even though, in principle, they would also be present. Still, they would require significant modifications to the current model setup to guarantee SM gauge invariance [91]. Additional couplings of the Z' boson to SM leptons could also be possible. However, such couplings would lead to additional constraints [66]. Moreover, it will not be considered.

.

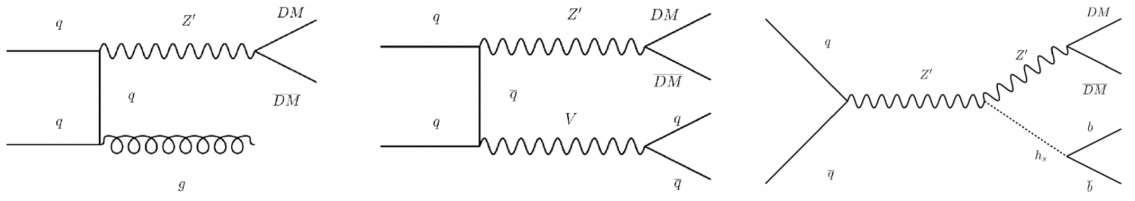

Figure 4.1: Process leasing to missing transverse momentum signatures. Left: a typical mono-jet process. Center and right: Monoh_s process.

In contrast to the conventional signatures of DM at the LHC, which involve missing transverse momentum due to the recoil of DM particles against an SM state from initial state radiation (as

shown in Figure 4.1 left), the scenario for the dark Higgs model involves the recoil of DM particles against a visibly decaying dark Higgs boson emitted in the final state radiation (Fig 4.1 right). For this analysis, the considered process is shown on the right side, where the dark Higgs boson decays into a pair of b quarks.

The difference with the dark Higgs model is that DM particles recoil against a visibly decaying dark Higgs boson from final state radiation. Compared with other searches like mono-jet [106] , the dark Higgs boson will decay into a pair of b quarks, leading to a characteristic highly boosted large-radius jet.

For the analysis, the selection of the bb channel over other processes is motivated by several key factors. First, in the presence of a light-dark Higgs boson, the decay mode $Z' \to \chi\chi S$ becomes available, offering a distinct three-body decay signature with relatively soft dark Higgs bosons and approximately back-to-back DM particles. This results in little missing transverse momentum in the decay, making it a unique and distinguishable signature. Furthermore, the three-body decay mode can have a substantial branching fraction, especially for specific couplings, allowing for efficient detection. Compared to other processes, such as initial state radiation, the bb channel proves advantageous as it provides high transverse momentum jets associated with large missing transverse momentum after the decay of the dark Higgs boson. This signature stands out in LHC searches and offers great sensitivity to uncover new physics. Additionally, the decay of the dark Higgs boson predominantly into heavy quarks produces a distinct fat jet containing two b-jets, allowing for effective discrimination from background processes and enhancing the sensitivity of the bb channel compared to common mono-X searches. Overall, the selection of the bb channel enables comprehensive exploration of the dark Higgs model and its implications for dark matter searches, offering unique signatures and enhanced sensitivity to uncover new physics beyond the Standard Model.

4.2 Event selection

Event selection is the foundation of any particle physics analysis and is crucial to our search for the Dark Higgs boson. This section will discuss the principles, criteria, and strategies we use to curate a sample of events that exhibit the characteristics of Dark Higgs interactions and are robust against various background processes. From defining specific object properties to delineating regions of interest, our event selection criteria are meticulously crafted to unveil the elusive signals hidden within the vast sea of collider data.

4.2.1 Object definitions

Accurate identification and characterization of particles and objects produced in high-energy proton-proton collisions are prerequisites for any particle physics analysis. In this subsection, we discuss the intricate task of object definitions within the CMS detector, which is the cornerstone of our event selection process. The successful reconstruction of muons, jets, and other fundamental particles is paramount in discerning the signatures of the Dark Higgs. Here, we elucidate the criteria and algorithms to define these objects, ensuring they adhere to stringent quality standards. The precision of our object definitions not only facilitates the identification of genuine particles but also shields against spurious signals, ensuring the fidelity of our analysis.

Missing Transverse Momentum and Hadronic Recoil

The E_T^{miss} used in the analysis is computed by taking the negative vector sum of the transverse momenta of all particle flow candidates reconstructed in a given event. Corrections to the momenta of AK4 PF CHS jets reconstructed in the event are further propagated to the E_T^{miss} (Type-1 T corrections). To mimic the E_T^{miss} in the signal regions, control regions are constructed by selecting events with a lepton corresponding to regions dominated by W boson production. The hadronic

recoil against the bosons can be computed by removing the $\vec{p}_T$ of the lepton from the E_T^{miss} computation. Formally, the recoil is given by

$$\vec{U} = \vec{p}_T^{\text{miss}} + \vec{p}_T^{\,l}, \tag{4.5}$$

where the last term is the $\vec{p}_T$ of the muon or electron, depending on the control region. Events in data can end up with large spurious E_T^{miss} due to detector noise and beam backgrounds. To remove these events, several E_T^{miss} filters, such as those listed below, have been recommended by the JetMET POG [2] and have been implemented in the analysis

- Good Vertex Filter (At least one good primary vertex in the event)
- Global Beam Halo SuperTight Filter
- HBHE Noise Filter
- HBHE NoiseIso Filter
- ECAL Dead Cells Trigger Primitive Filter
- Bad PF Muon Filter
- EE Bad SC Filter
- ECAL Bad Calibration Filter

AK4 Jets

We used the reconstructed jets clustering particle flow candidates in the event using the anti-k_T algorithm [38] with a distance parameter of 0.4 (AK4 jets). To mitigate the effects of pileup, we employ two techniques: Charged Hadron Subtraction (CHS) and L1FastJet corrections. CHS eliminates charged hadrons not originating from the primary vertex, where the proton-proton collision occurs. These extra tracks arise from additional proton-proton interactions within the same bunch crossing, known as pileup. L1FastJet corrections, on the other hand, address the smearing effect of pileup on jet energy measurements. Using the event-by-event energy density, L1FastJet corrections estimate and subtract the pileup energy contribution from each jet, ensuring more accurate jet measurements. We apply further corrections for jet pseudorapidity (η) and transverse momentum (L2, L3 corrections). We apply further residual jet energy corrections to jets in data to match their response with the simulation (L2L3Residual). These corrections have been derived on AK4 CHS jets [3]. We consider narrow jets (AK4 jets) in the analysis that meet the following requirements:

- $p_T > 30$ GeV, $|\eta| < 2.4$
- Pass standard tight jet identification criteria [109]
- Neutral hadron fraction < 0.8
- Neutral EM fraction < 0.99
- Charged hadron fraction > 0.1
- Pass the Loose PU ID if $p_T < 50$ GeV

The explicit selections on the energy fractions discussed above may be overridden by selections in the "tight jet identification" criteria. In regions where a lepton is required, we reject a jet if ΔR(jet, l) < 0.4. We also require jets to be well-separated from "fat jets" (see below), so we reject a jet if ΔR(jet, fat jet) < 1.5. In this way, all AK4 jets in the analysis are "extra jets" concerning the fat jet that will make up the Dark Higgs candidate in the event.

AK15 Fat Jets

To identify the dark Higgs candidate, we cluster jets from PUPPI-weighted PF candidates using the anti-k_T algorithm with a distance parameter of 1.5 (AK15 jets), also known as "fat jets." The fat jets must be "dark Higgs tagged" to be considered good candidates. The two main criteria for dark Higgs tagging are the jet mass and the B-hadron content of the jet.

We use the soft drop method to groom the fat jets to reduce the impact of pileup interactions, underlying events, and parton shower activity. We chose the soft drop parameters $Z_{\text{cut}} = 0.15$ and $\beta = 1$ to optimize the mass resolution after grooming.

We use deep-learning-based techniques to identify B-hadrons within the fat jets (Deep-Flavour). This technique is independent of the jet mass variable, which is important because we use the kinematics of the groomed jet throughout the analysis.

Fat jets considered in the analysis must have $p_T > 160$ GeV and $|\eta| < 2.4$. To remove jets due to HCAL noise or mismeasurement, we require that fat jets pass the tight jet ID requirements described in [16]. In regions where a lepton is required, we reject a fat jet if $\Delta R(\text{jet}, l) < 1.5$. Finally, we correct the jets using the L1, L2, L3, and L2L3Residuals corrections derived for AK8 jets.

Muons

We reject events containing muons in the analysis to suppress electroweak backgrounds. To be considered a "loose" muon, an object must pass the following selection:

- $p_T > 20$ GeV (2016 and 2017), 15 GeV (2018)
- $|\eta| < 2.4$
- "Loose Muon" identification criteria [112]
- $|\text{dxy}| < 0.5$ cm
- $|\text{dz}| < 1.0$ cm
- Particle-flow relative isolation in a cone of $\Delta\text{R} = 0.4$, $\text{I}_{\text{rel}} < 0.4$ (the "very loose" working point, with an efficiency above 98%)

We use selections that rely on muons to develop control regions for irreducible backgrounds. To be considered a "tight" muon, an object must pass the following selection:

- $p_T > 30$ GeV
- $|\eta| < 2.4$
- "Tight Muon" identification criteria [113] (which include $|\text{dxy}| < 0.2$ cm and $|\text{dz}| < 0.5$ cm selections, among others)
- Particle-flow relative isolation in a cone of $\Delta\text{R} = 0.4$, $\text{I}_{\text{rel}} < 0.15$ (the "tight" working point, with an efficiency of 95%)

We apply an event scale factor to MC to account for differences in muon reconstruction and identification between data and simulation. The derivation of this scale factor and the corresponding systematic is described in [16].

Electrons

As with muons, we use two definitions of electrons for veto and selection. Loose electrons must pass the following selections:

- $p_T > 10$ GeV
- $|\eta_{SC}| < 2.5$
- Loose identification requirements [107] (Electron_mvaFall17V2Iso_WP90 working point, an MVA-based identification with isolation inputs and an efficiency of 90%)
- Separated from loose muons (ΔR > 0.3)

Tight electrons must pass the following selections:

- $p_T > 29$ GeV (2016), 40 GeV (2017 and 2018)
- $|\eta_{SC}| < 2.5$
- Tight identification requirements [107] (Electron_mvaFall17V2Iso_WP80 working point, an MVA-based identification with isolation inputs and an efficiency of 80%)
- Separated from loose muons (ΔR > 0.3)

Loose leptons for selection (for the Dilepton control regions) have the exact requirements as the veto definition for loose electrons given above. An event scale factor is applied to account for differences in electron reconstruction and identification between data and MC, as described in [16].

Photons

We reject events with any photons passing the loose identification, defined as follows:

- $p_T > 15$ GeV
- $|\eta| < 2.5$
- Loose identification requirements [108] (94X-V2 ID at the loose working point, a cut-based identification with an average efficiency of 90%)
- A veto on electrons (i.e., an identified photon must not also be an electron)

Taus

We apply a loose selection to reject taus:

- $p_T > 18$ GeV
- $|\eta| < 2$
- Old DecayModeFinding (since we are neither using DeepTau v2 nor particularly high-p_T taus, there is no particular reason to use the new one.)
- MVA-based very loose isolation provided by the POG [114] (specifically, the byVLooseIsolationMVArun2017v2DBoldDMwLT2017 working point)
- Separated from loose muons (ΔR > 0.4) and loose electrons (ΔR > 0.4)

4.2.2 Datasets

This analysis uses certified events from the following data-taking periods:

- Run2016B–H datasets, 17 Jul 2018 reconstruction, corresponding to 35.9 fb^{-1}.
- Run2017B–F datasets, 31 Mar 2018 reconstruction, corresponding to 41.5 fb^{-1}.
- Run 2018A–C datasets, 17Sep2018 reconstruction and Run 2018D dataset, Prompt reconstruction, all corresponding to 59.74 $^{-1}$.

4.2.3 Trigger selection

Events with W bosons decaying into an electron and a neutrino are selected with unprescaled single electron triggers. These triggers have different requirements for electron identification and isolation, depending on their p_T threshold. Low-p_T electron triggers are inefficient for high-p_T electrons, so high-p_T electron triggers with different isolation requirements are also used.

In addition, a bug in the ECAL reconstruction during the 2017 run caused the wrong ADC gain switch to be used [75]. To address this bug, the EGamma POG requires a logical OR with the Photon200 path for analyses that use electrons with $p_T > 300$ GeV and 2017 data. This requirement is also used for 2018 data as a safety measure.

Table 4.1 shows the full list of triggers used for each year of data taking, along with the L1 seeds and the associated primary datasets.

The different reconstruction eras aim to achieve a uniform performance for the Run 2 dataset. All those datasets correspond to a total integrated luminosity of 137.2 fb^{-1}. The analysis uses control regions defined by muons and electrons, so the MET and electron/photon datasets are used. The MET dataset includes muon pass-through triggers, which allow us to select events with high E_T^{miss} or high muon recoil.

4.2.4 Event topology

The CMS detector at the CERN LHC records various particle interactions, each with its unique signature. Understanding and characterizing these event topologies is essential for our search for the Dark Higgs. This section will explore the distinct features of Dark Higgs events and how they differ from background processes.

The Feynman diagram for the signal is shown in Figure 4.1. The event topology in the dark Higgs model consists of missing transverse energy and a fat jet that contains two b quarks. Having such objects in the final state, the main corresponding background is Z+jets, W+jets, and $t\bar{t}$ production. The expected preselection for the final state is shown in Table 4.2. The preselection aims to build a well-defined set of events upon which the signal and control regions will be constructed.

4.2.5 Signal region and control regions

Distinguishing between signal and background events is essential for the Dark Higgs analysis. This subsection discusses the crucial concepts of signal and control regions. The signal region is designed to capture the characteristics of Dark Higgs interactions. In contrast, the control regions are used to validate our background estimation techniques and ensure the reliability of our analysis. By carefully defining these regions and their specific criteria, we create a robust framework to draw meaningful conclusions about the presence or absence of Dark Higgs signals while carefully monitoring the impact of background processes.

