



FACULTAD DE CIENCIAS  
UNIVERSIDAD DE CANTABRIA

---

**Search for dark matter production in  
association with top quarks in the  
dilepton final state at  $\sqrt{s} = 13$  TeV**

---

A thesis submitted in fulfillment of the requirements for the  
**Degree of Doctor of Philosophy**

---

Written by  
**Cédric Prieëls**

Under the supervision of  
**Jónatan Piedra Gómez**  
**Pablo Martínez Ruiz del Árbol**

Santander, June 2020





FACULTAD DE CIENCIAS  
UNIVERSIDAD DE CANTABRIA

---

**Búsqueda de materia oscura en  
asociación con quarks top en el estado  
final dileptónico a  $\sqrt{s} = 13$  TeV**

---

Memoria para optar al  
**Grado de doctor**

---

Escrita por  
**Cédric Prieëls**  
  
Bajo la supervisión de  
**Jónatan Piedra Gómez**  
**Pablo Martínez Ruiz del Árbol**

Santander, Junio 2020



# **Abstract**



# **Resumen**



## Acknowledgments



# Acronyms used

<b>ADMX</b>	Axion Dark Matter Experiment	<b>DY</b>	Drell-Yan
<b>ALICE</b>	A Large Ion Collider Experiment	<b>ECAL</b>	Electromagnetic Calorimeter
<b>AMS</b>	Alpha Magnetic Spectrometer	<b>EDM</b>	Event Data Model
<b>AOD</b>	Analysis Object Data	<b>EFT</b>	Effective Field Theory
<b>ATLAS</b>	A Toroidal LHC ApparatuS	<b>EWK</b>	Electroweak
<b>BDT</b>	Boosted Decision Trees	<b>FR</b>	Fake Rate
<b>BR</b>	Branching Ratio	<b>FSR</b>	Final State Radiation
<b>BSM</b>	Beyond the Standard Model	<b>GEM</b>	Gas Electron Multiplier
<b>BW</b>	Breit-Wigner	<b>GSF</b>	Gaussian Sum Filter
<b>CAST</b>	CERN Axion Solar Telescope	<b>HCAL</b>	Hadronic Calorimeter
<b>CERN</b>	European Council for Nuclear Research	<b>HLT</b>	High-Level Trigger
<b>CL</b>	Confidence Level	<b>HO</b>	Hadron Outer
<b>CMB</b>	Cosmic Microwave Background	<b>IACT</b>	Imaging Atmospheric Cherenkov Telescopes
<b>CMS</b>	Compact Muon Solenoid	<b>IAXO</b>	International AXion Observatory
<b>CSC</b>	Cathode Strip Chamber	<b>ISR</b>	Initial State Radiation
<b>CR</b>	Control Region	<b>KF</b>	Kalman Filter
<b>CSV</b>	Combined Secondary Vertex	<b>L1</b>	Level-1 Trigger
<b>CTA</b>	Cherenkov Telescope Array	<b>LAT</b>	Fermi Large Telescope
<b>DAQ</b>	Data Acquisition System	<b>LEP</b>	Large Electron Positron collider
<b>DAS</b>	Data Aggregation System	<b>LHC</b>	Large Hadron Collider
<b>DCS</b>	Detector Control System	<b>LNGS</b>	Laboratori Nazionali del Gran Sasso
<b>DQM</b>	Data Quality Monitoring	<b>LO</b>	Leading Order
<b>DM</b>	Dark Matter	<b>LS</b>	Long Shutdown
<b>DMWG</b>	Dark Matter Working Group	<b>MACHO</b>	Massive Compact Halo Object
<b>DNN</b>	Deep Neural Network	<b>MC</b>	Monte Carlo
<b>DT</b>	Drift tube	<b>MET</b>	Missing Transverse Energy

<b>MFV</b>	Minimal Flavour Violation	<b>RMS</b>	Root Mean Square
<b>ML</b>	Machine Learning	<b>RPC</b>	Resistive Plate Chamber
<b>MPI</b>	Multiple Parton Interaction	<b>SC</b>	Super Cluster
<b>MSSM</b>	Minimal Supersymmetric Standard Model	<b>SD</b>	Spin Dependent
<b>MVA</b>	Multi-Variate Analysis	<b>SF</b>	Scale Factors
<b>NFW</b>	Navarro-Frenk-White	<b>SI</b>	Spin Independent
<b>NLO</b>	Next to Leading Order	<b>SM</b>	Standard Model
<b>PDF</b>	Parton Density Function	<b>SPS</b>	Super Proton Synchrotron
<b>PF</b>	Particle Flow	<b>SR</b>	Signal Region
<b>POG</b>	Physics Object Group	<b>TEC</b>	Tracker EndCap
<b>PR</b>	Prompt Rate	<b>TIB/TBD</b>	Tracker Inner Barrel and Disks
<b>PS</b>	Proton Synchrotron	<b>TOB</b>	Tracker Outer Barrel
<b>PU</b>	Pile Up	<b>UE</b>	Underlying Event
<b>PUPPI</b>	Pileup Per Particle Identification	<b>UED</b>	Universal Extra Dimensions
<b>PV</b>	Primary Vertex	<b>VBF</b>	Vector Boson Fusion
<b>QCD</b>	Quantum ChromoDynamics	<b>WIMP</b>	Weakly Interactive Massive Particle
<b>QFT</b>	Quantum Field Theory	<b>WP</b>	Working Point

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Dark Matter case</b>	<b>5</b>
2.1	The Standard Model (SM) . . . . .	5
2.2	At the origins of dark matter . . . . .	6
2.2.1	Zwicky and the virial theorem . . . . .	7
2.2.2	Spiral galaxies rotation curves . . . . .	7
2.2.3	Cosmic Microwave Background (CMB) anisotropies . . . . .	8
2.2.4	Gravitational lensing . . . . .	10
2.3	Dark matter properties . . . . .	11
2.4	Dark matter candidates . . . . .	12
2.5	Dark matter searches . . . . .	19
2.5.1	Direct searches . . . . .	19
2.5.2	Indirect searches . . . . .	22
2.5.3	Collider production . . . . .	26
2.6	Dark Matter (DM) production at the Large Hadron Collider (LHC) . . . . .	29
2.6.1	The single top production channel . . . . .	31
2.6.2	The $t\bar{t}$ production channel . . . . .	32
2.6.3	The dilepton final state . . . . .	32
2.7	Previous relevant results . . . . .	33
<b>3</b>	<b>The experimental setup</b>	<b>37</b>
3.1	The LHC accelerator . . . . .	37
3.1.1	The LHC in a nutshell . . . . .	38
3.1.2	LHC key parameters . . . . .	39
3.2	The Compact Muon Solenoid (CMS) detector . . . . .	42
3.2.1	Tracker . . . . .	44
3.2.2	Electromagnetic Calorimeter (ECAL) . . . . .	46
3.2.3	Hadronic Calorimeter (HCAL) . . . . .	48
3.2.4	Solenoid . . . . .	50

3.2.5	Muon system . . . . .	51
3.2.6	Trigger system . . . . .	54
3.2.7	Data acquisition system . . . . .	56
<b>4</b>	<b>Objects reconstruction</b>	<b>57</b>
4.1	Particle Flow (PF) algorithm . . . . .	57
4.2	Primary Vertex (PV) definition . . . . .	58
4.3	Leptons reconstruction . . . . .	59
4.3.1	Muons . . . . .	59
4.3.2	Electrons . . . . .	61
4.4	Jets reconstruction . . . . .	63
4.4.1	B-tagging . . . . .	65
4.5	Missing Transverse Energy (MET) . . . . .	66
4.6	Top reconstruction . . . . .	68
4.6.1	Top reconstruction in practice . . . . .	69
<b>5</b>	<b>Data, signals and backgrounds</b>	<b>71</b>
5.1	The Monte Carlo (MC) simulation method . . . . .	71
5.2	Files format . . . . .	75
5.3	Analysis code . . . . .	76
5.4	Data samples . . . . .	76
5.5	Signal samples . . . . .	76
5.6	Backgrounds prediction . . . . .	76
5.6.1	Top production . . . . .	77
5.6.2	Drell-Yan estimation . . . . .	79
5.6.3	$t\bar{t} + V$ . . . . .	79
5.6.4	Non prompt contamination . . . . .	80
5.6.5	Smaller backgrounds . . . . .	85
5.6.6	Weights and corrections applied . . . . .	85
<b>6</b>	<b>Event selection</b>	<b>87</b>
6.1	Objects selection . . . . .	87
6.1.1	Triggers selection . . . . .	87
6.1.2	Electrons selection . . . . .	89
6.1.3	Muons selection . . . . .	90
6.1.4	Jet selection . . . . .	92
6.2	Signals regions . . . . .	93
6.2.1	$t/\bar{t}$ +DM region . . . . .	94