All regions (signal and control) share a common preselection, described in Table 4.2. This preselection ensures that the event topology is close to the one expected for the dark Higgs signal. To distinguish between the signal and control regions, we use the lepton and photon multiplicities

Year	HLT path	L1 seed	Primary data set
2016	HLT_PFMETNoMu90_PFMHTNoMu90_IDTight HLT_PFMETNoMu100_PFMHTNoMu100_IDTight HLT_PFMETNoMu110_PFMHTNoMu110_IDTight HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	L1_ETMHF70 L1_ETMHF70 L1_ETMHF70 L1_ETMHF70	MET
	HLT_Ele27_WPTight_Gsf	L1_SingleEG24	SingleElectron
	HLT_Ele105_CaloIdVT_GsfTrkIdT	L1_SingleEG30 L1_SingleJet170 L1_SingleTau100er2p1	
2017	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60	LL1_ETMHF70 L1_ETMHF80_HTT60er	MET
	HLT_Ele35_WPTight_Gsf HLT_Ele115_CaloIdVT_GsfTrkIdT	L1_SingleEG24 L1_SingleEG30 L1_SingleJet170 L1_SingleTau100er2p1	SingleElectron
	HLT_Photon200	L1 SingleEG30 L1 SingleJet170 L1_SingleTau100er2p1	SinglePhoton
2018	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	L1_ETMHF100 L1_ETM150	MET
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60	L1_ETMHF90_HTT60er	
	HLT_Ele32_WPTight_Gsf	L1_SingleIsoEG24er2p1 L1_SingleEG26er2p5 L1_SingleEG60	EGamma
	HLT_Ele115_CaloIdVT_GsfTrkIdT	L1_SingleEG34er2p5 L1_SingleJet160er2p5 L1_SingleJet180 L1_SingleTau120er2p1 L1_SingleEG60	
	HLT_Photon200	L1_SingleEG34er2p5 L1_SingleJet160er2p5 L1_SingleJet180 L1_SingleTau120er2p1 L1_SingleEG60	

Table 4.1: HLT paths and the associated L1 seeds used in the analysis for the different years of data taking.

and the number of b-tagged AK4 jets that do not overlap with the fat jet. This minimizes the extrapolation uncertainties.

The analysis considers the smallest possible number of control regions to constrain the shapes and yields of the three main standard model backgrounds to the search: W and Z production with jets and $t\bar{t}$ production. We define these regions based on two criteria: whether the main candidate jet in the event passes the DeepAK15 selection and the number of well-reconstructed leptons (electrons or muons) in the event. Signal events are expected to populate the {pass, 0l{ category, while background events are expected to populate the other categories ({pass, 1l}, {fail, 0l}, and {fail, 1l}). We use the distribution of events in these categories to estimate the remaining standard model background in the signal-populated category (See fig 4.2).

To constrain the background shapes and yields, it is defined the following control regions on top of the event preselection:

- Zero lepton: Jet HEM mitigation, no loose leptons (muons or electrons), no loose photons,

AK15 jet p_T threshold	> 160 GeV
E_{miss} threshold (signal region)	> 250 GeV
Recoil $\lvert\mathbf{U}\rvert$ threshold (control region)	> 250 GeV
min $\Delta\phi(\mathbf{U}$, AK15)	> 1.5
Loose tau leptons	Vetoed

Table 4.2: Preselection summary.

selected by Emiss triggers. In this region, the 2018 period also requires the additional E_T^{miss} HEM mitigation described above.

- Single muon: Presence of one tight muon, no loose electrons, no loose photons, no extra b-tagged jets, jet HEM mitigation, selected by E_T^{miss} triggers.
- Single electron: Presence of one tight electron, no loose muons, no loose photons, no extra b-tagged jets, jet HEM mitigation, selected by single electron triggers. In this region, it also requires $E_T^{\text{miss}} > 100$ GeV and $\Delta\phi(E_T^{\text{miss}}, \text{jet}) > 0.7$ to improve E_T^{miss} modeling.
- Top muon: Presence of one tight muon, no loose electrons, no loose photons, at least one extra b-tagged jet, jet HEM mitigation, selected by E_T^{miss} triggers.
- Top electron: Presence of one tight electron, no loose muons, no loose photons, at least one extra b-tagged jet, jet HEM mitigation, selected by single electron triggers. In this region, it also requires $E_T^{\text{miss}} > 100$ GeV and $\Delta\phi(E_T^{\text{miss}}, \text{jet}) > 0.7$ to improve E_T^{miss} modelling.

All the regions are divided into two subregions based on whether or not the fat jet passes the ZHbbvsQCD b-tag threshold of 0.65. The ZHbbvsQCD tagger is a deep learning-based tagger used to identify jets likely to originate from the decay of the Higgs boson into two b quarks. By selecting events with a high ZHbbvsQCD tagger score, the background can be reduced, and the search statistics can be increased.

The pass subset of the zero lepton region is the signal region, where the search for evidence of dark Higgs events is performed. The challenge of the analysis is to predict the distributions of the standard model backgrounds in the signal region.

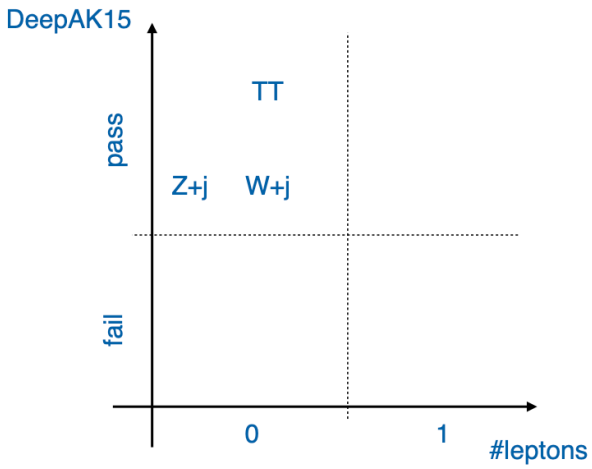

Figure 4.2: Overall picture of the regions

4.3 Signal and background estimation

To separate the nature of the Dark Higgs, it is important to distinguish the signal from the myriad of background processes that pervade high-energy particle collisions. These background sources, arising from well-understood and often abundant physics phenomena, can mimic or obscure the elusive signatures of exotic particles. To rigorously ascertain the existence of the Dark Higgs, we must carefully estimate and mitigate these background contributions.

This section is the linchpin of the analysis, shedding light on the techniques and strategies that enable us to disentangle genuine Dark Higgs events from the background clutter. Here, we comprehensively explore the diverse approaches employed to quantify the expected backgrounds, ranging from data-driven techniques to sophisticated Monte Carlo simulations.

4.3.1 Monte Carlo Signal Generation

The MC samples for the mono-h_s process shown in Figure 4.1 left are generated with the MadGraph5 aMC@NLO v2.6.0 [19] using NNPDF v3.0 [50] for 2016 production and NNPDF v3.1[50] for 2017/2018 production. Pythia v8.2.4.0 [120] with the CUEP8M1 and CP5 tunes were used for hadronization and underlying events generation for 2016 and 2017/2018 productions. The samples are generated at leading order (LO). The h_s boson mass hypotheses for this search are 50, 70, 90, 110, 130, and 150 GeV. The mass grid in Table 4.3 shows the values of the mass of the DM particle and the mass of the Z' bosons scanned in the signal generation.

It is important to note that the hs and the SM Higgs boson have identical branching ratios for a given mass value due to their mixing. The decay of hs into $b\bar{b}$ is the most prevalent for $m_{hs} < 135$ GeV, but it remains significant for masses between 135 GeV and 160 GeV. For this investigation, we have considered the masses of the hs boson to be 50, 70, 90, 110, 130, and 150 GeV.

Z' mass [GeV]	DM mass [GeV]
195	100
200	150
295	150
300	100
495	250
500	150, 500
995	500
1000	150, 1000
1995	1000
2000	500, 1500
2495	1250
2500	750
2995	1500
3000	1000

Table 4.3: Z' and DM Masses generation

A distance parameter k_T between partons in the $\eta - \phi$ space is used for the MLM jet matching procedure [85], which helps to improve the accuracy and reliability of predictions for multi-jet processes. A requirement on k_T to be larger than a value called "xqcut" is imposed. The xqcut value is the required parton separation at the parton level. Events that satisfy such requirements are passed to Pythia for parton showering. After showering, but before hadronization and decay, the final-state partons are clustered into jets. These jets are then compared to the original partons from the matrix element event. A jet is considered to be matched to a parton if the k_T between the parton and the jet is smaller than a value called "qcut." The event is rejected unless each jet

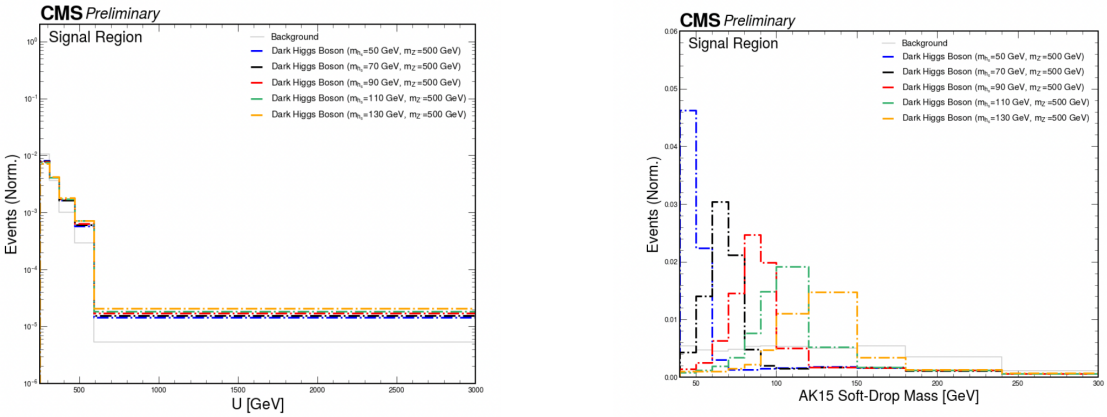

Figure 4.3: The unstack U (left) and leading AK15 m_{SD} (right) distributions.

is matched to a parton. For the dark Higgs model generation, the matching parameters xqcut and qcut have been set to 15 GeV and 20 GeV, respectively.

4.3.2 Background estimation

In general lines, the background to this search is those from standard model processes that have real E_T^{miss} in the final state:

- Z$\to \nu\nu$+jets.
- W$\to l\nu$+jets.
- semileptonic and fully leptonic t$\bar{t}$.
- semileptonic and fully leptonic diboson WW, WZ, and ZZ.
- multijet production, a.k.a "QCD".
- γ + jets.

The last two backgrounds come primarily from mismeasurements of the reconstructed particle momenta in the event. Finally, DY+jets samples with electrons or muons in the final state are used for control region analysis. Some examples of these processes are presented in Figures 4.4, 4.5, and 4.6.

In Fig. 4.3 the unstack leading AK15 m_{SD} and U distributions are shown for different dark Higgs boson signal processes, compared to the one expected for the sum of all backgrounds. In these plots, the 2018 MC simulation is used. Similar distributions are observed using the 2016 and 2017 simulations.

A data-driven method estimates the main background (W and Z + jets, and $t\bar{t}$). Meanwhile, the rest of the background presented in Tables 4.4 and 4.5 are from Monte Carlo. The data-driven method estimates the background directly from the experimental data without relying on theoretical calculations or simulations. This is done by defining control regions in the data where the signal contribution is expected to be negligible, but the background is dominant. The signal can be estimated from the background contribution by comparing the properties of these control regions to the signal region (where both signal and background are present).

A pass-fail transfer factor is used to predict the distributions of the W, Z+jets, and tt backgrounds in the signal region (i.e., the pass subset of the zero-lepton region). Transfer factors are

Dataset name	σ [pb]
/QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1735000
/QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	366800
/QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	29370
/QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6524
/QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1064
/QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	121.5
/QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	25.42
/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	$136.02^{\dagger}$
/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	$80.95^{\dagger}$
/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	$35.85^{\dagger}$
/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	$35.85^{\dagger}$
/ZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8	0.08912
/ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8	0.04865
/ggZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8	0.014366
/ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8	0.007842
/WminusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8	0.100
/WplusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8	0.159
/ttHTobb_M125_13TeV_powheg_pythia8	$0.506 * 0.5824$
/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	$831.76^{\dagger}$
/WW_TuneCUETP8M1_13TeV-pythia8	$118.7^{\dagger}$
/WZ_TuneCUETP8M1_13TeV-pythia8	$47.2^{\dagger}$
/ZZ_TuneCUETP8M1_13TeV-pythia8	$16.6^{\dagger}$
/DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	148.0
/DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	40.94
/DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	5.497
/DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	1.367
/DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	0.6304
/DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	0.1514
/DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	0.003565
/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	1343
/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	359.6
/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	48.85
/WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	12.05
/WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	5.501
/WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	1.329
/WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8*	0.03216
/ZJetsToNuNu_HT-100To200_13TeV-madgraph*	280.5
/ZJetsToNuNu_HT-200To400_13TeV-madgraph*	77.7
/ZJetsToNuNu_HT-400To600_13TeV-madgraph*	10.71
/ZJetsToNuNu_HT-600To800_13TeV-madgraph*	2.559
/ZJetsToNuNu_HT-800To1200_13TeV-madgraph*	1.183
/ZJetsToNuNu_HT-1200To2500_13TeV-madgraph*	0.292
/ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph*	0.0069

Table 4.4: List of background MC 2016 samples. Datasets marked with * are LO in QCD and EWK but will have NLO corrections applied.

Dataset name	σ [pb]
/ZJetsToNuNu_HT-100To200_13TeV-madgraph*	302.8
/ZJetsToNuNu_HT-200To400_13TeV-madgraph*	92.59
/ZJetsToNuNu_HT-400To600_13TeV-madgraph*	13.18
/ZJetsToNuNu_HT-600To800_13TeV-madgraph*	3.257
/ZJetsToNuNu_HT-800To1200_13TeV-madgraph*	1.49
/ZJetsToNuNu_HT-1200To2500_13TeV-madgraph*	0.3419
/ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph*	0.005146
/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8*	1319
/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8*	1395
/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8*	407.9
/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8*	57.48
/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8*	12.87
/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8*	5.366
/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8*	1.074
/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8*	0.008001
TTJets_TuneCP5_13TeV-madgraphMLM-pythia8	831.76†
ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8	80.95†
ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8	136.02†
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85†
ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85†
ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	3.36†
WW_TuneCP5_13TeV-pythia8	118.7†
WZ_TuneCP5_13TeV-pythia8	43.5†
ZZ_TuneCP5_13TeV-pythia8	19.1†
QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8	23700000
QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8	1547000
QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	322600
QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	29980
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	6334
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	1088
QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	99.11
QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	20.23
GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8	18700
GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	8640
GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2185
GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	260
GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	85.31
DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8*	143.0
DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8*	161.1
DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8*	48.66
DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8*	6.968
DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8*	1.743
DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8*	0.8052
DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8*	0.1933
DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8*	0.003468

Table 4.5: List of background MC samples, 2017/2018 versions. These samples were produced with the CP5 tune, so the cross-sections differ slightly. Datasets marked with * are LO in QCD and EWK but will have NLO corrections applied. Cross-sections marked with † are (N)NLO.

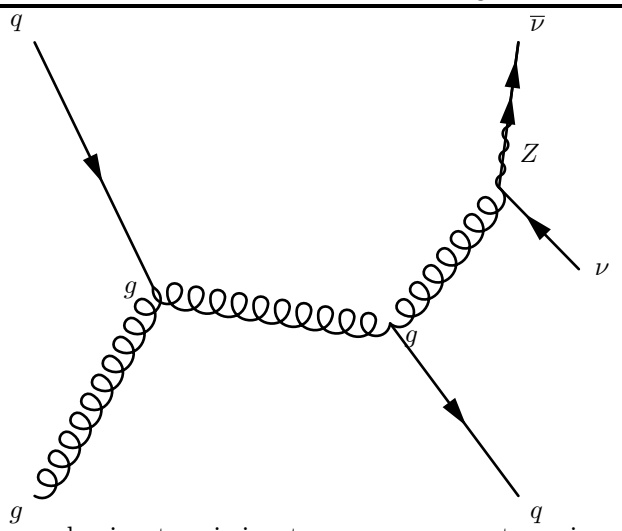

Figure 4.4: Example process leasing to missing transverse momentum signatures including a Z decay.

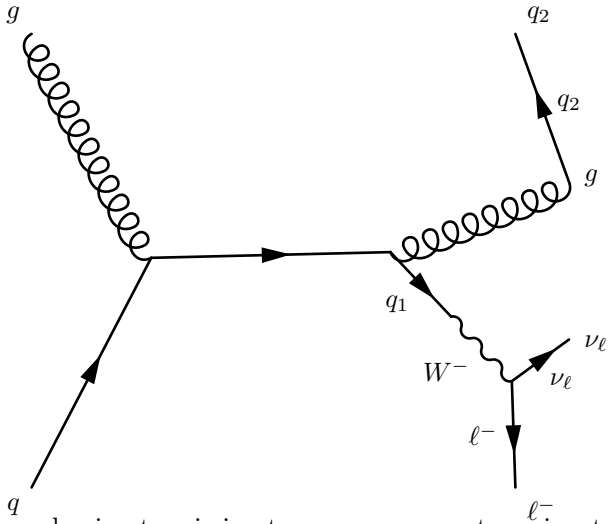

Figure 4.5: Example process leasing to missing transverse momentum signatures including a W^- decay.

calculated in the zero-lepton region using the data in the fail subset, which is dominated by W, Z+jets events. We use the single-lepton regions to constrain the transfer factors for the W+jets background and the top-lepton regions to constrain the transfer factors for the $t\bar{t}$ background. Finally, we use the implicit W-Z link from electroweak symmetry to constrain the transfer factors for the Z+jets background.

For the estimation, the rhalphabet method is used. The method is as follows. First, it defines a signal region and a control region. Then, essential variables for discriminating between the signal and the background are defined. After, it is used a data-driven method to estimate the transfer function from the control region to the signal region. The transfer function is a function that maps the distribution of the variables in the control region to the distribution of the variables in the signal region. Finally, a transfer function is used to estimate the background in the signal region.

To estimate the Z+jets background component in the signal region, we assume it is connected to the same component in the zero-lepton region via a transfer factor, denoted as $TF(m_{SD}, U)$. Since the difference between the signal region and the zero lepton region is the category of fat jet tagging (pass or fail), we will call this transfer factor the pass-fail ratio.

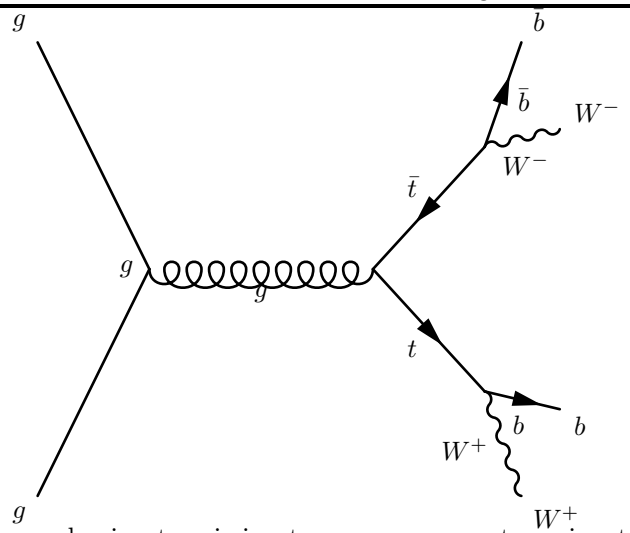

Figure 4.6: Example process leasing to missing transverse momentum signatures including a $t\bar{b}$ decay.

The pass-fail ratio is between two 2D binned distributions in m_{SD} and recoil U. Therefore, it is also a 2D binned function divided into the 5×11 bins discussed above. For the polynomial correction, we assume it is defined over the 2D bin barycentres and estimate it on the Bernstein basis. Therefore:

$$TF(m_{SD}, U) = \sum_{k=0}^{n_{m_{SD}}} \sum_{l=0}^{n_U} a_{k,l} b_{k,l}(m_{SD}) b_{l,n_U}(U), \tag{4.6}$$

where $a_{k,l}$ are the free parameters of the polynomial correction, and $b_{k,n}(m_{SD})$ and $b_{l,n_U}(U)$ are the Bernstein basis functions. The pass-fail ratio has to account for two effects: First, it covers any residual tagger-kinematics correlation, including dependence on the varying heavy-flavor fraction. Second, it covers the effects of data-simulation discrepancies in the zero-lepton region. Given this double duty, we factorize it into two parts: a simulation-based pass-fail ratio incorporating systematic effects and a polynomial with free parameters. We, therefore, write the final form of the pass-fail ratio as:

$$R^{Z+jets}(m_{SD}, U) = \sum_{l=0}^{n_U} a_{k,l} b_{k,n_{m_{SD}}}(m_{SD}) b_{l,n_U}(U) \times \epsilon^{Z+jets}(m_{SD}, U), \tag{4.7}$$

and the Z+jets distribution in the signal region as

$$N^{Z+jets}_{0l,pass}(m_{SD}, U) = N^{Z+jets}_{0l,fail}(m_{SD}, U) \times R^{Z+jets}(m_{SD}, U). \tag{4.8}$$

Using the same approach as for Z+jets, the W+jets background estimation is made. The distribution in the signal region is expressed as

$$N^{W+jets}_{0l,pass}(m_{SD}, U) = N^{W+jets}_{0l,fail}(m_{SD}, U) \times R^{W+jets}(m_{SD}, U). \tag{4.9}$$

However, there are additional constraints for this process. It is assumed that Z+jets and W+jets are related together through a W-Z link, and it is based on the fact that the production of Z bosons and W bosons is closely related:

$$N^{W+jets}_{0l,fail}(m_{SD}, U) = \frac{N^{Z+jets}_{0l,fail}(m_{SD}, U)}{E^{W-Z}(m_{SD}, U, \theta)}, \tag{4.10}$$

where the systematic uncertainties are incorporated via nuisance parameters θ. It is also possible to constrain the W+jets distribution to improve the sensitivity of the search by using the single-lepton region in pass and fail categories:

$$N^{W+jets}_{1l,pass/fail}(m_{SD}, U) = \frac{N^{W+jets}_{0l,pass/fail}(m_{SD}, U)}{T^{W+jets}_{pass/fail}(m_{SD}, U, \theta)} \tag{4.11}$$

We can summarize these transfer factors as follows:

- R: A transfer factor that connects the irreducible Z+jets background in the signal region to the Z+jets background in the zero-lepton fail control region. This is a pass-fail ratio related to the inefficiency of the DeepAK15 tagging.

- E: A transfer factor that connects the W+jets and Z+jets backgrounds in the zero-lepton fail control region. Considering the lepton from W+jets is already lost, the ratio of contributions of each process in this region is dominated by electroweak symmetry considerations. The effect of W/Z NLO correction uncertainties appears mostly through this link.

- T: A transfer factor that constrains the W+jets background in the signal region by comparing it to the W+jets background in the single-lepton regions, separately in the pass and fail categories. This is directly related to the lepton misreconstruction and misidentification probabilities.

Finally and such as W+jets and Z+jets, it is estimated the $t\bar{t}$ distribution, where it is used data in a dedicated $t\bar{t}$ enriched control region to predict it on the signal region:

$$N^{t\bar{t}}_{t\bar{t}_l CR_{pass}}(m_{SD}, U) = \frac{\mu^{t\bar{t}}_{SR_{pass}}(m_{SD}, U)}{TF^{t\bar{t}}_{t\bar{t}_l CR_{pass}}(m_{SD}, U, \theta)}. \tag{4.12}$$

The $t\bar{t}$ contamination in the W_μ and W_e control regions in the pass category is also used to constrain the $t\bar{t}$ background. This is done by subtracting the estimated $t\bar{t}$ contribution from the total number of events in these control regions:

$$N^{t\bar{t}}_{W_\mu CR_{pass}}(m_{SD}, U) = \frac{\mu^{t\bar{t}}_{SR_{pass}}(m_{SD}, U)}{TF^{t\bar{t}}_{W_\mu CR_{pass}}(m_{SD}, U, \theta)}. \tag{4.13}$$

These transfer factors and the control and signal regions for the background estimation can be visualized in Fig 4.7.

4.4 Systematic Uncertainties

Particle physics analyses require precise measurements such as the Dark Higgs. This requires a deep understanding of the signal and background processes and the systematic uncertainties that can affect the results. Systematic uncertainties are factors beyond the statistical fluctuations inherent in any scientific experiment. They can arise from imperfections in the experimental apparatus, modeling of physical processes, or data analysis techniques. This section will categorize and quantify them precisely and elucidate their nature, sources, and mitigation strategies. By doing so, we can ensure that our scientific findings are robust, accountable, and transparent.

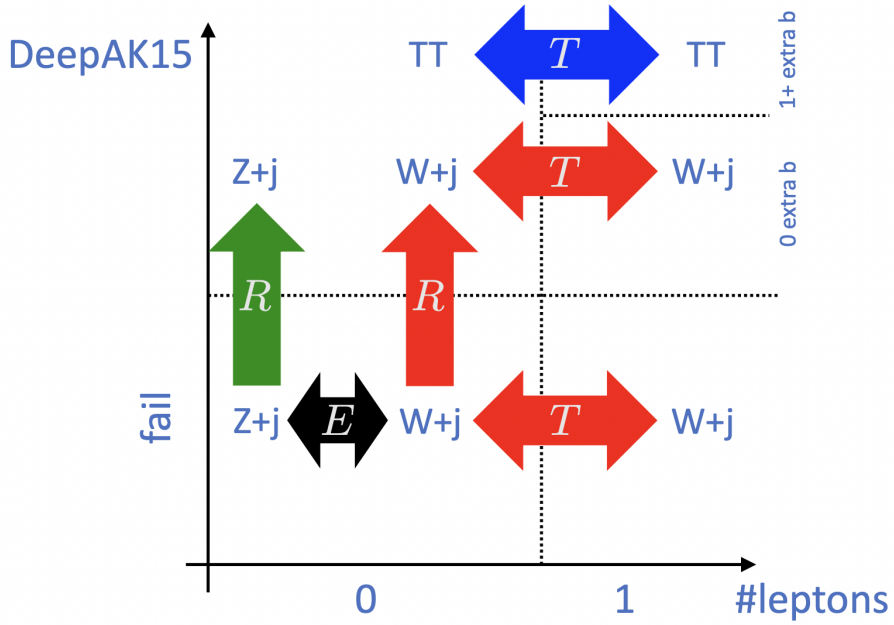

Figure 4.7: Transfer factors amongst the regions.

4.4.1 Heavy-flavor fraction uncertainty

To constrain the signal region using our control regions, we need to know the heavy-flavor fraction of the background accurately. This is necessary to build transfer factors between regions with different b-tagging requirements. However, the heavy-flavor fraction is not well-modeled by theory, so we use experimental results to estimate its uncertainty.

CMS has measured the differential cross-section of W+jets as a function of Njets [44]. From this study, we estimate that the uncertainty on the cross-section of W+fat jets is less than 20%.

Next, using CMS measurements of W+c [47] and W+$b\bar{b}$ [45] production, we extract the following uncertainties on heavy-flavor production:

$$\frac{\delta\sigma(W+c, W\to\ell\nu)}{\sigma(W+c, W\to\ell\nu)} = \frac{7.65\text{ pb}}{107.7\text{ pb}}, \quad \frac{\delta\sigma(W+b\bar{b}, W\to\ell\nu)}{\sigma(W+b\bar{b}, W\to\ell\nu)} = \frac{0.206\text{ pb}}{1.06\text{ pb}} \tag{4.14}$$

Then, the uncertainties are propagated as:

$$\begin{aligned}\frac{\delta\sigma(W+\text{HF})}{\sigma(W+\text{fat jet})} &= \left[\left(\frac{\delta\sigma(W+\text{HF})}{\sigma(W+\text{HF})}\right)^2 + \left(\frac{\delta\sigma(W+\text{fat jet})}{\sigma(W+\text{fat jet})}\right)^2\right]^{1/2} \\ &= \left[\frac{\delta\sigma(W+c)^2+\delta\sigma(W+b\bar{b})^2}{\left(\sigma(W+c)+\sigma(W+b\bar{b})\right)^2} + \left(\frac{\delta\sigma(W+\text{fat jet})}{\sigma(W+\text{fat jet})}\right)^2\right]^{1/2}\end{aligned}$$

Using the discussed values, the result can be written as:

$$\frac{\delta\sigma(W+\text{HF}, W\to\ell\nu)}{\sigma(W+\text{fat jet}, W\to\ell\nu)} \approx 0.21. \tag{4.15}$$

A similar calculation is done for Z+jets, where the estimated uncertainty for Z+fat jet production [48] is 15%, and the resulting uncertainty on the heavy-flavor fraction is 0.22.

4.4.2 b-tagging Scale Factors

The BTV POG provides scale factors and associated uncertainties for the b-tagging of narrow jets. These scale factors are propagated into event scale factors for this analysis, along with two shape variations: narrow jet b-tag and narrow jet mis-tag ([1]).