6.2.2	$t\bar{t}$ +DM region . . . . .	94
6.3	Control regions . . . . .	97
6.3.1	Drell-Yan (DY) control region . . . . .	97
6.3.2	Top control region . . . . .	97
6.3.3	$t\bar{t}V$ control region . . . . .	97
6.3.4	Same sign control region . . . . .	97
6.4	Background-signal discrimination . . . . .	102
6.4.1	Discriminating variables . . . . .	102
6.4.2	Neural network . . . . .	103
<b>7</b>	<b>Results and interpretations</b>	<b>105</b>
7.1	Systematics and uncertainties . . . . .	105
7.2	Results . . . . .	105
<b>8</b>	<b>Conclusions</b>	<b>107</b>
8.1	Future prospects . . . . .	107
<b>Appendices</b>		<b>109</b>
<b>A</b>	<b>Samples used</b>	<b>111</b>
A.1	Data samples . . . . .	111
A.2	Signal samples . . . . .	111
A.3	Backgrounds samples . . . . .	111
<b>B</b>	<b>Neural network optimization</b>	<b>119</b>
<b>Bibliography</b>		<b>127</b>



# Chapter 1

## Introduction

The SM of particle physics introduced in Section 2.1 is nowadays the most accepted mathematical model used to describe the elementary particles and three of the four fundamental forces of nature (electromagnetic, weak and strong interactions). This model is quite simple in concept, but has been able to describe most of the phenomena observed in nature so far with an incredible level of precision, and made a lot of predictions that have now been proven to be true, such as the postulate of the Higgs mechanism [1, 2] followed by the discovery of the Higgs boson itself [3, 4] by the CMS and A Toroidal LHC ApparatuS (ATLAS) experiments analyzing the proton-proton collisions produced by the Large Hadron Collider (LHC) at a center of mass energy  $\sqrt{s} = 13$  TeV, announced at the European Council for Nuclear Research (CERN) on the 4th of July 2012.

However, as accurate as it seems to be, this theory is known to have several shortcomings which require further investigation. Eventual exotic particles which do not fit in the current model could be the sign of new physics and have therefore been extensively searched for over the course of the last decades in order to enhance our understanding of the Universe and all its constituents.

In this context, the first serious DM hypothesis was introduced in the 1970s because of gravitational anomalies observed by several astrophysicists, as a way to explain the apparent non-luminous missing mass in the Universe [5]. Indeed, the visible mass in most galaxies appears to be way too low to explain several astrophysical processes, such as the rotation curves of the galaxies [6], which seems to be incompatible with the well established laws of gravitation. Some additional measurements of the gravitational lensing (in the Bullet Cluster, for example [7]) and the anisotropies observed in the CMB [8] are other evidences for the existence of DM, as explained in Section 2.2.

As far as we currently know from cosmological measurements, ordinary baryonic matter only constitutes around 5% of the Universe, while DM represents around 26% of the energy density of the Universe (the rest is being considered as dark energy) [9]. Understanding the nature and properties of this new kind of exotic matter is therefore crucial to try and understand the physics in the Universe, with many large experiments around the world currently involved in such searches.

Nowadays, the existence of DM is well established in the physics community, even though it has never been observed directly, since our only evidences so far for its existence come from its large-scale gravitational effects. While its mass, spin, nature and basic properties are still unknown and extensively studied, one of the best DM candidate is the so-called Weakly Interactive Massive Particle (WIMP), predicted to interact both gravitationally and weakly with SM particles. This would allow direct and indirect detection of such candidates, used as the driving process of many of experiments over the last decades, trying to find the hint of a possible interaction between standard baryonic particles and eventual DM particles, or even between several DM particles themselves. Dark matter production through the use of a particle accelerator colliding SM particles together,

such as the LHC, is also a possibility, and will be considered as the main channel towards the eventual detection of this exotic matter throughout this work. The production through colliders is actually able to provide constraints on low dark matter masses as well, in a region where both the direct and indirect searches are less efficient, which makes the LHC a perfect tool to study this kind of Beyond the Standard Model (BSM) physics. These searches will be summarized in Section 2.5.

However, observing DM is still extremely difficult, mainly because it barely interacts with ordinary baryonic matter, except through gravity (we have to assume that it does interact with SM at least weakly for the sake of this work though, as we would not be able to discover it as an individual particle if it were not the case). This means that nowadays, all the experiments searching for DM have only been able to put constraints on the DM particle mass and on the interaction cross sections between the dark and standard sectors. Actually, even if the collisions between protons produced by the LHC do have a sufficient amount of energy to produce this kind of particles, we would not expect them to interact with our detector, making their detection even harder. The eventual presence of such matter has to be inferred from the study of the interaction between SM particles and CMS itself, since a typical DM-like event consists of at least one energetic SM particle produced in association with a large imbalance in the transverse momentum due to the presence of an eventual DM candidate that was able to escape our detection.

In the context of this work, DM is searched for in association with one or two top quarks which play the role of the SM particle allowing us to trigger the event. This is indeed a perfect channel for this kind of searches if we assume that the interaction between the dark and standard sectors respect the principle of Minimal Flavour Violation (MFV), which can be consistently defined independently of the structure of the new physics model [10]. In this case, this interaction should follow the same Higgs-like Yukawa coupling structure as the usual SM baryonic particles, which actually depend on the mass: the heavier the SM particle considered is, the easier it is for it to couple with the dark sector. This makes the top quark, the most massive of all the fundamental particles observed by far, an excellent object to study in this context.

However, this also means that the phenomenology of this quark is mostly driven by its large mass and that it decays before hadronization can occur, usually into a W boson and a bottom quark. The final state of the process we are interested in is then made of some b jets, leptons and/or quarks and is categorized depending on this number of b-tagged jets and on the decay of the W itself. This work will actually be focused on the two leptons final state, also known as the dileptonic channel, mostly since this channel does not have lots of background processes raising to a similar final state, even though its branching ratio is the smallest, as will be explained in Section 2.6. Additionally, leptons are by reconstruction much cleaner than jets: their identification and momentum calculation is easier to perform, and the uncertainties associated to these measurements will be smaller.

The LHC has now been running for 10 years, and several similar searches have already been carried out and published in the past by the CMS and ATLAS collaborations, at different center of mass energies. First of all, at 8 TeV, several searches for a pair of top quarks were published by the CMS (in association with DM in the semileptonic [11] and dileptonic [12] final states) and ATLAS collaborations [13]. Then, at 13 TeV, the ATLAS collaboration published on one hand several studies, considering different final states and different luminosities ( $13.3 \text{ fb}^{-1}$  and  $36.1 \text{ fb}^{-1}$ ) [14, 15, 16]. On the other hand, the CMS collaboration published a few extremely important papers for this study [18, 19]. For the first time in 2019, the results obtained by the  $t/\bar{t}+\text{DM}$  and  $t\bar{t}+\text{DM}$  analyses have also been combined and published using the 2016 data [20]. Our main objective is now to repeat and improve this analysis while considering the full Run II dataset, globally improving the analysis strategy and including the dileptonic final state for the first time in this combination.

After a general introduction about DM in general in Chapter 2, the experimental setup will be detailed in the Chapter 3. This will include a discussion about the LHC itself, along with a complete description of CMS, the detector used to collect the data that will be analyzed throughout this work. This data has been collected during the years 2016, 2017 and 2018 and corresponds to an integrated luminosity of  $\sim 137 \text{ fb}^{-1}$ , collected during the Run II of operation of the LHC and at a center of mass energy  $\sqrt{s} = 13 \text{ TeV}$ . In particular, a particular care will be given to the explanation of the Particle Flow (PF) algorithm, used to reconstruct the different objects used and that will be defined in the Chapter 4, while the estimation of the different backgrounds and the selection of interesting events will be detailed throughout the Chapters 5 and 6.

Distinguishing between the signal we are searching for and backgrounds having a much higher cross-section and kinematically really close, such as the SM  $t\bar{t}$  without production of DM is not a straightforward task (sometimes a production of missing transverse energy due to the presence of physical neutrinos is even obtained). To isolate the signal and to obtain some discrimination between these kind of processes, an algebraic reconstruction of the event and top-notch Machine Learning (ML) techniques are used in this work, in order to train a network of neurons to perform this task. The main objective is to make them learn how to combine the discriminating power of a set of input variables in order to create a single output variable describing the probability of a single event to be classified as signal or background. All this process will be detailed in Section 6.4.

Finally, a statistical interpretation of our data will be performed and different sources of systematic uncertainties will be considered in Chapter 7. This will allow us to set upper limits on the cross section production value of DM particles in our particular channel and for the simplified models considered in this analysis. The conclusions of this work and some additional future prospects will then finally be presented in Chapter 8.



# Chapter 2

## The Dark Matter case

In this chapter, some general explanations about the Standard Model (SM) will first of all be given in Section 2.1 as a general introduction about today's most accepted mathematical model describing all the known particles and their interactions. The case for Dark Matter (DM) will then be presented in Section 2.2, along with a summary of the main evidences, mostly astrophysical, which lead to the introduction of this kind of Beyond the Standard Model (BSM) physics. Then, the main properties expected by such exotic matter will be introduced in Section 2.3 and nowadays's most accepted DM candidates, such as the Weakly Interactive Massive Particle (WIMP) briefly introduced previously will be presented in Section 2.4.