The scale factors and uncertainties provided by the POG cover the pT range probed in this analysis, up to 1000 GeV. For jets beyond this pT range, the recommendation is to use the scale factor from the p_T boundary and double the uncertainty ([1]).

As recommended by the BTV POG, we derive the MC efficiency for b-tagging jets in the phase space used by this analysis. The efficiencies are derived using the $t\bar{t}$ Monte Carlo sample and applying a U $>$ 200 GeV preselection. The jets are split into three categories: b jets (matched to a truth b), c jets (matched to a truth c), or light-flavor jets (unmatched).

4.4.3 Lepton Efficiency Scale Factors

We apply data-to-MC scale factors to events in the control regions used in the analysis. These scale factors are derived from the efficiencies of the electron and muon selections, measured in both data and simulation. The uncertainty on the lepton efficiency scale factor is correlated between the single muon/electron and dimuon/dielectron control regions. The uncertainties on the lepton efficiency scale factor are:

- Electron ID and tracking (1% and 0.5% per lepton leg, respectively)
- Muon ID and tracking (1% and 0.5% per lepton leg, respectively)
- Tau veto (3%)

A more thorough description of the lepton uncertainties is provided in [15].

4.4.4 Trigger Efficiencies

Single-electron triggers are used to select events with electrons. Their efficiencies in data are parametrized as a function of electron pT and applied to MC. The derivation of these efficiencies is described in [15].

E_T^{miss} triggers are used to select events with real E_T^{miss} or muons. Their efficiency is parametrized as a function of hadronic recoil U in such events. Specifically, E_T^{miss} computed without muons. However, the efficiency has a small dependence on the number of muons in the event due to muons faking jets in the HLT and lowering the trigger efficiency. At U = 200 GeV, the effect is small ($\approx$ 2.5%) and decreases at larger U. Since we are operating in U $>$ 250 GeV, this uncertainty is not applicable. The prescription used is the same as in [15].

4.4.5 QCD Normalization Uncertainty

This analysis estimates the QCD background from Monte Carlo simulation in all regions. A 100% uncertainty is applied to the normalization of the QCD, correlated between regions with the same source of the fake. This means that one uncertainty is used to QCD in the signal region, a separate uncertainty is applied to QCD in muon control regions, and similarly for electrons.

The 100% uncertainty is estimated from a QCD-enriched region defined by the following requirements:

- Veto electrons, muons, taus, photons
- Require $E_T^{\text{miss}} > 200$ GeV
- Require min $\Delta\phi(E_T^{\text{miss}}, \text{jet}) < 0.1$

4.4.6 Additional Uncertainties

In addition to the uncertainties on the luminosity measurement (2.5%) and theoretical cross-section for MC-driven processes, we also consider the following rate uncertainties:

- MET bin migration (5%): This uncertainty is derived by propagating the jet energy scale and resolution uncertainties to the MET [110, 111], as done in [15].
- AK15 jet energy scale (4%): This uncertainty is derived by propagating the AK8 jet energy scale uncertainties from the POG, with a factor of two added to cover the extrapolation from AK8 to AK15.

The next chapter explains that a DeepAK15 Scale Factor uncertainty is also calculated in a separate process.

Chapter 5

Scale Factors Determination for Double b Tagging

This chapter focuses on correcting the efficiency of the DeepAK15 tagger that identifies a fat jet resulting from the decay of the dark Higgs boson candidate by scale factors. Scale factors are calculated by applying the tagger in jets of a b-quark enriched sample from data from LHC Run 2 and Monte Carlo simulation. The results are the profile of scale factors that are applied as a correction in the analysis for the dark Higgs search.

It is important to correct the efficiency of the DeepAK15 tagger for this analysis because the dark Higgs boson candidate is expected to decay into two b-quarks, so it is important to accurately identify fat jets from b-quarks as the signal process. The b-quark enriched sample used to calculate the scale factors is selected by requiring the presence of a b-tagged jet or a lepton from the decay of a b-hadron. The uncertainties associated with the scale factors are due to the statistical uncertainty of the b-quark enriched sample and the systematic uncertainty of the simulation.

The scale factors are applied to the analysis for the dark Higgs search by multiplying them with the weights of the signal and background Monte Carlo simulations. This ensures that the simulations accurately represent the efficiency of the DeepAK15 tagger in the data.

5.1 Introduction to double b tagging

Identifying and characterizing particles, especially b quarks, is essential to our search for the Dark Higgs in the complex landscape of high-energy particle collisions. B quarks, which often arise from the decay of particles of interest, hold crucial clues to the presence of exotic phenomena. This section will explore double b tagging, a sophisticated technique central to our analysis.

Double b tagging is a powerful tool that allows us to increase the purity of our b quark samples and improve the sensitivity of our analysis to Dark Higgs events. This technique relies on the DeepAK15 algorithm to selectively identify events containing b quark pairs (bb events). The ability to distinguish between genuine b quarks and other particles is essential for our study.

Like the DeepAK8 algorithm [7], the DeepAK15 algorithm is a new approach to using particle-level information directly through custom machine learning (ML) methods. It is a versatile multi-class classifier that specializes in identifying hadronically decaying particles. It categorizes events into five primary classes: W, Z, H, top quark, and other particles. Each of these main classes is further subdivided into minor categories, corresponding to the various decay modes of each particle, such as $Z \to bb$, $Z \to cc$, and $Z \to qq$.

DeepAK15 defines two distinct lists of inputs for each jet under consideration (see Fig 5.1). The first, called the "particle" list, consists of up to 100 jet constituent particles sorted by their decreasing transverse momentum. This ensures that even if fewer than 5% of jets have more than

100 reconstructed particles, restricting the list to the 100 hardest particles results in negligible performance loss. The "particle" list is rich with essential information, including measured properties of each particle, such as pT, energy deposition, charge, angular separation from the jet axis or subjet axes, and more. Additional information derived from the tracking detector, such as track displacement and quality, is also included for charged particles. These inputs are essential for the algorithm to extract features relevant to the substructure of a jet, especially regarding the presence of heavy flavor (b or c) quarks. In total, 42 variables are included for each particle in the "particle" list.

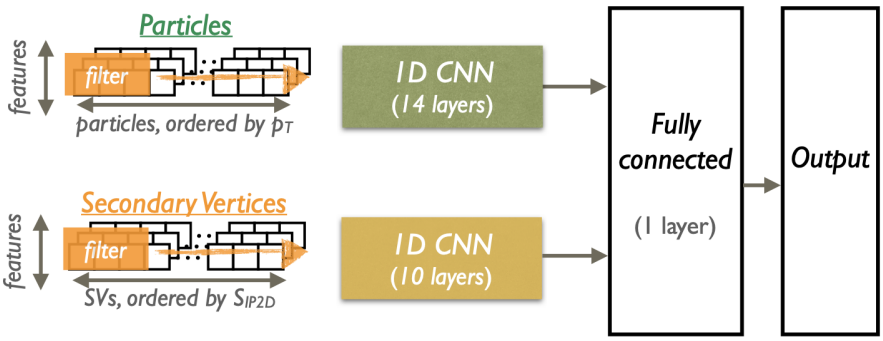

Figure 5.1: The network architecture of DeepAK8 [7]

In addition to the "particle" list, DeepAK15 also employs a secondary vertex (SV) list, which can contain up to 7 secondary vertices, each characterized by 15 distinct features. These features encompass SV kinematics, displacement information, and quality criteria. The SV list significantly contributes to the extraction of features that are closely tied to the heavy flavor content within the jet.

One significant challenge posed by the direct use of particle-level information is the substantial increase in the number of inputs and the correlations between these inputs. This requires an algorithm that can efficiently process the inputs while exploiting the correlations. To meet this requirement, DeepAK15 employs a customized deep neural network (DNN) architecture, meticulously crafted to process inputs efficiently and harness their crucial correlations. This architectural sophistication is a defining feature of DeepAK15, enabling it to excel in its mission of distinguishing b quark pairs (bb events) with remarkable precision in the context of our Dark Higgs analysis.

As we delve into the intricacies of double b tagging, we will uncover its significance in our quest to unveil the elusive Dark Higgs. We will also explore the methodology used to determine the scale factors that refine the accuracy of our double b tagging, enabling us to extract meaningful physics results from the CMS detector at the CERN LHC.

5.2 Methodology for determining double b tagging scale factors

Determining scale factors for double b tagging is essential for our analysis, as it enables us to distinguish Dark Higgs signals from background processes precisely. This subsection describes our rigorous methodology for determining scale factors. The method is based on the BTV group performing cc-tagger calibration for AK15 jets, as explained in [72].

We begin by considering an enriched QCD sample with a high proportion of b quarks. This sample is carefully designed to mimic the conditions of the data samples collected by the CMS detector. To further increase the purity of b quarks in our data sample, we use the BtagMu path, which selects events with a high probability of containing b quarks.

The DeepAK15 algorithm is applied within this enriched sample to identify events with pairs of b quarks (bb events). DeepAK15 is a powerful machine learning algorithm that exploits the distinctive features of b quarks and interactions to distinguish them from other particles.

Once we have identified bb events, we use the combined tool [84] by using the Coffea framework [83] to determine scale factors directed from the fit. The combined tool is a statistical framework that combines information from data and simulations to produce precise and accurate results. It performs a comprehensive analysis of the bb events in our enriched sample using the precision of DeepAK15 tagging. The scale factors that correct the double b tagging algorithm are obtained by comparing the observed and expected yields of bb events.

5.2.1 Proxy jets

To correct the DeepAK15 algorithm, ideally, a pure sample with the $b\bar{b}$ final process is required. However, it is difficult and sometimes impossible to obtain such a sample. For this reason, the strategy relies on using proxy jets.

Proxy jets correct the DeepAK15 algorithm without a pure jets sample with the $b\bar{b}$ final process. Proxy jets are chosen to have similar properties to the signal jets, such as the same transverse momentum and pseudorapidity.

One example of proxy jets is jets from $Z \rightarrow b\bar{b}$ decays. $Z \rightarrow b\bar{b}$ events are relatively easy to identify and select, and the b-jets in these events are expected to have similar properties to the b-jets in the signal events. Another example of proxy jets is jets from multijet events. Multijet events are prevalent in high-energy particle collisions, and they can be used to calibrate the DeepAK15 algorithm for a wide range of jet transverse momenta and pseudorapidities.

Once the DeepAK15 algorithm is corrected using proxy jets, the obtained corrections can be transferred to the signal jets. This is done by applying the same correction factors to the signal jets as the proxy jets.

The proxy jets utilized for this analysis are from the process g$\rightarrow b\bar{b}$, jets from a QCD sample. This sample contains significant statistics due to the number of events and considerable purity.

5.2.2 Triggers and MET Filters

The following triggers are used to select events for the double b tagging scale factors determination:

- HLT_BTagMu_AK4DiJet170_Mu5
- HLT_BTagMu_AK8Jet300_Mu5
- HLT_BTagMu_AK8Jet500_Mu5

The trigger HLT_BTagMu_AK4DiJet170_Mu5 is a high-level trigger (HLT) used in the CMS experiment, and it is designed to select events that contain a b-tagged jet with transverse momentum greater than 170 GeV, a muon with transverse momentum greater than 5 GeV, and a second jet with a transverse momentum greater than 170 GeV. The trigger HLT_BTagMu_AK8Jet300_Mu5 is designed to select events that contain a b-tagged jet with a transverse momentum greater than 300 GeV and a muon with a transverse momentum greater than 5 GeV. Similar to the trigger AK8Jet500.

These triggers are chosen to have a high efficiency for the signal events and a low rate of background events.

The MET triggers for the analysis are:

- goodVertices
- globalSuperTightHalo2016Filter
- HBHENoiseFilter

- HBHENoiseIsoFilter
- EcalDeadCellTriggerPrimitiveFilter
- BadPFMuonFilter
- ecalBadCalibFilterV2 (for 2017 and 2018)

5.2.3 Jet Selection

The following criteria are used to select jets for the double b tagging scale factors determination:

AK15 jet selection

- $p_T > 250$ GeV
- $m_{SD} > 50$ GeV

Required including two subjects that match a global muon (reconstructed using both the tracker and the muon system of a particle detector)

- within $\Delta R < 2.4$ from subject axis
- muon $p_T > 7$ GeV
- $p_T(\text{muon})/p_T(\text{subject}) < 0.7$

The selection criterion pT (muon)/pT (subject) < 0.7 is designed to select muons that have a lower transverse momentum than the jet that they are associated with. This is because muons that originate from the decay of a heavy particle are expected to have a lower transverse momentum than the jet they are associated with.

By selecting muons with a low transverse momentum relative to the jet they are associated with, this selection criterion can help reduce the background from other sources, such as cosmic rays and muons produced in the decay of light quarks.

Secondary Vertex (SV) selection These criteria are chosen to ensure that the selected jets are high quality and likely to contain b quarks. The selection is as follows:

- $\Delta R < 1.5$ matching with a selected AK15 jet.

5.2.4 Datasets

The following datasets are used to determine the double b tagging scale factors:

Data

- 2016: BTagMu/Run2016B-H,
- 2017: BTagMu/Run2017B-F,
- 2018: BTagMu/Run2018A-C, Run2018D-PromptReco.

The letters on the run labels indicate the data-taking period. Using PromptReco implies the dataset was processed with the PromptReco reconstruction software.

Monte Carlo

- QCD_HT binned samples
  - 200-300, 300-500, 500-700, 700-1000, 1000-1500, 1500-2000, 2000-inf
- Privately produced signal samples ($h_s \to b\bar{b}$)
  - Only for QCD proxy validation.

5.2.5 Fit and Combine Tool

To ensure precise and accurate results, a robust methodology is necessary for determining the scale factors for double b tagging. The Combine Tool is a crucial framework of this methodology and is central to the scale factor determination process. This tool performs several essential functions that contribute to the precision of our results.

The input of the combine tool is a datacard file that defines the details of the experiment. The datacard begins with an optional description, followed by the declaration of the number of observables, background sources, and systematic uncertainties.

The datacard then specifies the number of events observed in each channel, where the "bin" line labels the channels, and the "observation" line lists the observed event counts. Following this, the expected event yields for each bin and process are provided and organized in columns. The "bin" line identifies the channel, the "process" line labels the sources, and the "rate" line indicates the expected event yield.

The final section of the datacard describes the systematic uncertainties. Each source of uncertainty is labeled, and its type of distribution (e.g., log-normal or log-uniform) is specified. The impact of each uncertainty on the event rates for different processes and channels is quantified in the corresponding columns. For the analysis, shape uncertainties are utilized. These represent the uncertainty in the form or distribution of events in a histogram for a specific process or channel in particle physics analyses.

One of the primary functions of the Combine Tool is to perform a statistical fit on both the observed data and simulated samples. By modeling the observed data and comparing it to Monte Carlo simulations, the tool provides invaluable insights into the performance of the DeepAK15 tagger in a data-like environment.

The model used for the analysis is the "FitDiagnostics," whereby taking the total background, the total signal, and the data shapes from FitDiagnostics output, we can compare the post-fit predictions from the S+B fit (first case) and the CR-only fit (second case). The $b\bar{b}$ is considered the signal for this part of the analysis. Meanwhile, other jet processes like c, $c\bar{c}$, b, and other light flavors are regarded as background.

Once the statistical fit is complete, the tool calculates the scale factors, a rate parameter inside the datacards. Rate parameters are used to model the normalization of specific processes or bins as functions of other parameters. These factors are determined by analyzing the observed yields of $\bar{b}$ events and comparing them with the expected yields from our simulations. They are correcting the double b tagging algorithm to accurately reflect the CMS detector's conditions.

Our scale factor determination is based on observed LHC Run 2 data obtained through the BTagMu path and trigger selections. The DeepAK15 tagging algorithm's precision is used to identify bb events.

A datacard example is presented as follows:

```
# Simple Counting Experiment
# Analysis for Signal Process (Sig) and Background Process (Bkg)
imax 1   # Number of channels
jmax 1   # Number of background sources
kmax 1   # Number of nuisance parameters (systematic uncertainties)
------------------------------------------------------------
# Channel information
bin 1
observation 10  # Number of observed events in the channel
------------------------------------------------------------
# Process information
bin         1
process     Sig Bkg
process     0   1
```

```
rate        5.0 5.0  # Expected yields for Signal and Background
------------------------------------------------------------------
# Systematic uncertainties
lumi lnN 1.1 1.1  # Luminosity uncertainty
shapeUnc shape 1.0 - # Shape uncertainty
------------------------------------------------------------------
```

5.2.6 Corrections and Uncertainties

When determining scale factors, there are two primary sources of uncertainty. Statistical uncertainty comes from the statistics of the bb-quark enriched sample obtained from data, affecting the scale factors' precision. Meanwhile, systematic uncertainty arises from the uncertainties in the simulation process, including variations due to physical processes, detector responses, and other systematic effects that can affect the determination of the scale factor. Table 5.1 summarizes the uncertainties for this analysis.

Uncertainty Source	**Type**	**Signal**	**Background**
svmassYearfail_mcstat_binNumber	Shape	-	1.0
svmassYearpass_mcstat_binNumber	Shape	1.0	-
fracYear	lnN	1.5	1.5
jesYear	lnN	1.02	1.02
lumiYear	lnN	1.027	1.027
puYear	lnN	1.05	1.05

Table 5.1: Summary of Uncertainties

The uncertainties svmassYearfail_mcstat_binNumber and svmassYearpass_mcstat_binNumber specify that certain bins in the analysis are affected by shape uncertainties. The sf_weightYear_effect_svmass2017pass_genbb is the rate param uncertainty to determine the scale factors. The frac uncertainty refers to the quark flavor uncertainty. Jes stands for "Jet Energy Scale." It refers to uncertainties related to the energy measurement of jets. Lumi refers to "Luminosity" uncertainty. Luminosity represents the number of particle collisions occurring in an accelerator experiment over a given period. Finally, pu refers to "Pileup" uncertainty. Pileup refers to additional, unrelated proton-proton collisions that occur along with the primary collision of interest in a particle physics experiment. It can affect the reconstruction and identification of particles.

To calculate the uncertainties for the scale factor determination, we calculate one uncertainty by considering the total process in a single bin instead of calculating uncertainties from each process in a single bin. This means instead of calculating statistical uncertainties by considering each process where each process has a different uncertainty, we first calculate the sum of processes' yields and errors. Then, a single uncertainty will apply to each process.

The BBlite method (Barlow and Beeston method [59]) is a systematic uncertainty estimation method that combines the statistical uncertainties of multiple processes into a single uncertainty. This is done by calculating the sum of the processes' yields and errors in each template bin. A single uncertainty is then applied to each process in each bin. This method is more conservative than the traditional method of calculating statistical uncertainties but is also more accurate.

Initially, the code was:

```
1 merged_central, merged_error2 = bkg_hists["svtemplate"].integrate("gentype").
      integrate("process").values(sumw2=True)[()]
2 merged_central=merged_central[:, category_map[category]]
3 merged_error2=merged_error2[:, category_map[category]]
```

Where *gentype* refers to data or background, and the *process* are BTagMu or QCD. The *category* refers to pass or fail the requirement. However, this method has limitations in the analysis due to it does not consider the list for MC (bb, b, cc, c, others) and the number of bins (from 0 to 21). This has implications for the merged_central and merged_error arrays.

To fix this problem, we utilize a pair of functions to consider these parameters:

```
def get_mergedMC_stat_variations(dictionary, category, mc_list):
    """
    Get the merged central and error squared values for a list of MC processes.

    Args:
        dictionary: A dictionary containing the MC data.
        category: The category to get the data for.
        mc_list: A list of MC processes to get the data for.

    Returns:
        A tuple of the merged central and error squared values.
    """

    templ=template(dictionary, 'QCD', mc_list[0], category, read_sumw2=True)
    merged_central=np.zeros_like(templ[0])
    merged_error2=np.zeros_like(templ[3])
    for mc in mc_list:
        templ=template(dictionary, 'QCD', mc, category, read_sumw2=True)
        for i in range(len(templ[0])):
            if templ[0][i] <= 1e-5 or templ[3][i] <= 0.:
                continue
            merged_central[i] += templ[0][i]
            merged_error2[i]  += templ[3][i]
    return merged_central, merged_error2
```

The function then iterates over the list of MC processes and gets the template for each process. For each bin in the template, the function checks if the central value and error squared value are greater than zero. If they are, the function adds the central value and error squared value to the corresponding arrays in "merged_central" and "merged_error2".

Then, we used a function that works by iterating over all bins in the template and calculating the effect of the systematic uncertainty on the central value of the template in each bin:

```
def addBBliteSyst(templ, param, merged_central, merged_error2, epsilon=0):
    for i in range(templ.observable.nbins):
        if merged_central[i] <= 0. or merged_error2[i] <= 0.:
            continue
        if templ._nominal[i] <= 1e-5:
            continue
        effect_up = np.ones_like(templ._nominal)
        effect_down = np.ones_like(templ._nominal)
        effect_up[i] = 1.0 + np.sqrt(merged_error2[i])/merged_central[i]
        effect_down[i] = max(epsilon, 1.0 - np.sqrt(merged_error2[i])/
            merged_central[i])
        templ.setParamEffect(param[i], effect_up, effect_down)
```

This function takes the following arguments:

- templ: The template to apply the systematic uncertainty.
- param: The parameter to apply the systematic uncertainty.
- merged_central: The merged central values of the processes.
- merged_error2: The merged error squared values of the processes.
- epsilon: A small value to avoid zero division errors.

To calculate the effect-up value, take the square root of the total error of the bin's processes and divide it by the average central value of those processes. Then add this value to the central value of the bin's template. Conversely, to obtain the effect-down value, subtract the square root of the total error of the bin's processes from the average central value of those processes.

5.3 Results and implications of double b tagging scale factors

In the previous sections, we provided a detailed explanation of determining scale factors crucial for the DeepAK15 double b tagging algorithm's accuracy. This section focuses on the results obtained from our scale factor determination process and their significant impact on our analysis of the elusive Dark Higgs boson.

In our efforts, we were able to derive scale factor profiles that correct the DeepAK15 tagger's efficiency with great precision. These profiles are the foundation of our analysis, allowing us to accurately identify fat jets originating from b-quarks, which are key signals in our investigation.

The scale factor profiles are often presented as a function of various kinematic variables, allowing us to make specific corrections based on the event and particle characteristics we are analyzing. This detailed approach ensures that our analysis remains sensitive to subtle differences in the DeepAK15 tagger's behavior across different energy regimes and event types.

By applying these scale factors to our analysis, we can ensure that the DeepAK15 tagger performs consistently between our data and Monte Carlo simulations. This alignment provides the simulations to reflect the tagger's performance in real data conditions accurately. As a result, our results are more accurate and reliable.

The obtained scale factors are presented in Table 5.2, and the plots are presented in Fig 5.2, 5.3, 5.4, 5.5, 5.6, and 5.7.

Year	Scale Factor
2016	8.5349e-01 ± 1.52e-02
2017	9.3763e-01 ± 5.87e-03
2018	9.8177e-01 ± 1.02e-02

Table 5.2: Determinated Scale Factors for the year.

The scale factors are close to 1, so the algorithm efficiently identifies double b-jets. The uncertainties on the scale factors are relatively small, which means that the scale factors are well-measured.

- The scale factor for 2016 is 0.85349 ± 0.0152. This means that the b-tagging algorithm is 85.349% efficient at identifying b-jets in the 2016 data, with an uncertainty of 1.52%.
- The scale factor for 2017 is 0.93763 ± 0.00587. This means that the b-tagging algorithm is 93.763% efficient at identifying b-jets in the 2017 data, with an uncertainty of 0.587%.
- The scale factor for 2018 is 0.98177 ± 0.0102. This means that the b-tagging algorithm is 98.177% efficient at identifying b-jets in the 2018 data, with an uncertainty of 1.02%.

The mass of the first secondary vertex (SV1) is represented by the variable $\log(m_{SV1,d_{xy}sigmax}/\text{GeV})$, a logarithmic function of the distance between the SV1 and primary vertex in the transverse plane. Suppose the distance between the primary and secondary vertices is more significant. In that case, it is more likely that the secondary vertex came from the decay of a b-quark with a longer lifespan than other quarks.

The signed d_{xy} distance measures the displacement of the secondary vertex from the primary vertex in the transverse plane and is used to determine if the secondary vertex came from the decay of a b-quark. If the displacement is in the same direction as the muon, it is more likely

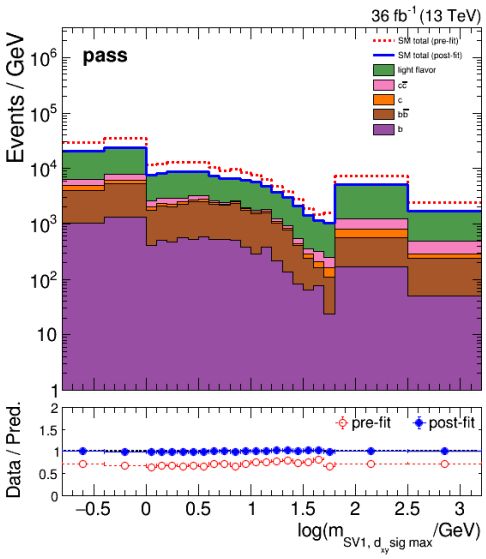

Figure 5.2: Results 2016 Pass

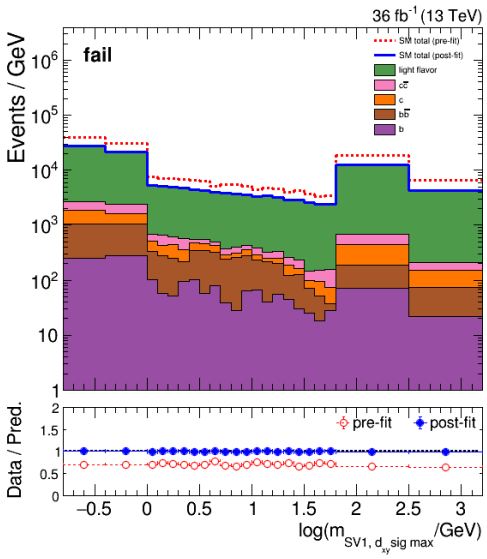

Figure 5.3: Results 2016 Fail

that the secondary vertex was produced by the decay of a b-quark since they are often created in pairs with muons and travel in the same direction. To differentiate between signal events (where the secondary vertex likely came from a b-quark decay) and background events (where the secondary vertex is more likely to have originated from the decay of a light quark), the variable $\log(m_{SV1,d_{xy}sigmax}/\text{GeV})$ is used.

The variable has both pre-fit and post-fit distributions generated by the combined tool. The pre-fit distribution shows the variable's distribution before the tool fits the data. In contrast, the post-fit distribution shows the variable's distribution after it has been fitted. The combined tool uses a model with signal and background components to fit the data, with parameters such as the number of signal and background events determined through the process. The post-fit distribution is then calculated based on these parameters. The post-fit distribution is considered more accurate because it considers the information from the fit. The post-fit distribution will shift towards higher variable values if the fit shows more signal events than background events.