The main ways we have to search (direct, indirect and collider searches) for this new exotic physics along with the main experiments dedicated to such searches will then be shown in Section 2.5. Finally, the searches performed in colliders such as the Large Hadron Collider (LHC) and our particular channels of interest (DM produced in association with either one or two top quarks) will be detailed in Section 2.6 and the latest similar results and exclusion plots published over the course of the previous years by the ATLAS and CMS collaborations will be shown in Section 2.7.

### 2.1 The Standard Model (SM)

The Standard Model (SM) of particle physics is a relativistic Quantum Field Theory (QFT) theory which is able to mathematically summarize our current understanding of all the particles surrounding us and to describe three out of the four main interactions between these particles (mainly the weak, strong and electromagnetic forces, while an explanation of the origin of the gravitational interaction is still out of the reach of this theory nowadays).

If the neutrinos are considered to be normal Dirac fermions (and the question whether or not they could actually be Majorana particles is still well under discussion [22]), then the SM contains 26 free parameters, mainly the masses of the 12 predicted fermions, as shown in Figure 2.1, along with the couplings describing the strengths of the three interactions, two parameters describing the Higgs potential, eight mixing angles and, maybe, the phase of the eventual strong CP violation that we will mention in Section 2.4 (but in any case this parameter is expected to be close to 0). This is a quite high number of free parameters, but the value of most of them has now been derived experimentally, to match the observations made over the last decades.

The SM Lagrangian density in a differential volume element  $\mathcal{L} = \int L d^3x$ , accounting for the kinetic and potential energy of a system, takes the (very) simplified form given by Equation 2.1, where

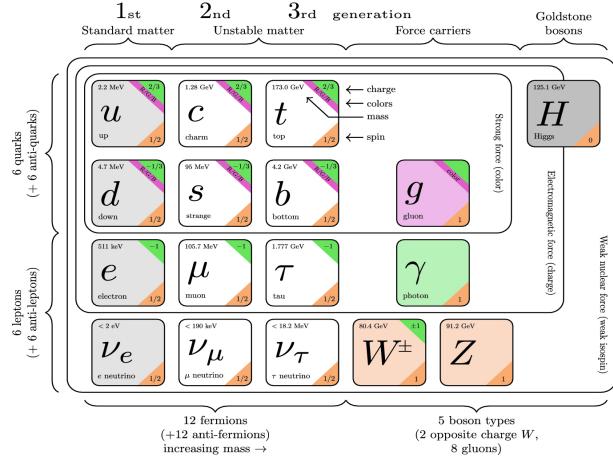


Figure 2.1: Representation of the 12 fermions of the SM [21].

the  $F_{\mu\nu}$  is the field strength tensor accounting for the different interactions,  $\psi$  is the interacting field describing quarks and leptons and  $\phi$  is the Brout-Englert-Higgs field while  $y_{ij}$  are the Yukawa couplings to this field, which depend on the mass of the particle considered and which will be described later on in Section 2.5.3.

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + \bar{\psi}_iy_{ij}\psi_j\phi + |D_\mu\phi|^2 - V(\phi) \quad (2.1)$$

A complete description of this model is unfortunately out of the scope of this work but it is important to note that the SM, although mostly experimental, is still working extremely well today. Indeed, it managed to make a lot of predictions over the years, which is the best you can obtain from a mathematical model, and most of its predictions revealed themselves to be true (we can quote for example that it successfully predicted the existence of the gluons, the top quarks, along with the  $W$ ,  $Z$  and Higgs bosons [23]).

However, even though it now appears to be complete since all its predictions have been discovered, we know that the SM is not the final theory of particle physics since it is not able to explain all the observations which have been made. There are still today some open questions and many Beyond the Standard Model (BSM) theories trying to explain such observations which do not fit inside the SM, such as an eventual inclusion of the gravitational interaction within this model. We can quote for example as successful BSM theories the possible existence of the supersymmetry, telling us that each particle should have a superpartner whose spin differs by  $1/2$  or the possible existence of Dark Matter (DM) particles, the main subject of the discussion of the following sections.

## 2.2 At the origins of dark matter

The origin of the concept of dark matter can be traced back to the 17 and 18th centuries, shortly after Newton's works on gravitation, even though this concept changed quite a lot over the years. Back then, DM was more considered to be ordinary matter which simply did not emit any kind of electromagnetic radiation, being therefore invisible and dark, but which does have a strong impact in the gravitational point of view because of its mass. It was for example considered in the 20th century to be found in massive astronomical objects able to absorb the light or other objects

located behind them, such as black holes (and even some actual theories still believe this is true!).

### 2.2.1 Zwicky and the virial theorem

In the 20th century, the first experimental evidences for the existence of dark matter were shown. In 1933, Fritz Zwicky managed to determine the mass of the Coma Cluster using the virial theorem [24], which states that in a cluster in equilibrium under its own gravitation the kinetic energy must be comparable to its gravitational binding energy.

Mathematically, the virial theorem can be written in Equation 2.2, where the brackets represent the mean value of the quantity obtained over time or position, the universal constants of gravitation is  $G = 6.67 \cdot 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$  and where the gravitation potential energy expression can be simplified assuming a spherical distribution of the masses and the same average density everywhere in the cluster considered.

$$2\langle T \rangle + \langle V \rangle = 0, \text{ where } \begin{cases} T = \frac{1}{2} \sum_i M_i v_i^2 = \frac{1}{2} M \langle v^2 \rangle \\ V = -4\pi G \int_0^R M \rho r dr \propto \frac{GM^2}{R} \end{cases} \quad (2.2)$$

Solving these simple equations gives us an approximate value of the mass of the cluster in Equation 2.3, where  $R$  is the radius of the cluster and  $\langle\langle v^2 \rangle\rangle$  is the squared velocity of all the galaxies averaged over position and time.

$$M \propto \frac{\langle\langle v^2 \rangle\rangle R}{G} \quad (2.3)$$

From this simple expression and astronomical observations, Zwicky then managed to compute the average mass to light ratio of its galaxies and concluded that the value obtained was around 400-500 times larger than the mass previously estimated by Edwin Hubble, who simply considered the number of visible galaxies within this cluster for his calculations. One plausible explanation for this discrepancy is to introduce the concept of DM, which contributes to the mass of the cluster without increasing the galactic luminosity.

Zwicky's results were actually quite controversial since they were based on statistical calculations relying on different hypotheses not always justified, such as the fact that the galaxies in the cluster must be gravitationally bound with each other and they were actually proven to be overestimated later on [25], but additional observations came to enforce to validity of his conclusions anyway.

### 2.2.2 Spiral galaxies rotation curves

Despite being controversial and slightly off, Zwicky's results were soon followed by a series of additional astronomical observations leading to the same conclusion, the possible existence of non-luminous matter in all the galaxies, called dark matter. The most famous of these results is the study of the observed and expected rotation curves of the stars within spiral galaxies such as the Milky Way in the 1970s [6].

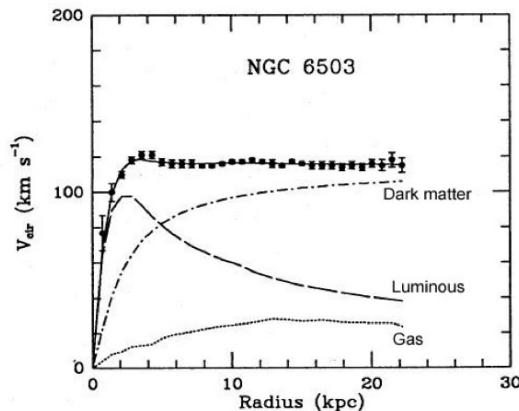
According to this study, if we assume that we can apply Newton's universal laws of gravitation at

the galactic scale, then the stars within this kind of galaxies should rotate with a velocity depending on the radius to the galactic center obtained by the usual equation for centripetal acceleration in a gravitational field and represented in Equation 2.4, where  $M(r)$  accounts for the total mass encountered in a radius  $r$ .

$$v_{\text{rotation}}(r) = \sqrt{\frac{GM(r)}{r}} \quad (2.4)$$

At first approximation, one can assume that most of the mass within this kind of galaxies comes from the inner core, which means that, at large radius, the velocity of individual stars is expected to decrease proportionally to  $r^{-1/2}$ . Any deviation to this rule suggest that either our understanding of gravity at large scales or our basic understanding of galaxies as a celestial body made of stars, dust and gas, has to be revised.

Actually, observations made by Vera Rubin and her team in the early 1970s with a new spectrograph designed to study the velocity curves of spiral galaxies with a degree of accuracy never reached before, did not confirm these expectations [26]. Indeed, according to these results, from a given value of the radius, the velocity curve appears to be flat instead of decreasing, as illustrated in Figure 2.2. This is another hint that can motivate the introduction of the concept of DM.



K.G. Begeman, A.H. Broels, R.H. Sanders. 1991. Mon.Not.RAS 249, 523.