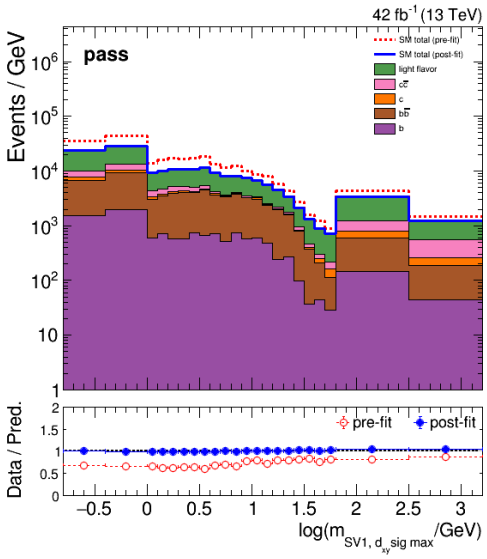

Figure 5.4: Results 2017 Pass

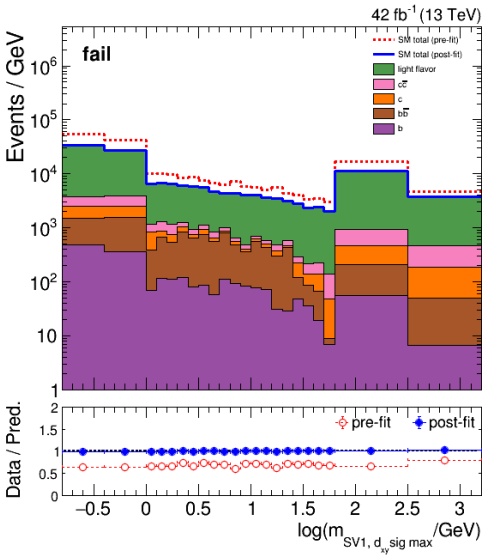

Figure 5.5: Results 2017 Fail

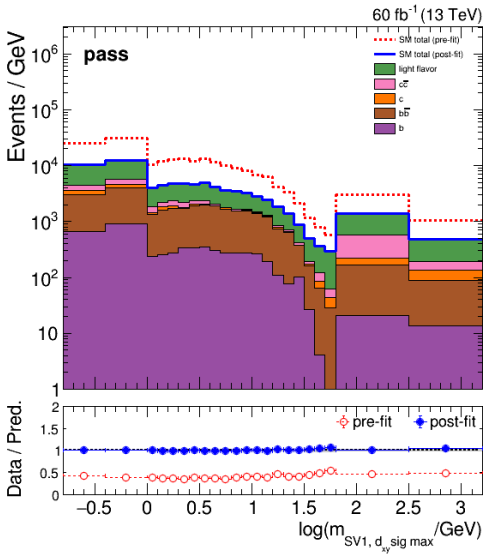

Figure 5.6: Results 2018 Pass

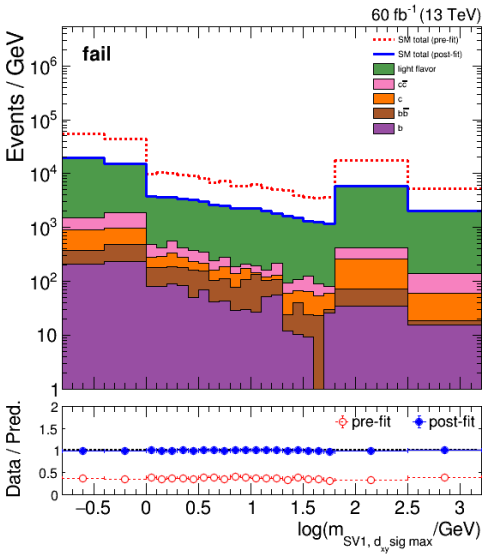

Figure 5.7: Results 2018 Fail

Chapter 6

Results

Exclusion limits are calculated in the two-dimensional parameter space of the dark matter and mediator masses, m_{DM} and m_{med}. The coupling between the mediator and Standard Model quarks is set to a constant value of $g_q = 0.25$, and the mediator-DM coupling is set to $g_\chi = 1.0$. The resulting exclusion limits at 95% confidence level (CL) on the signal strength μ are shown in Figures 6.1, 6.2, 6.3, 6.4, 6.5, and 6.6 for different hypotheses of the dark Higgs boson mass.

For small values of $m_{DM} \approx 1$ GeV, different values of the mediator mass m_{med} are expected to be excluded as a function of the dark Higgs boson mass. The excluded value of m_{med} reduces with increasing values of m_{DM}, as the branching fraction for decays of the mediator into DM candidates is reduced. The results are compared to those from direct detection experiments.

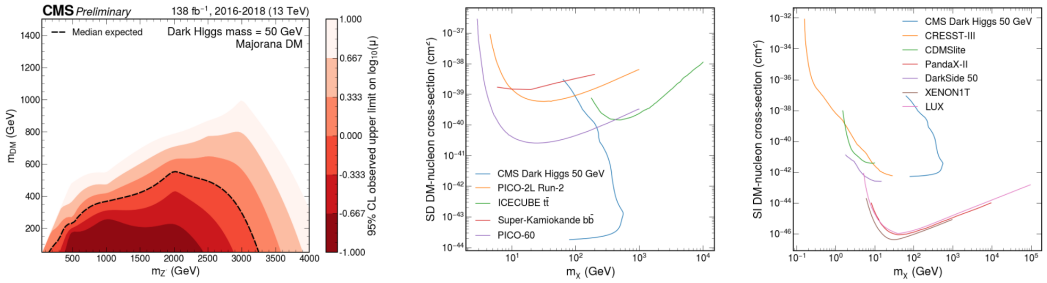

Figure 6.1: Expected exclusion limits for a dark Higgs boson mass of 50 GeV. Left: 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of m_{med}. The black solid line indicates the exclusion boundary $\mu = 1$. Middle: 90% CL on the spin-dependent DM-nucleon cross-section as a function of m_{DM}. Right: 90% CL on the spin-independent DM-nucleon cross-section as a function of m_{DM}.

CL stands for confidence level. Given the collected data, it measures the probability that a given result is true [21]. In the context of statistical hypothesis testing, the CL is the probability of rejecting a true null hypothesis.

Limits are used to quantify the uncertainty in a measured value. In this context, limits are often used to constrain the parameters of a model, such as the mass of a new particle or the cross-section of a process. It is important to note that these are just expected exclusion limits.

Direct detection of dark matter involves observing the interaction between the DM particle and the nucleon in the detector material. The nucleus or nucleon in the target experiences elastic scattering, resulting in recoil and energy measurement. This measurement is used to determine the scattering cross-section and the mass of the dark matter particle [23]. Meanwhile, the indirect detection of dark matter is based on the idea that dark matter particles can annihilate with each other to produce standard model particles, such as photons, neutrinos, and antimatter. The cross-

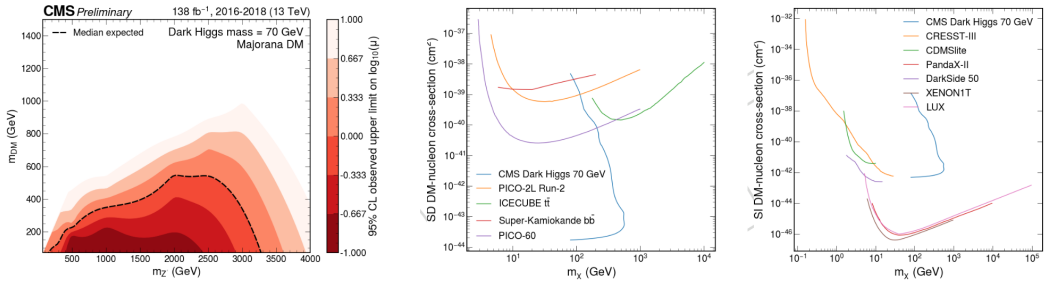

Figure 6.2: Expected exclusion limits for a dark Higgs boson mass of 70 GeV. Left: 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of m_{med}. The black solid line indicates the exclusion boundary $\mu = 1$. Middle: 90% CL on the spin-dependent DM-nucleon cross-section as a function of m_{DM}. Right: 90% CL on the spin-independent DM-nucleon cross-section as a function of m_{DM}.

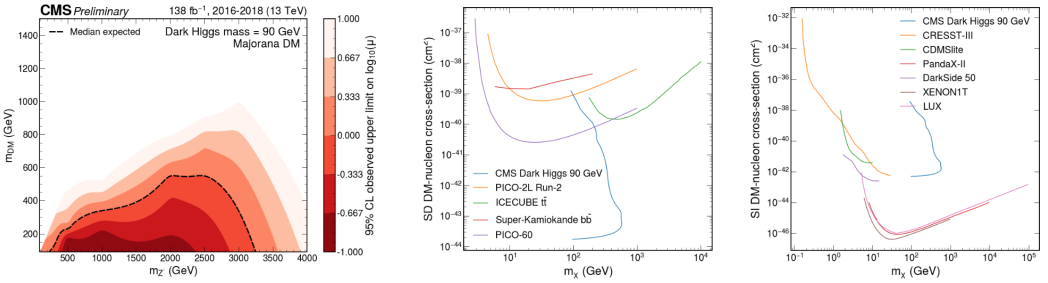

Figure 6.3: Expected exclusion limits for a dark Higgs boson mass of 90 GeV. Left: 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of m_{med}. The black solid line indicates the exclusion boundary $\mu = 1$. Middle: 90% CL on the spin-dependent DM-nucleon cross-section as a function of m_{DM}. Right: 90% CL on the spin-independent DM-nucleon cross-section as a function of m_{DM}.

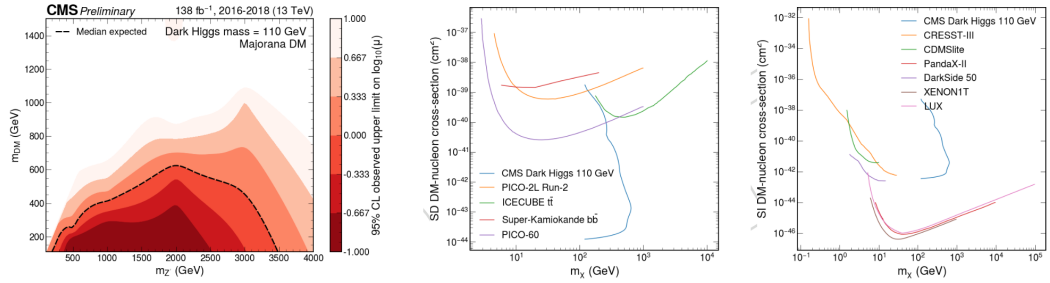

Figure 6.4: Expected exclusion limits for a dark Higgs boson mass of 110 GeV. Left: 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of m_{med}. The black solid line indicates the exclusion boundary $\mu = 1$. Middle: 90% CL on the spin-dependent DM-nucleon cross-section as a function of m_{DM}. Right: 90% CL on the spin-independent DM-nucleon cross-section as a function of m_{DM}.

section for dark matter annihilation is a key parameter in indirect detection experiments [81].

The process of obtaining DM-nucleon cross sections from collider cross sections is explained in references [36], [60], and [31]. The couplings, mediator mass, and DM mass play important roles.

The cross-sections are compared with the results from other experiments for direct and indirect dark matter detection, as presented in [119] for CMS results. More results are presented on [126], [71], [20], and [117].

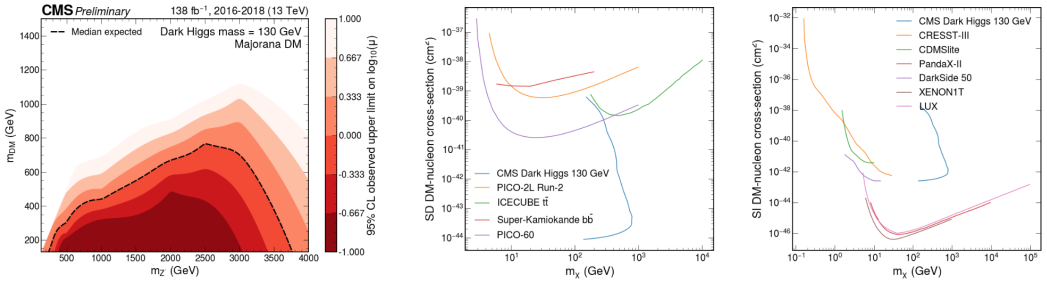

Figure 6.5: Expected exclusion limits for a dark Higgs boson mass of 130 GeV. Left: 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of m_{med}. The black solid line indicates the exclusion boundary $\mu = 1$. Middle: 90% CL on the spin-dependent DM-nucleon cross-section as a function of m_{DM}. Right: 90% CL on the spin-independent DM-nucleon cross-section as a function of m_{DM}.

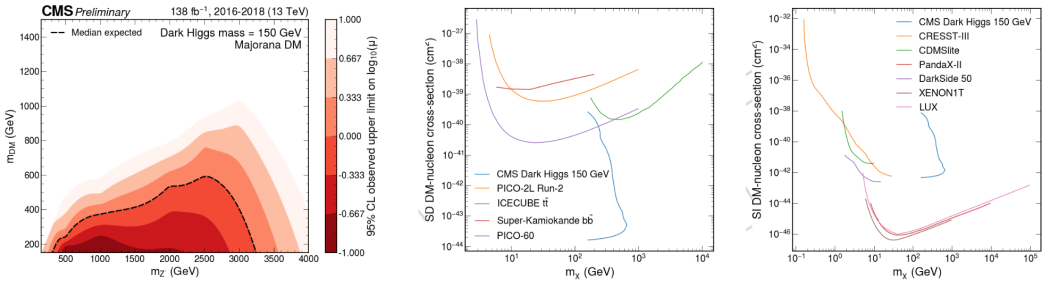

Figure 6.6: Expected exclusion limits for a dark Higgs boson mass of 150 GeV. Left: 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of m_{med}. The black solid line indicates the exclusion boundary $\mu = 1$. Middle: 90% CL on the spin-dependent DM-nucleon cross-section as a function of m_{DM}. Right: 90% CL on the spin-independent DM-nucleon cross-section as a function of m_{DM}.

Conclusions

This thesis has presented the status of a dark Higgs boson analysis in the b-quark final state channel. The analysis was performed using the full Run 2 dataset collected by the CMS experiment, corresponding to an integrated luminosity of 137 fb^{-1}. Expected exclusion limits were set on the dark Higgs boson production cross section at the 95% confidence level.

Determining scale factors was a relevant part of the analysis, as it was necessary to identify the fat jets properly. One challenge was the presence of a large number of systematic uncertainties. To address this, a BBLite fitting technique was used. The combine tool allowed for the simultaneous determination of the signal and background scale factors and the nuisance parameters. The scale factors were determined with high precision, and they played an important role in the success of this analysis.

Appendix A

RPC Lab and RH sensor

The objective was to develop a low-cost relative humidity sensor connected to the RPC gas line on the CMS RPC Ecocampus Laboratory at BUAP and measure the relative humidity with high accuracy and precision.

The utilized materials are:

- Arduino Uno.
- Relative humidity and temperature sensor DHT11.
- Jumper wires 15 cm in length.
- 16 mm brass connections. (reductions, T, Nuts)
- 8 mm brass connections. (reductions, Nuts, and ferrules).
- 16 mm copper pipe.
- 8 mm copper pipe.
- Glue.
- Brass plug.

Methodology

Construction. First, the jumper wires connected the Arduino to the sensor (See Fig A.1 for wires connections) to test whether it worked and compared it with a digital temperature sensor (Fig A.2). A protoboard was used to fix the sensor to connect the Arduino. However, it was used only for this step.

After comparing both sensors, the next step was to put them into a container where the gas could go through a gas line supply without gas leaks. The 16 mm copper pipe has a length of 10 cm, and the 8 mm copper pipe has a distance of 5 cm. The sensor should be between the T connector and the 16 mm copper pipe; thus, the wires must be fixed to avoid disconnecting. Then, the jumper wires were soldered using a tin solder at the sensor, and the other side was free to be connected to the Arduino.