Figure 2.2: Expected and observed rotation curves of the galaxy NGC 6503 [6]. The black dots correspond to the data and the *luminous* line corresponds to the expected rotation curve decreasing as  $r^{-1/2}$ , as expected from Newtonian dynamics.

### 2.2.3 Cosmic Microwave Background (CMB) anisotropies

The CMB is a mostly uniform background of primary radio waves emitted when the Universe became transparent around 380.000 years after the Big Bang and was discovered accidentally in the 1940s [27]. Studying it is extremely important, as it is actually made of the oldest and cleanest electromagnetic radiation we can find in the Universe. Precise measurements of this radiation are actually critical in many different fields of physics, since any proposed model of the Universe must be able to explain this radiation, its temperature and anisotropies.

Recent measurements determined that the CMB can be considered as emitting a thermal black body spectrum at a temperature of  $(2.72548 \pm 0.00057)\text{K}$  [28]. However, we now know that this temperature is actually not constant as some small anisotropies (at the  $10^{-5}$  level) depending on the value of the angular angle of observation can be observed, as represented on Figure 2.3.

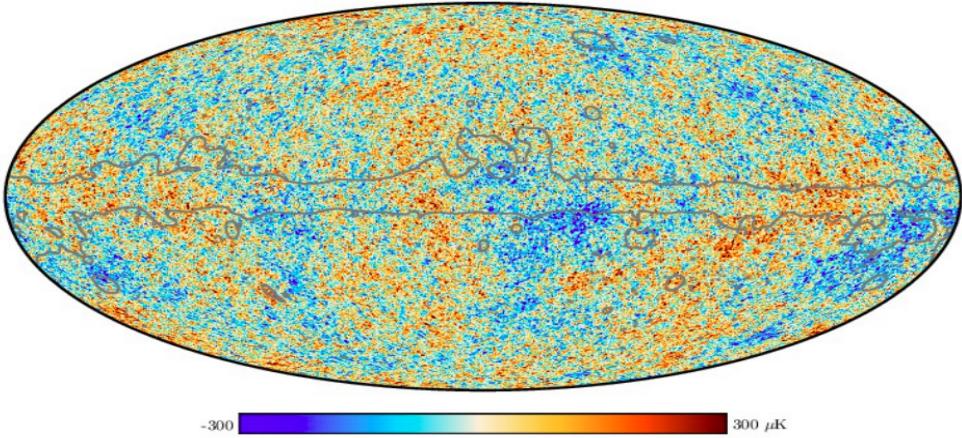


Figure 2.3: Anisotropies at the  $10^{-5}$  level in the temperature of the CMB, as observed by the Planck satellite in 2018 [29].

We see these fluctuations projected over a 2D sphere, and it is therefore natural to introduce at this point Laplace's spherical harmonics,  $Y_{lm}(\theta, \phi)$ , a complete set of orthogonal functions obtained by solving Laplace's equation  $\nabla\psi = 0$  on a sphere and defined by a few parameters such as  $l$ , the multipole, representing a given angular angle in the sky ( $l=100$  corresponds to  $\sim 1^\circ$ ) and  $m$ , the number of poles, such as  $-l \leq m \leq l$  [30].

It is possible to show that these spherical harmonics form a complete orthonormal basis on this space and therefore that any function can be defined on the sphere may be expanded into these harmonics using coefficients called  $a_{lm}$ . The temperature fluctuations, whose value depend on the two usual spherical angles  $\theta$  and  $\phi$  can therefore be expanded using these generic functions, according to Equation 2.5.

$$\frac{\Delta T}{T}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} a_{lm} Y_{lm}(\theta, \phi) \quad (2.5)$$

The information about the anisotropies can actually be extracted from the variance of these harmonic coefficients  $a_{lm}$  of the expansion since the CMB is assumed to be a gaussian random field. The power spectrum of the CMB can therefore be extracted according to Equation 2.6, and from which most of the cosmological information of the CMB is derived.

$$D_l = \frac{l(l-1)C_l}{2\pi} = \sum_m |a_{lm}^2| \quad (2.6)$$

Interestingly enough, this spectrum is directly affected by the value of the density of the dark matter in the Universe, and this is something that will be often discussed in the remaining of this chapter. Doing a multi-parameters fit on the observed data represented in the power spectrum (cf. Figure 2.4) is then able to give us directly the energy density of baryonic  $\Omega_b$  and dark  $\Omega_\chi$  matter, along with other important parameters of the  $\Lambda$ CDM cosmological model. Today's most precise measurements have been obtained in 2018 using the Plank satellite, and lead to the determination of these two quantities:  $\Omega_bh^2 = 0.02220 \pm 0.00020$  and  $\Omega_\chi h^2 = 0.1185 \pm 0.0015$  [31]. This is one of the strongest constraint we have on DM so far, since any DM candidate will need to comply with this measurement.

By dividing these results with the value of the scaling factor for Hubble expansion rate  $h = 0.674$  [32], we can obtain from these numbers a proportion of 4.9% of ordinary baryonic matter and 26.1% of dark matter in the Universe, as announced in the introduction.

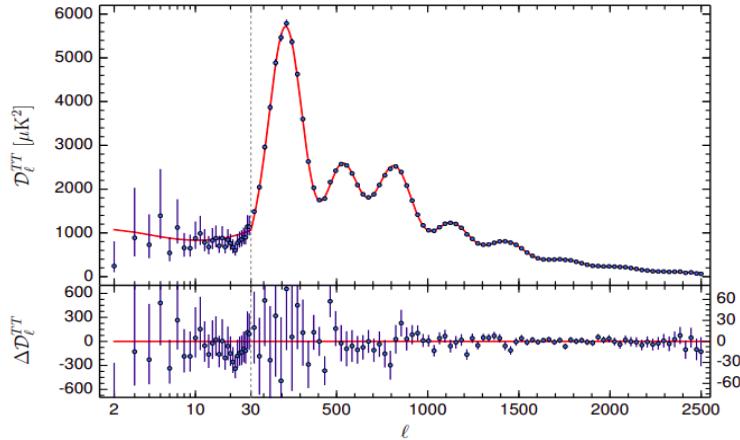


Figure 2.4: Power spectrum of the CMB obtained by Planck, representing the fluctuations of the temperature of the radiation with respect to the angular angle of observation [31].

#### 2.2.4 Gravitational lensing

The last evidence supporting the existence of dark matter has been obtained by observing several clusters of galaxies in the Universe, such as the Bullet Cluster, and by studying their mass distribution through gravitational lensing.

The gravitational lensing effect is a consequence of the general relativity, a theory developed by Einstein as a way to represent gravity using the geometry of spacetime, stating that massive objects lying between distant sources and an observer should act as a lens and bend the light emitted by the source. This deviation of the light is actually proportional to the mass of the object in between the source and the observer, meaning that the gravitational lensing can give us a way to measure the mass distribution of massive objects, such as galaxy clusters. This mass distribution can then be compared to the luminous distribution of the cluster, to see if we can observe a discrepancy between the two measurements, which could be another hint of the existence of DM.

The bullet Cluster is particularly interesting in this context since it actually provides an evidence for the eventual existence of DM which does not rely on any mathematical assumption (other than the general relativity principle) and cannot at principle be explained by alternate laws of gravitation. In this case, some observations clearly showed that the spatial deviations between the center of the total mass and the center of the baryonic luminous mass cannot be explained with an alteration of the gravitational force law alone, with a statistical significance of  $8\sigma$  [33].

As seen in Figure 2.5, the image taken by Chandra clearly shows an offset between the visible plasma of the cluster and the actual mass distribution measured through gravitational lensing (green contours). The center of the luminous mass of the cluster does not seem to match the one obtained considering its non-luminous counterpart as well, which is another evidence of the possible existence of DM within galaxy clusters.

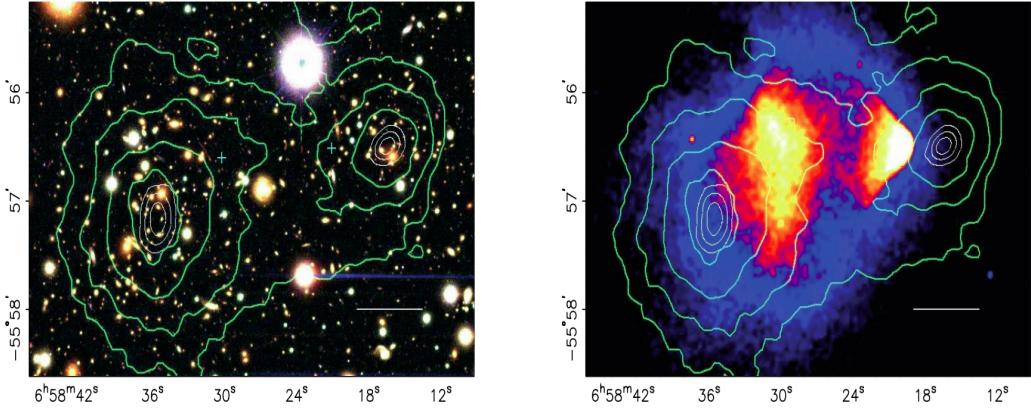


Figure 2.5: Mass distributions obtained by the Magellan in the visible (on the left) and Chandra on the X-rays spectrum (on the right) telescopes of the Bullet Cluster. Being shifted compared to each other, this is yet another clear evidence for the existence of DM [33].

## 2.3 Dark matter properties

All the previous observations allow us to list some of the most important properties that the perfect dark matter candidate should have. Even though several theories exist, each giving slightly different properties to the DM, we will consider in this work the following mostly accepted properties for such particles:

- First of all, we will assume that DM is a **particle**. As far as we know, the Universe is simply composed of particles so there is no objective reason to think that DM, being matter with a certain mass, might not be made of some kind of indivisible particles at some level.
- Then, the perfect DM candidate should of course be **dark**. This means that it should not interact at all in with electromagnetic radiation such as light, and that it should therefore be **electrically neutral**. However, it has to interact at least gravitationally because of the evidences for the evidence of such a particle explained before, mostly relying on gravitational effects, and we have to assume in this work that it interacts weakly as well to have a chance to discover it within particle accelerators.
- It has to be **non-baryonic**, mainly because the energy density for the baryonic matter obtained by observing the power spectrum of the CMB is too low to account for DM as well, as explained in Section 2.2.3. Indeed, according to these results, baryonic matter account for around 5% while dark matter accounts for more than 25% of the energy density of the Universe.
- We will also only consider **cold** dark matter, since the widely accepted  $\Lambda$  $CDM$  cosmological model is actually based on this assumption. By cold, we do not refer to the temperature of these particles but actually to their size, and therefore to the velocity by which they can travel in the Universe. Large scale structures of the Universe such as we can observe them today cannot actually not be explained if DM is made of hot/relativistic particles, as represented in Figure 2.6. However, although not really as popular, it is important to note that alternative models with warm DM have also been developed and still exist today.
- It is interesting to report as well that DM particles are expected to be found near the

electroweak symmetry breaking scale, between **10 GeV and 1 TeV**. This is a consequence of the expected production mechanism of such particles, the so-called thermal freeze-out.

This principle tells us that at some point in the history, DM was supposed to be in thermal equilibrium with other primordial SM particles, meaning that its production and annihilation rates were equal, as shown in Equation 2.7.

$$\chi\bar{\chi} \leftrightarrow e^+e^-, \mu^+\mu^-, q\bar{q}, W^+W^-, ZZ \quad (2.7)$$

However, because of the expansion of the Universe, at some point DM particles were simply too far apart from each other and these reactions maintaining this equilibrium were not efficient enough any more. At this stage, the abundance of DM became fixed: this is the freeze-out, as represented in Figure 2.7.

This principle is interesting because, as a rule of thumb, we can say that if a particle interacts heavily, it will stay longer in equilibrium and its freeze-out abundance will therefore be smaller, so there is a mathematical relation between the strength of the SM/DM interaction  $g$ , the mass of the DM particle  $m_\chi$  and its relic abundance  $\Omega_\chi$  that has been precisely measured, as expressed in Equation 2.8 [34].

$$\Omega_\chi \propto \frac{m_\chi^2}{g^4} \quad (2.8)$$

By using a typical value for  $g$  of the order of the Fermi coupling constant  $G_0^F \simeq 4.54 \cdot 10^{14} \text{ J}^{-2}$  we can see that, in order to observe a freeze-out abundance comparable to the one observed recently by the Planck satellite, the DM candidate should have a mass between 10 GeV and 1 TeV as previously stated. The measurement of the CMS power spectrum is therefore able to put constraints on the DM cross-section with the baryonic sector, and all the DM candidates quoted in Section 2.4 will have to respect this criteria.

In any case, DM is not expected to have a mass lower than 300 eV since at this scale, the phase space density that would be needed to explain this relic abundance of DM would simply violate the Pauli-exclusion principle [37].

- Finally, the DM particles should be **long-lived**. Indeed, we expect that they were produced 13.8 billion years ago during the Big Bang, but it seems they are still present in the Universe since we still see their effect today. They should then be stable particles, or their lifetime should at least be larger than the age of the Universe itself.

All these properties narrow quite a lot the number of possible DM candidates, as we will now see in the following section.

## 2.4 Dark matter candidates

Several different categories of particles could pretend to be good candidates for dark matter but only the most interesting ones will be quoted here, since an extensive list of all the different possible candidates is out of the scope of this work. Two SM particles will first of all be investigating, before introducing some BSM theories giving us additional DM candidates with the expected properties.

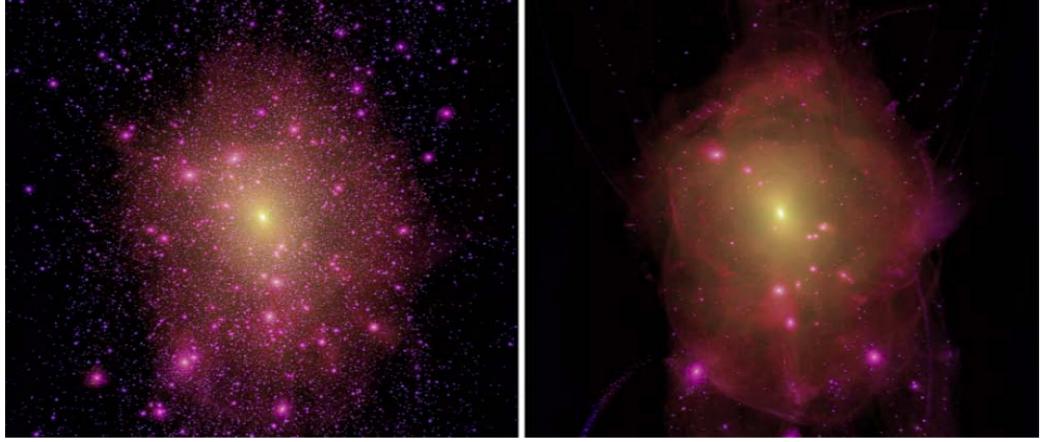


Figure 2.6: Computer simulations for cold (on the left) and warm (on the right) DM scenarios and their impact on a galactic halo at 0 redshift [35].

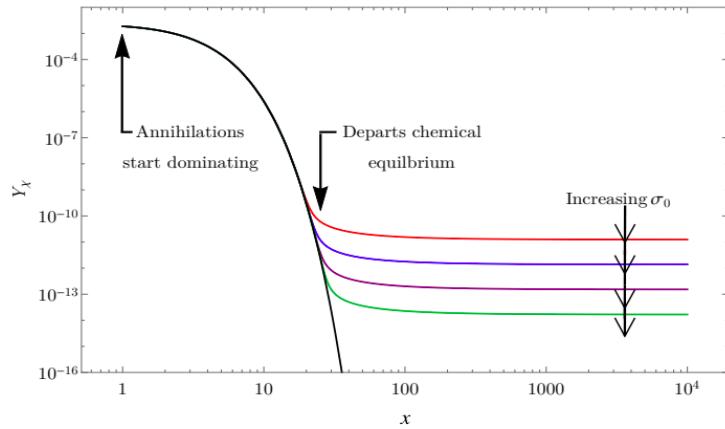


Figure 2.7: Schematic representation of the freeze-out process, representing the abundance of a 500 GeV DM as  $Y_\chi$  with respect to the time and the impact of increasing cross-section annihilation values on this freeze-out abundance [36].

### Massive Compact Halo Objects (MACHOs)

The first obvious DM candidate are the so-called MACHOs. These objects are massive astronomical non-luminous bodies (such as black holes) made of baryonic matter and very hard to detect, that could be responsible of the gravitational lensing observed and that could explain the apparent missing mass in the Universe. However, as we saw in Section 2.3, DM is not expected to be made of such ordinary matter, mainly because observational data of the CMB and the deduced baryonic density of energy in the Universe is able to rule out this possibility.

Several different experiments did try to search for such DM anyway, and managed to constrain the properties of this kind of objects. The main way to search for such massive objects is through their gravitational lensing effect (actually, we talk about microlensing in this case since this effect is small) since, according to the general relativity principle, they should bend the light of luminous objects located behind them, such as stars, and this bending actually depends on the mass of the eventual MACHO. Experiments like the MACHO project and EROS observed in this context  $\Theta(10^7)$  stars for several years, looking for microlensing events in order to constrain this particular

DM model. Results published in 2000 from the MACHO project, after studying almost 12 million stars, actually observed between 13 and 17 such events, lower than expected if DM was only made of MACHOs.

The collaboration actually managed to exclude at the 95% Confidence Level (CL) the possibility of the dark halo to be entirely made of such baryonic particles [38]. On the other hand, the EROS collaboration only observed 1 microlensing after studying more than 30 millions of bright stars during 6.7 years, while  $\sim 39$  events were expected [39]. Both results have also been combined in order to obtain the exclusion plot represented in Figure 2.8. From this study, MACHOs with low masses ( $10^{-7}M_{\odot} < m < 10^{-3}M_{\odot}$ ) should make up less than 25% of the dark matter halo for most models considered at the 95% CL [40].