Therefore, the next step is to put the sensor inside the T and the copper pipe. However, the wires are tiny, and it will cause gas leaks (See Fig A.3). Moreover, manipulating the wires can disconnect them from the sensor, so fixing and closing the T connector where the wires go out was necessary.

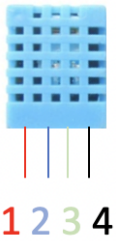

Figure A.1: DTH11 Sensor connections; Red color (1) for Vcc, blue color (2) for data, green color (3) not used, and black color (4) for ground.

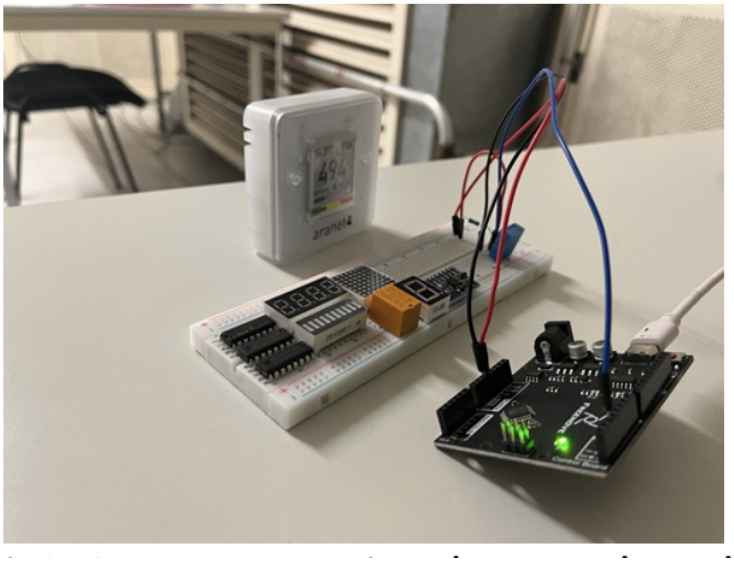

Figure A.2: Sensor Test using the protoboard.

For these reasons, a gas plug was connected to the T connector to seal it. To permit the pass of the cables, the connector was drilled with a few millimeters drill bit, as shown in Fig A.4. Then, the wires passed through the hole and were sealed with glue, then one day was left to dry it. After fixing the wires, it is possible to put all the connectors together. If the copper pipe were cut with the wrong tool, a new cut must be done with the tube cutter to avoid gas leaks. Once all the components are connected, a last connection test is done. Then, the sensor is prepared to be tested, passing gas through it (See Fig A.5).

Methodology

The requirements used for the test are listed below

- Nitrogen gas.
- A steal bubbler to humidify the gas.
- A thermometer.

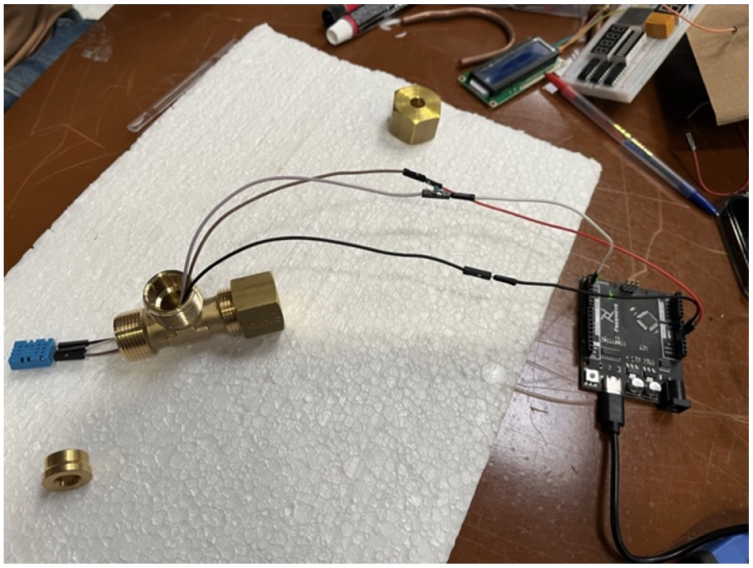

Figure A.3: Sensor assembly.

- An industrial sensor to measure the dew point (Dewpoint transmitter DMt1432 WAISALA 16003, 220V).
- 6 mm plastic pipes with 6 mm connectors.
- A refrigerator.

Note: The reason for using nitrogen gas was due to the gas pressure needed to generate a bubbler inside the steal bubbler. Lower pressure cannot produce it.

The setup is as follows: The 6 mm plastic pipe was connected to the gas rack at the nitrogen line. The gas line entered the refrigerator, and the output was connected to the industrial sensor. Finally, the industrial sensor output was connected to a gas exit. Inside the refrigerator, the gas line was connected to the bubbler, and the output connected to the sensor, and then the gas line left the refrigerator. A thermometer was inside the refrigerator to measure the temperature. The water inside the bubbler was one litter.

The simplified Magnus-Tetens formula [95] was used for the test, which indicates the change in relative humidity by changing two parameters: Temperature and dew point. This setup changes the temperature using the refrigerator; the thermometer measured that, and the industrial sensor measured the dewpoint. The Magnus-Tetens formula is obtained from the Clausius-Clapeyron equation, and it is written as follows:

$$RH = 100 \cdot \frac{\exp\left(\frac{17.625 \cdot TD}{243.04+TD}\right)}{\exp\left(\frac{17.625 \cdot T}{243.04+T}\right)}, \tag{A.1}$$

TD refers to the dew point in Celsius degrees, and T for temperature.

The test stated a gas flux of 5 L/h at 0.66 mBar pressure. The initial temperature was 18.8 C. Then, the temperature was modified by the refrigerator and measured, as well as the dew point. The results are presented in the table below.

Figure A.4: A gas plug is being drilled to permit the passing of the wires.

Results

Using the Magnus-Tetens formula and the sensor's measured results, a plot shows the temperature on the x-axis and the dewpoint on the y-axis. On the right axis, a color bar shows the relative humidity from 0 to 1, where 1 represents 100 percent relative humidity. The dot represents the calculated relative humidity by the theory; the X represents the measured relative humidity by the sensor, and the triangle represents a point measured using the CMS RPC gas mixture to compare it with the results. The measurements are presented in Table A.1 and Fig A.6.

Conclusions

A cheaper (compared to industrial ones that cost more than 100 EUR) and portable system was constructed and tested. However, the price limits the precision of the sensor in a range of 20 to 85 relative humidity percent. Meanwhile, it has a fast response, and the operational range will be 30-50 relative humidity. Therefore, no problems are expected. The commercial refrigerator temperature has limitations, and the method needs to be improved. It is impossible to achieve all temperature ranges with the refrigerator, and a different method is required.

Preliminary results show a good agreement with the predicted results. The difference between the measured and the predicted outcome is less than 2 percent, excluding the measurement with

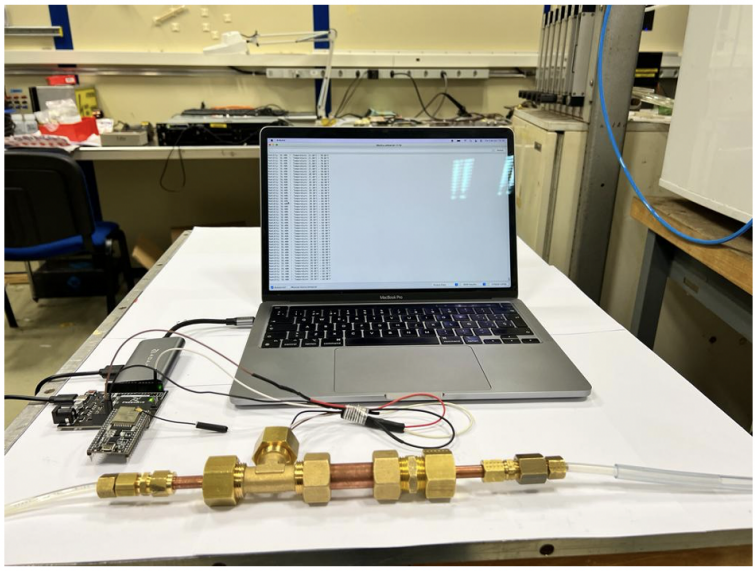

Figure A.5: Sensor connections for gas test.

Measurement	Predicted	The new system	Difference
36.15 %	35.5 %	35.5 %	0.65 %
31.47 %	30.4 %	30.4 %	1.07 %
29.31 %	30.4 %	30.4 %	1.09 %
25.27 %	25.0 %	25.0 %	0.27 %
98.75 %	85.0 %	85.0 %	13.75 %
38.68 %	43.0 %	43.0 %	26.32 %

Table A.1: The predicted results by the Magnus-Tetens formula and the measured results are presented. The measurement differences are listed on the right side.

the RPC gas mixture. This variation must be explained by setup configuration. An investigation is required.

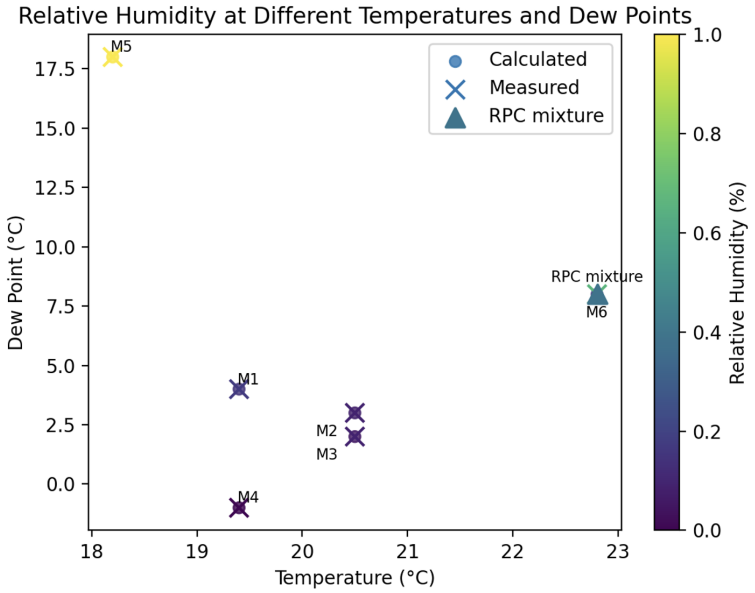

Figure A.6: Plotted results for the predicted and the measured relative humidity.

Bibliography

[1] Cms btag. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80X.

[2] Cms jetmet pog. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2.

[3] Cms jetmet pog.

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

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

[6] The Phase-2 Upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, 2017. This is the final version, approved by the LHCC.

[7] Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment. Technical report, CERN, Geneva, 2019.

[8] Pileup mitigation at CMS in 13 TeV data. Technical report, CERN, Geneva, 2019.

[9] 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. *Physics Letters B*, 716(1):1–29, 2012.

[10] K Abe and (forthe XMASS Collaboration). The xmass experiment. *Journal of Physics: Conference Series*, 120(4):042022, jul 2008.

[11] et al Abercrombie D. Dark matter benchmark models for early LHC run-2 searches: Report of the ATLAS/CMS dark matter forum. *Physics of the Dark Universe*, 27:100371, jan 2020.

[12] Yiming Abulaiti. Status of searches for dark matter at the LHC. Technical report, CERN, Geneva, 2022.

[13] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, jul 2005.

[14] Wolfgang Adam, Boris Mangano, Thomas Speer, and Teddy Todorov. Track Reconstruction in the CMS tracker. Technical report, CERN, Geneva, 2006.

[15] N. AK4churin et al. Search for dark matter produced with an energetic jet, or a hadronically decaying w or z boson, at $\sqrt{s} = 13$ tev with the full 2016 dataset, 2016. CMS Note 2016/473.

[16] N. Akchurin et al. Search for dark matter produced with an energetic jet, or a hadronically decaying w or z boson, at $\sqrt{s} = 13$ tev with the full 2016 dataset. CMS Note, 2016.

[17] R.L. Workmanet al.(Particle Data Group). H mass. prog. theor. exp. phys. 2022, 083c01, 2022.

[18] Alexandre Alves, Stefano Profumo, and Farinaldo S. Queiroz. The dark z′ portal: direct, indirect and collider searches. *Journal of High Energy Physics*, 2014(4), apr 2014.

[19] J. Alwall and et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, 2014.

[20] C. et al. Amole. Dark matter search results from the complete exposure of the pico-60 c_3f_8 bubble chamber. *Phys. Rev. D*, 100:022001, Jul 2019.

[21] The CMS Experiment at CERN. Higgs boson – terms and definitions.

[22] Yang Bai, Patrick J. Fox, and Roni Harnik. The tevatron at the frontier of dark matter direct detection. *Journal of High Energy Physics*, 2010(12), dec 2010.

[23] Amit Dutta Banik and Debasish Majumdar. Extension of minimal fermionic dark matter model: a study with two higgs doublet model. *The European Physical Journal C*, 75(8), aug 2015.

[24] David Barney. Presentation for public - Introduction to CMS for CERN guides. 2013.

[25] L. Baudis, J. Hellmig, H. V. Klapdor-Kleingrothaus, B. Majorovits, Yorck Alexander Ramachers, and H. Strecker. Status of the Heidelberg Dark Matter Search (HDMS) experiment. In *1st International Conference on Particle Physics and the Early Universe*, pages 219–223, 1997.

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

[27] Anna Benecke. *Searches for new heavy bosons and vector-like quarks with the CMS experiment at $\sqrt{s}$ = 13 TeV and novel pileup mitigation techniques.* PhD thesis, Hamburg Germany, 2020.

[28] Lars Bergström. Dark matter candidates. *New Journal of Physics*, 11(10):105006, oct 2009.

[29] R Bernabei, P Belli, F Cappella, R Cerulli, C J Dai, A d'Angelo, H L He, A Incicchitti, H H Kuang, X H Ma, F Montecchia, F Nozzoli, D Prosperi, X D Sheng, and Z P Ye. Results from the dama/libra experiment. *Journal of Physics: Conference Series*, 203(1):012003, jan 2010.

[30] Andreas Birkedal, Konstantin Matchev, and Maxim Perelstein. Dark matter at colliders: A model independent approach. *Physical Review D*, 70(7):077701, 2004.