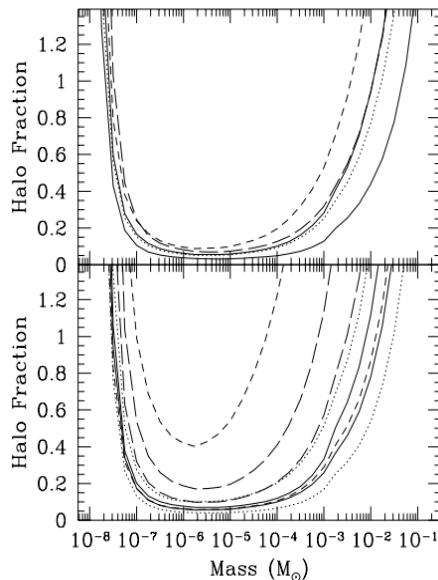


Figure 2.8: Halo fraction upper limit at the 95% CL compared to the mass of the lensing object for different MACHO models considered by the EROS (on the top) and the MACHO (on the bottom) collaborations [40].

## Active neutrinos

SM *active* neutrinos  $\nu$  (as opposed to *sterile* neutrinos, that will be the subject of the discussion in the next section) have been considered as good DM candidates for a long time as well, since they are electrically neutral and long-lived SM particles, two important properties of any DM candidate. They have a few particular properties that might be interesting in this context.

- First of all, even though it has still not been measured precisely, the sum of their mass has recently been measured to be lower than 0.17 eV at the 95% CL from cosmological studies [44]. Even though this is not quite understood, this is incredibly low compared to other SM particles, this particularity usually being referred to as the *mass puzzle* of the neutrinos.
- Their low mass has a consequence in the sense that it means that the gravitational interaction between two neutrinos is usually considered to be negligible and that we can consider that they only interact weakly, making it hard to study their properties. Their actual cross-section of interaction, as represented in Figure 2.9 and which of course depends on their

energies and on the channel of interaction (neutral or charged current considered), is therefore usually extremely low, making it hard to detect neutrinos.

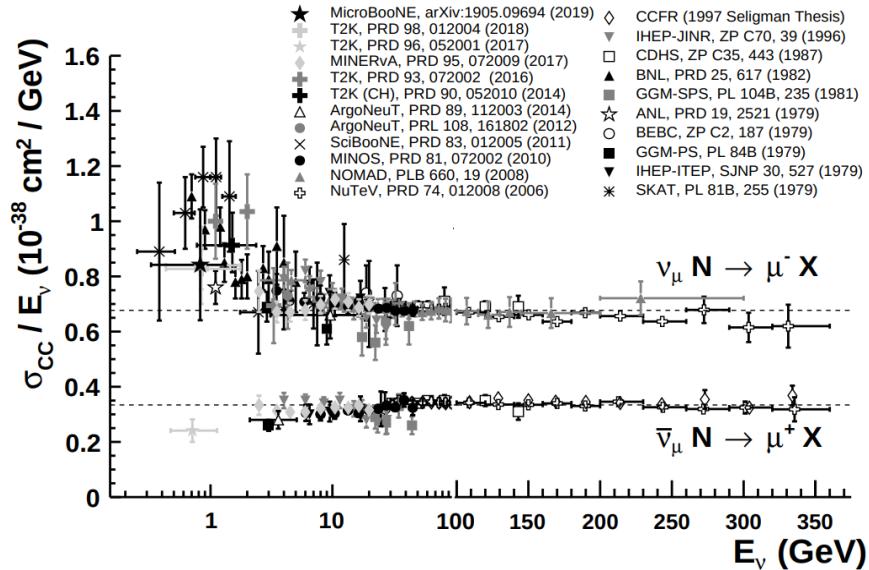


Figure 2.9: Neutrino cross section of interaction from the charged current as measured by different experiments over a large range of energies, for both neutrinos  $\nu$  and antineutrinos  $\bar{\nu}$  [41].

This figure clearly shows that an approximate value of the cross section for such neutrinos is of the order of magnitude of  $10^{-38} \text{ cm}^2 \text{ GeV}^{-1}$ , which is typically tens of orders of magnitude lower than the interaction cross section of a photon ( $\sigma_\gamma \sim 10^{-25} \text{ cm}^2$  [42]). This means that the typical neutrinos of a few MeV produced by nuclear reactors have a mean free path of approximately 10 light years in steel.

- Neutrinos are also the only SM particle only observed in their left-handed chirality state, while anti-neutrinos can only be observed in their right-hand state. This could mean two things: either right-handed neutrinos do not exist in nature for some reason, or we have just not been able to detect them because their interaction with baryonic matter is too weak. Right-handed neutrinos, also referred to as *sterile* neutrinos, do not fit in the current SM but could actually also be a strong BSM DM candidate, as we will discuss in the next subsection.
- Finally, neutrinos can oscillate, a quantum phenomena according to which the flavor of a neutrino can spontaneously change with time. This means that the three interacting states  $\nu_\alpha$  observed are actually composed of several mass eigenstates  $\nu_i$ , as related in Equation 2.9, where these states can be related using the Pontecorvo-Maki-Nakagawa-Sakata matrix ( $V_\nu$ ) $_{\alpha i}$ .

$$\nu_\alpha = (V_\nu)_{\alpha i} \nu_i \quad (2.9)$$

This effect does not however at principle have anything to do with the fact that neutrinos can be considered DM candidates, so this subject will not be discussed further in this work.

However, two physical reasons can explain why we do not really believe that DM could be made of neutrinos any more. First of all, their relative abundance does not match the expected one for DM from the freeze-out mechanism explained in Section 2.3. Indeed, their freeze-out abundance can be computed quite easily from Equation 2.10 [43], where the sum of the masses of the three

neutrino flavors has been calculated to be lower than 0.17 eV [44] instead of the 11.5 eV expected to obtain the correct DM relic abundance as observed today from the power spectrum of the CMB.

$$\Omega_\nu h^2 = \sum_{i=1}^3 \frac{m_{\nu_i}}{93 \text{ eV}} \quad (2.10)$$

Additionally, for several reasons explained in Section 2.3, a good DM candidate is expect to be cold, i.e. non-relativistic. However, being extremely light, neutrinos are expected to be ultra-relativistic and could therefore not be responsible of the emergence of large scale structures as observed today. We can therefore most probably rule out the possibility of DM being made of SM neutrinos.

### Sterile neutrinos

The most obvious SM particles being rules out as DM candidates, it is now time to introduce some of the most famous BSM theories introducing additional particles that could have the properties searched for. The first one of these theories introduce the so-called sterile neutrinos, usually referring to right-handed SM neutrinos, as discussed in the previous subsection.

If they exist, sterile neutrinos are expected to interact in an even weaker way than SM active neutrinos, they could be very long-lived as well and in principle, nothing prevents us from considering that they could have a mass superior to 0.4 keV, giving therefore the correct DM relic abundance [37]. A superior bound of 50 keV on such particles can also be obtained considering limits on the observation of the monochromatic decay  $\gamma$  line originating from the one loop radiative decay  $N \rightarrow \nu + \gamma$  of such particles.

Several experiments are already searching for such particles at this level of energy. Most of these experiments focus on the analysis of  $\gamma$ -rays and are actually searching for this particular monochromatic line resulting of the decay of sterile neutrinos. Two independent groups actually announced in 2014 the observation of an unidentified emission line at 3.57 keV (Figure 2.10) which did not match any known atomic emission line and which is actually consistent with an eventual DM signal [45, 46], since most of the possible instrumental contamination effects have been ruled out over the course of the last few years.

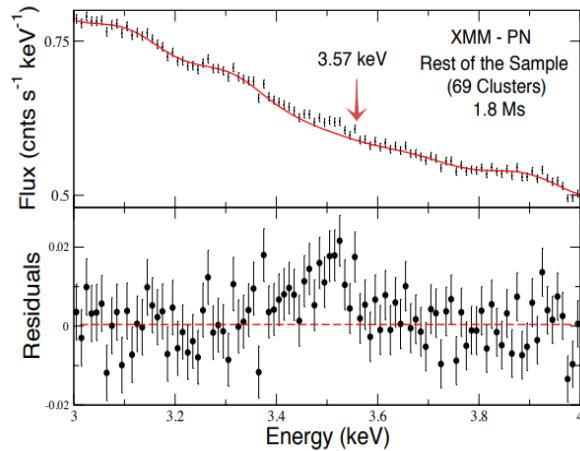


Figure 2.10: 3.57 keV emission line detected with a  $4.5\sigma$  CL by the XMM-Newton telescope in 2014, which could be a hint of the presence of DM [45].

However, some additional studies of the galactic center pointed out the fact that this observation might actually come from the observation of a potassium K XVIII transition line [47]. Recent observations actually ruled out at the 99% CL an eventual DM origin for this particular line [48], but further studies are still ongoing.

## Axions

Axions could also explain the particle nature of DM, since their existence could explain 100% of the DM in the Universe, unlike most of the other candidates presented so far. Axions are hypothetical stable neutral particles, with masses of the order of the meV, introduced as a consequence of the strong-CP violation issue of Quantum ChromoDynamics (QCD). This issue is the following: the usual  $\Theta$  term of the QCD Lagrangian shown in Equation 2.11 [49] should be responsible of breaking the CP symmetry, but this effect has actually never been observed so far: this is the so-called the strong CP problem.

Two ways to explain that we never observed this phenomena exist: the first is to assume that one of the quarks of the SM is massless but this does not match the current observations and measurements. The second consists in assuming that the parameter  $\Theta$ , the QCD vacuum angle, is small enough so that this term becomes negligible. However, by definition, the  $\Theta$  angle should be between 0 and  $2\pi$ , so there is no physical reason for this parameter to be small, unless some new physics can be introduced, as the one developed in 1977 by Peccei and Quinn [50] by relaxing  $\Theta$  from a parameter to a dynamic variable and absorbing it through the introduction of a new pseudoscalar particle that was called the axion.

$$\mathcal{L}_\Theta = \frac{\Theta}{32\pi^2} \epsilon_{\mu\nu\rho\sigma} G_a^{\mu\mu} G_a^{\rho\sigma} \quad (2.11)$$

By definition, it is possible to show that axions satisfy two of the previous criteria for a good DM candidate, since they are non-relativistic and their abundance might be enough to account for the dark matter energy density observed, since their actual abundance can easily be computed from Equation 2.12 [51], from which we could conclude that an axion having a mass of  $\sim 20$   $\mu\text{eV}$  could account for the DM relic density of the Universe, as observed today.

$$\Omega_a \simeq \left( \frac{6\mu\text{eV}}{m_a} \right)^{7/6} \quad (2.12)$$

Several axions searches experiments have therefore been set up, such as the Axion Dark Matter Experiment (ADMX), a resonant microwave cavity installed at the University of Washington, CERN Axion Solar Telescope (CAST), a CERN experiment observing the Sun which came online in 2002 and which managed in 2014 to turned up definitely the existence of solar axions [52], or the International AXion Observatory (IAXO), whose aim would be to search for axions with a much better signal to ratio noise than CAST. All the results obtained by these experiments along with their future projections are represented in Figure 2.11.

## Weakly Interactive Massive Particles (WIMPs)

The actual DM candidate that will be mostly considered throughout this work are the so-called Weakly Interactive Massive Particles (WIMPs), which are expected to interact, even though very

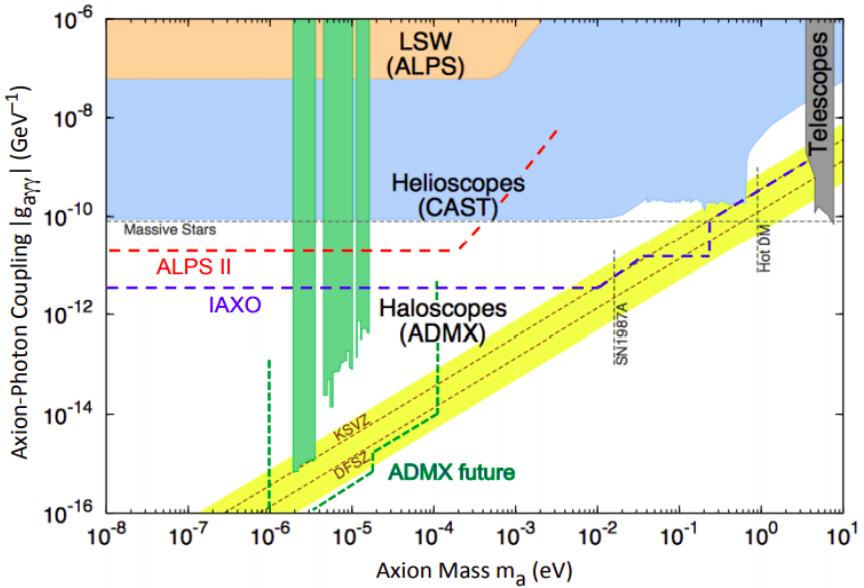


Figure 2.11: Axions exclusion plot summary and projected coverage of axion searches experiments, such as ADMX, CAST and IAXO [51].

weakly, with ordinary baryonic matter and which have an expected mass in the range of 100 GeV to 1 TeV for reasonable electroweak production cross-section values, right where we expect DM to be found from its relic density: this is the so-called "WIMP miracle", an important concept that can be translated mathematically as well. Indeed, because of the freeze-out scenario explained in Section 2.3, we can find an expression relating the relic abundance of DM  $\Omega_\chi$  with its annihilation cross section  $\langle\sigma_A v\rangle$  through Equation 2.13 [59].

$$\Omega_\chi h^2 \sim \frac{3 \cdot 10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle\sigma_A v\rangle} \quad (2.13)$$

This equation implies that, since we do know the current abundance of DM in the Universe, the total annihilation cross section of DM should be equal to  $\sim 3 \cdot 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ , which corresponds to the typical value given by WIMPs for a range of dark matter masses matching the expected one.

Several strategies can be used to detect such particles, as we will now see in the Section 2.5. This kind of particle basically arises in various BSM theories, such as the lightest supersymmetric particle in SUSY. According to this theory, each SM particle should have a superpartner whose quantum numbers would be identical except for their spins, which would differ by one half. All of these superpartners would then be potentially new and undiscovered particles, giving us a perfect DM candidate in most of the Minimal Supersymmetric Standard Model (MSSM) theories, the neutralino  $\chi$ .

The WIMPs are also interesting in the sense that introducing them in the terms of this supersymmetric theory would not only give us a strong DM candidate, but would also solve the hierarchy problem, the apparent large discrepancy between multiple aspects of the weak and gravitational forces ( $10^{24}$  less times stronger).

## 2.5 Dark matter searches

As previously stated, several cosmological evidences allow us to introduce the concept of dark matter, but its properties such as its mass, coupling and interaction cross-section are difficult to study in this context. Several different ways can then be used in order to try and detect DM particles in order to study them, as represented in Figure 2.12, strategies which can usually be divided into three categories: the direct and indirect searches, mostly relying on the production of baryonic matter from the interaction between two DM particles or on the observation on the interaction between the dark and baryonic sectors, and the production in colliders, usually able to probe lower DM candidates masses and which will actually be the main focus of this work.

A discussion about these different detection strategies along with the main results obtained by different experiments in each case will now be presented.

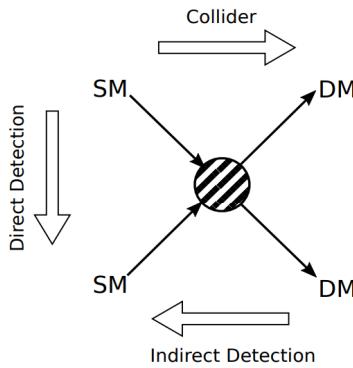


Figure 2.12: Schematic view of the three main DM detection strategies: direct, indirect and collider production searches [53].

### 2.5.1 Direct searches

From cosmological observations, we know that we live in a halo of dark matter. In this case, WIMPs should cross the Earth every day and, even if they interact only weakly, we should be able to directly detect them through their interaction with ordinary baryonic matter, for example because of their scattering with the nuclei of these particles. Indeed, the transfer of momentum between these two particles in this case might be detectable with the correct experimental device, typically placed deep underground to have the lowest possible background, which is the main source of perturbations of such experiments.

To study this particular category of searches, let's first of all introduce the rate of expected WIMP scattering off a target nucleus of mass  $m_N$  with Equation 2.14, rate which ends up being described by a simple steeply falling exponential function as shown in Figure 2.13 [54], where  $E_{nr}$  is the nuclear recoil energy measured,  $m_\chi$  is the WIMP mass,  $\sigma$  its cross section,  $\rho_0 = 0.3 \text{ GeV cm}^{-3}$  is the local dark matter density and  $f(v)$  the normalized WIMP velocity distribution.

$$\frac{dR}{dE_{nr}} = \frac{\rho_0 M}{m_N m_\chi} \int_{v_{\min}}^{v_{\text{esc}}} v f(v) \frac{d\sigma}{dE_{nr}} dv \propto \exp\left(-\frac{E_{nr}}{E_0} \frac{4m_\chi m_N}{(m_\chi + m_N)^2}\right) \quad (2.14)$$

From this relation, the Equation 2.15 can be easily derived, representing this time the number of