[31] Antonio Boveia, Oliver Buchmueller, Giorgio Busoni, Francesco D'Eramo, Albert De Roeck, Andrea De Simone, Caterina Doglioni, Matthew J. Dolan, Marie-Helene Genest, Kristian Hahn, Ulrich Haisch, Philip C. Harris, Jan Heisig, Valerio Ippolito, Felix Kahlhoefer, Valentin V. Khoze, Suchita Kulkarni, Greg Landsberg, Steven Lowette, Sarah Malik, Michelangelo Mangano, Christopher McCabe, Stephen Mrenna, Priscilla Pani, Tristan du Pree, Antonio Riotto, David Salek, Kai Schmidt-Hoberg, William Shepherd, Tim M.P. Tait, Lian-Tao Wang, Steven Worm, and Kathryn Zurek. Recommendations on presenting lhc searches for missing transverse energy signals using simplified s-channel models of dark matter. *Physics of the Dark Universe*, 27:100365, 2020.

[32] James Braun and Daan Hubert. Searches for wimp dark matter from the sun with amanda, 2009.

[33] Philippe Brax, Sylvain Fichet, and Philip Tanedo. The warped dark sector. *Physics Letters B*, 798:135012, nov 2019.

[34] O. Buchmueller et al. Characterising dark matter searches at colliders and direct detection experiments: vector mediators. *Journal of High Energy Physics*, 2015(1):037, 2015.

[35] Oliver Buchmueller, Caterina Doglioni, and Lian-Tao Wang. Search for dark matter at colliders. *Nature Physics*, 13(3):217–223, mar 2017.

[36] Oliver Buchmueller, Matthew J Dolan, S. A. Malik, and Patrick Meade. Characterising dark matter searches at colliders and direct detection experiments: vector mediators. *Journal of High Energy Physics*, 2015(1):37, 2015.

[37] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. 2008(04):063–063, apr 2008.

[38] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, apr 2008.

[39] Linda Carpenter, Anthony DiFranzo, Michael Mulhearn, Chase Shimmin, Sean Tulin, and Daniel Whiteson. Mono-higgs-boson: A new collider probe of dark matter. *Physical Review D*, 89(7), apr 2014.

[40] David G. Cerdeño. Dark matter 101 from production to detection. 2018.

[41] Mikael Chala, Felix Kahlhoefer, Matthew McCullough, and et al. Constraining dark sectors with monojets and dijets. *Journal of High Energy Physics*, 2015(7):89, 2015.

[42] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical design report. CMS. CERN, Geneva, 2002.

[43] CMS Collaboration. Commissioning of the CMS high-level trigger with cosmic rays. *Journal of Instrumentation*, 5(03):T03005–T03005, mar 2010.

[44] CMS Collaboration. Differential cross section measurements for the production of a w boson in association with jets in proton-proton collisions at $\sqrt{s}$=7 tev. *CMS Physics Analysis Summary*, CMS-PAS-SMP-12-023, 2012.

[45] CMS Collaboration. Measurement of the production cross section for a w boson and two b jets in pp collisions at $\sqrt{s}$=7 tev. *CMS Physics Analysis Summary*, CMS-PAS-SMP-12-026, 2012.

[46] CMS Collaboration. Performance of τ lepton reconstruction and identification in CMS. *Journal of Instrumentation*, 7(01):P01001–P01001, jan 2012.

[47] CMS Collaboration. Measurement of associated w + charm production in pp collisions at $\sqrt{s}$ s=7 tev. *CMS Physics Analysis Summary*, CMS-PAS-SMP-13-149, 2013.

[48] CMS Collaboration. Measurements of jet multiplicity and differential production cross sections of z+jets events in proton-proton collisions at $\sqrt{s}$=7 tev. *CMS Physics Analysis Summary*, CERN-PH-EP-2014-205, 2014.

[49] CMS Collaboration. The Phase-2 Upgrade of the CMS Beam Radiation Instrumentation and Luminosity Detectors. Technical report, CERN, Geneva, 2021. This is the final version, approved by the LHCC.

[50] NNPDF collaboration and et al. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 2015(4):40, 2015.

[51] The CMS Collaboration. 34(6):995, apr 2007.

[52] The CMS Collaboration. 3(08):S08004, aug 2008.

[53] The CMS collaboration. Performance of CMS muon reconstruction in pp collision events at$\sqrt{s} = 7$TeV. *Journal of Instrumentation*, 7(10):P10002–P10002, oct 2012.

[54] The CMS collaboration. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(02):P02006–P02006, feb 2015.

[55] The MAJORANA Collaboration. The majorana experiment, 2011.

[56] S. Coutu, T. Anderson, C. Bower, J. Gennaro, M. Geske, D. Müller, J. Musser, S. Nutter, N.H. Park, M. Schubnell, G. Tarlé, S. Wakely, and A. Yagi. Searching for tev cosmic electrons with the crest experiment. *Nuclear Physics B - Proceedings Supplements*, 215(1):250–254, 2011. Proceedings of the 12th Topical Seminar on Innovative Particle and Radiation Detectors (IPRD10).

[57] W. de Boer, George Alverson, Pran Nath, and Brent Nelson. Indirect dark matter searches in the light of ATIC, FERMI, EGRET and PAMELA. In *AIP Conference Proceedings*. AIP, 2010.

[58] Wim de Boer and David B. Cline. Indirect dark matter signals from EGRET and PAMELA compared. In *AIP Conference Proceedings*. AIP, 2009.

[59] Hans Dembinski and Ahmed Abdelmotteleb. A new maximum-likelihood method for template fits. *European Physical Journal C*, 82(11):1043, November 2022.

[60] Abdelhak Djouadi, Oleg Lebedev, Yann Mambrini, and Jérémie Quevillon. Implications of lhc searches for higgs-portal dark matter. *Physics Letters B*, 709(1):65–69, 2012.

[61] Daniel Dominguez. Particles of the standard model of particle physics, 2015. Last accessed 25 April 2023.

[62] Fiorenza Donato. Indirect searches for dark matter. *Physics of the Dark Universe*, 4:41–43, 2014. DARK TAUP2013.

[63] Michael Duerr, Alexander Grohsjean, Felix Kahlhoefer, Bjoern Penning, Kai Schmidt-Hoberg, and Christian Schwanenberger. Hunting the dark higgs. *Journal of High Energy Physics*, 2017(4), apr 2017.

[64] Michael Duerr, Felix Kahlhoefer, Kai Schmidt-Hoberg, et al. How to save the wimp: Global analysis of a dark matter model with two s-channel mediators. *Journal of High Energy Physics*, 2016(9):42, 2016.

[65] M. Eidemüller, H.G. Dosch, and M. Jamin. The field strength correlator from qcd sum rules. *Nuclear Physics B - Proceedings Supplements*, 86(1):421–425, 2000. Proceedings of the QCD 99 Euroconference.

[66] Andreas Ekstedt, Rikard Enberg, Gunnar Ingelman, et al. Constraining minimal anomaly free u(1) extensions of the standard model. *Journal of High Energy Physics*, 2016(11):71, 2016.

[67] W Erdmann. Vertex reconstruction at the cms experiment. 110(9):092009, may 2008.

[68] A. Albert et al. Search for secluded dark matter towards the galactic centre with the ANTARES neutrino telescope. *Journal of Cosmology and Astroparticle Physics*, 2022(06):028, jun 2022.

[69] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, oct 2017.

[70] C. Bucci et al. First results from the cuoricino experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 520(1):132–134, 2004. Proceedings of the 10th International Workshop on Low Temperature Detectors.

[71] M. Aaboud et al. *Physical Review D*, 94(3), aug 2016.

[72] Pierluigi Bortignon et al. Particlenet cc-tagger calibration for ak15 jets using the g - cc method, 12 2020.

[73] S. Cebrián et al. First results of the igex dark matter experiment at the canfranc underground laboratory. *Nuclear Physics B - Proceedings Supplements*, 95(1):229–232, 2001. Proceedings of the 6th International Conference on Frontiers in Particle Astrophysics and Cosmology.

[74] S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.

[75] Sam Harper et al. E/gamma triggers in 2017, 11 2017.

[76] Sarah A. Malik et al. Interplay and characterization of dark matter searches at colliders and in direct detection experiments. *Physics of the Dark Universe*, 9-10:51–58, sep 2015.

[77] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008.

[78] Jan Eysermans. *Search for Additional Higgs Bosons with the CMS Detector at the CERN LHC*. PhD thesis, Puebla, México, 2019.

[79] Cristina Ferro. B-tagging in cms, 2012.

[80] Mads T. Frandsen, Felix Kahlhoefer, Anthony Preston, Subir Sarkar, and Kai Schmidt-Hoberg. LHC and tevatron bounds on the dark matter direct detection cross-section for vector mediators. *Journal of High Energy Physics*, 2012(7), jul 2012.

[81] Jennifer M. Gaskins. A review of indirect searches for particle dark matter. *Contemporary Physics*, 57(4):496–525, jun 2016.

[82] Stefano Giagu. WIMP Dark Matter Searches With the ATLAS Detector at the LHC. *Front. in Phys.*, 7:75, 2019.

[83] Lindsey Gray, Nicholas Smith, Mauro Verzetti, Joosep Pata, Anna Woodard, Pratyush Das, Nick Amin, Andreas, areinsvo, and Lukas. Coffeateam/coffea v0.6.6, July 2019.

[84] Higgs PAG. https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit.

[85] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert. Matching parton showers and matrix elements. In *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*, pages 288–289, 2005.

[86] Andreas Hoecker. Physics at the lhc run-2 and beyond, 2016.

[87] Nadège Iovine and Juan A. Aguilar. Indirect search for dark matter in the galactic centre with icecube, 2021.

[88] Thomas Jacques, Andrey Katz, Enrico Morgante, and et al. Complementarity of dm searches in a consistent simplified model: the case of z'. *Journal of High Energy Physics*, 2016(10):71, 2016.

[89] N. Jet rechurin et al. Search for dark matter produced with an energetic jet, or a hadronically decaying w or z boson, at $\sqrt{s} = 13$ tev with the full 2016 dataset. CMS Note 2016/473, 2016.

[90] JuiJen and Wang. Miniclean dark matter experiment, 2017.

[91] Felix Kahlhoefer, Kai Schmidt-Hoberg, Thomas Schwetz, et al. Implications of unitarity and gauge invariance for simplified dark matter models. *Journal of High Energy Physics*, 2016(2):16, 2016.

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

[93] Vyacheslav Klyukhin. Design and Description of the CMS Magnetic System Model. *Symmetry*, 13(6):1052, 2021. 17 pages, 9 figures, 67 references.

[94] Vyacheslav Klyukhin, Benoit Curé, Nicola Amapane, Austin Ball, Andrea Gaddi, Hubert Gerwig, Alain Hervé, Richard Loveless, and Martijn Mulders. Using the standard linear ramps of the cms superconducting magnet for measuring the magnetic flux density in the steel flux-return yoke. *IEEE Transactions on Magnetics*, 55(2):1–4, 2019.

[95] Mark G. Lawrence. The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bulletin of the American Meteorological Society*, 86(2):225–234, 2005.

[96] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[97] Mario De Leo. Galaxy rotation curve, 2018, November 14. Wikipedia.

[98] Heather E. Logan. Tasi 2013 lectures on higgs physics within and beyond the standard model, 2022.

[99] A. Metz and A. Vossen. Parton fragmentation functions. *Progress in Particle and Nuclear Physics*, 91:136–202, nov 2016.

[100] Remigius K Mommsen. The Data Acquisition System of the CMS Experiment at LHC. Technical report, CERN, Geneva, 2010.

[101] NASA SCIENCE. Discoveries - highlights shining a light on dark matter, 2021, Dec 22. [Online; accessed September 2, 2022].

[102] NASA SCIENCE. Dark energy, dark matter, 2022. [Online; accessed September 2, 2022].

[103] Zian Pan. The solution to galaxy rotation curve problem——dark matter and mond. *Journal of Physics: Conference Series*, 2220(1):012008, mar 2022.

[104] C. Patrignani. Review of particle physics. *Chinese Physics C*, 40(10):100001, oct 2016.

[105] M.I. et al Pedraza-Morales. RPC upgrade project for CMS Phase II. *JINST*, 15(05):C05072, 2020.

[106] Alexey A. Petrov and William Shepherd. Searching for dark matter at LHC with mono-higgs production. *Physics Letters B*, 730:178–183, mar 2014.

[107] CMS EGamma POG. Multivariate electron identification run2, 2023.

[108] CMS EGamma POG. Multivariate photon identification run2, 2023.

[109] CMS JetMET POG. Cms jetmet pog, 2023.

[110] CMS JetMET POG. Jec data mc, 2023.

[111] CMS JetMET POG. Jet resolution, 2023.

[112] CMS Muon POG. Swguidemuonidrun2: Loose muon, 2023.

[113] CMS Muon POG. Swguidemuonidrun2: Loose muon, 2023.

[114] CMS Tau POG. Tau id recommendation 13 tev, 2023.

[115] A. Riess. dark matter, 2022, February 3. Encyclopedia Britannica.

[116] Pierre Salati. Indirect dark matter searches in the light of the recent ams-02 observations, 2016.

[117] Pat Scott. Neutrino telescope searches for dark matter in the sun, 2017.

[118] Mariana Filipova Shopova. Performance of the CMS RPC system – from LS1 upgrade to RUN2 at LHC, 2018. Presented 12 Jun 2018.

[119] Albert M et al Sirunyan. Search for dark matter produced in association with a leptonically decaying Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81(1):13, 2021. Replaced with the published version. Added the journal reference and the DOI. All the figures and tables, including additional supplementary figures, can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/EXO-19-003 (CMS Public Pages).

[120] T. Sjöstrand and et al. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, 2015.

[121] Tracy R. Slatyer. Tasi lectures on indirect detection of dark matter, 2017.

[122] Neil J. C. Spooner. Direct dark matter searches. *Journal of the Physical Society of Japan*, 76(11):111016, nov 2007.

[123] Linda Stoel, Michael Barnes, Wolfgang Bartmann, F Burkart, Bethany Goddard, W Herr, Thomas Kramer, A Milanese, G Rumolo, Elena Shaposhnikova, and Geneva Switzerland. High energy booster options for a future circular collider at cern. 05 2016.

[124] The CMS experiment. Cms detector, 2022. [Online; accessed November 24, 2022].

[125] The CMS experiment. Cms hadronic calorimeter detector, 2022. [Online; accessed October 24, 2022].

[126] Nicolò Trevisani. Collider Searches for Dark Matter (ATLAS + CMS). *Universe*, 4(11):131, 2018.