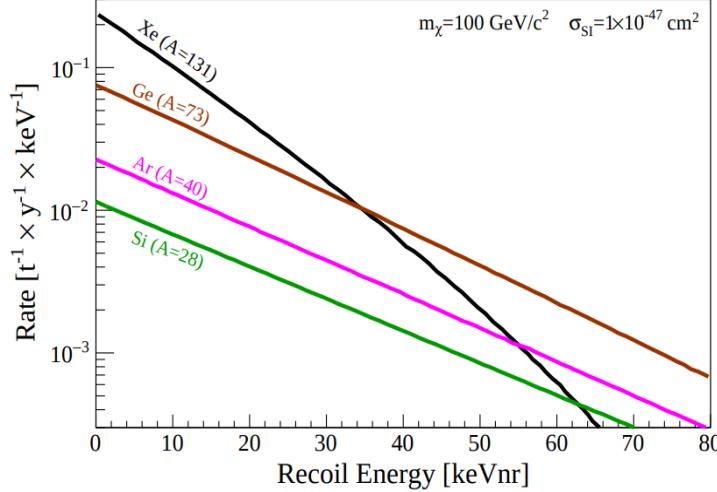


Figure 2.13: Nuclear recoil spectra induced in different materials for a given DM WIMP of 100 GeV, assuming a WIMP-nucleon Spin Independent (SI) cross section [54].

expected DM events in an experiment running during a time  $T$ , where  $\epsilon(E_{nr})$  is the efficiency of the detector for a given recoil energy.

$$N = T \int_{E_{\min}}^{E_{\max}} \epsilon(E_{nr}) \frac{dR}{dE_{nr}} dE_{nr} \quad (2.15)$$

The maximal velocity  $V_{\text{esc}}$  used as superior bound of the integral in Equation 2.14 has actually been measured to be in the range [498 – 608] km/s at the 90% CL [55], since any particle having a velocity higher than this would not be bound any more to the gravitational potential of a galaxy. This has an important consequence: all the direct and indirect detection experiments actually need to take into account is the annual modulation of the observed count rate, due to the movement of the Earth around the Sun, as shown in Figure 2.14 [56], since this velocity is not negligible compared to the escape velocity  $v_{\text{esc}}$ .

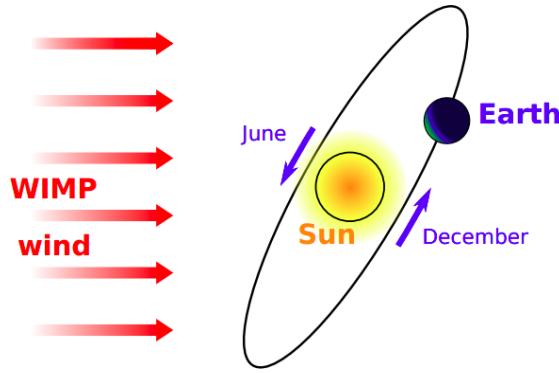


Figure 2.14: Schematic representation of the annual modulation of the WIMP wind introduced by the motion of rotation of the Earth around the Sun [56].

From our perspective, it seems indeed that the velocity of the speed of WIMP particles arriving

is changing depending on the month of the year, since the Earth is sometimes moving in the direction of the WIMP source, and is sometimes moving away: the maximal velocity is reached around June. This is extremely important to take into account this effect since, as we saw on the previous equations such as Equation 2.13, the count rate of incoming particles  $N(t)$  actually depends on this velocity, and this modulation then introduces a periodical modulation that we need to take into account, as shown in Equation 2.16, where the periodical part usually introduces a  $\sim 5\%$  deviation [54].

$$N(t) = B + N_0 + N_m \cos(\omega(t - t_0)) \quad (2.16)$$

This effect is also important because an experiment performed during a long period of time can actually help us finding an eventual hint of DM particles, since our signal is expected to follow this periodical deviation while the background is expected to be constant. Moreover, this WIMP wind is expected to come from a particular region of the sky while the backgrounds are expected to be distributed uniformly, so this gives a clear way to isolate the signal.

Finally, it is also important to note that two different kinds of direct searches can be defined, depending on the category of the scattering between the DM and the nucleus: the Spin Independent (SI) (proceeding through the scalar term) and Spin Dependent (SD) (proceeding through the axial term of the Lagrangian) searches, since the interaction cross section  $\sigma$  of Equation 2.14 is expected to be different for DM particles having a spin 0 or not, as shown in Equation 2.17, where  $F$  is a factor accounting for the dependence of the scattering on the energy. This means that results obtained by either hypothesis can usually not be compared with each other.

$$\frac{d\sigma}{dE_{nr}} \propto \sigma_{SI} F_{SI}^2(E_{nr}) + \sigma_{SD} F_{SD}^2(E_{nr}) \quad (2.17)$$

As previously stated, many experiments are dedicated to the direct search of DM particles, but in order to isolate an eventual DM signal, an environment with an ultra-low background is usually required, which can usually be reduced either by placing the detector underground (to reduce the contamination due to the cosmic rays), by increasing the statistics or by choosing carefully the active material of the detector (to reduce the internal background coming from the detector itself). The impact these kind of parameters have on the final limits can be seen in Figure 2.15.

These detectors try to detect the scattering of an unknown exotic particle with an ordinary nucleus, which typically can give rise to different categories of signals. Some detectors try for example to detect the direct ionization of the target atom, while others focus on the emission of light coming from the desexcitation of the scattered nucleus, and some even search for the heat produced by these collisions under the form of phonons in a crystal. All these different search strategies have been summarized schematically in Figure 2.16.

As of today, no direct experiment has been able to detect serious hints for the existence of DM, and they have only been able to set limit on the scattering cross section depending on the models parameters, as seen on Figure 2.17 for multiple experiments at once.

However, the DAMA experiment at the Laboratori Nazionali del Gran Sasso (LNGS) did find an interesting result by showing the hints of an annual modulation signal compatible to the expected one due to the WIMP wind in the 2-6 keV energy range, as seen on Figure 2.18 [58]. Further investigations about this modulation are still ongoing today, since no systematic effect able to account for the observed modulation amplitude and to simultaneously satisfy all the requirements of the signature has been found so far.

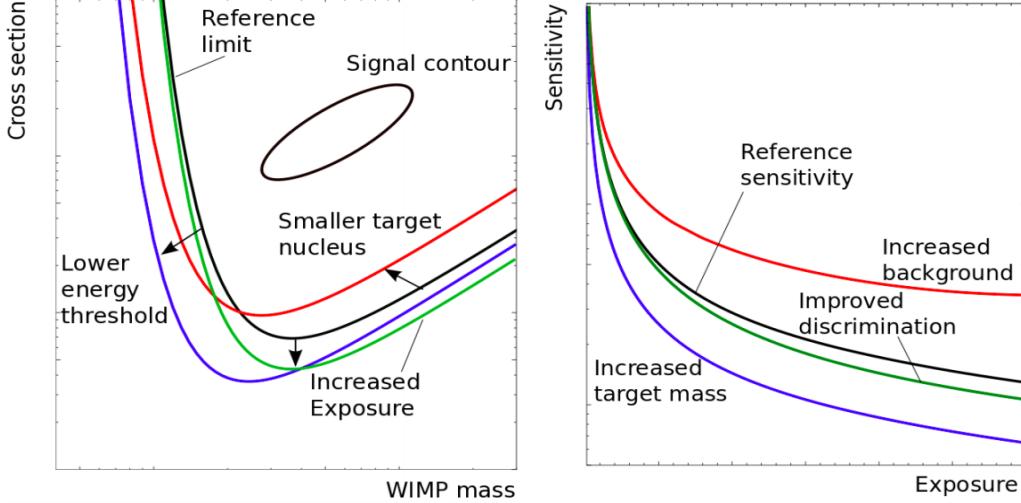


Figure 2.15: Impact of different experimental parameters on the final limits depending on the cross section and WIMP mass (on the left) or sensitivity and exposure (on the right), with respect to the expected limits (black curve) [57].

### 2.5.2 Indirect searches

The indirect detection of DM particles consists basically in searching for SM products coming from the annihilation of two DM particles or from its eventual direct decay, usually under the form of a flux of  $\gamma$ -rays, neutrinos, cosmic-rays or anti-matter appearing as an excess over the expected background. Indeed, many extensions to the SM, such as the supersymmetry SUSY or Universal Extra Dimensions (UED) do provide solid DM candidates (the lightest supersymmetric particle and the lightest Kaluza-Klein state, respectively) expected to interact with each other, resulting in the immediate production of SM (un)stable particles that can be detected by telescopes either placed on the ground or directly in space. Another point to make is that indirect searches are also usually affected by the annual fluctuation induced by the movement of the Earth around the Sun, as explained in the previous Section 2.5.1.

Indirect searches are actually extremely useful since they are sensitive to the DM annihilation cross section, mass and the density profile of DM halos  $\rho(\vec{r}(s, \Omega))$ , usually represented by a Navarro-Frenk-White (NFW) profile, as shown in Equation 2.18, where  $r_s = 20$  kpc is the scale radius of the DM halo for the Milky Way and  $r$  is the distance to the center of the cluster considered, assumed to be spherical in this case [60].

$$\rho(r) \propto \frac{r_s}{r \cdot \left(1 + \left(\frac{r}{r_s}\right)^2\right)} \quad (2.18)$$

The flux coming from the annihilation of two DM particles is then expected to be proportional to its annihilation cross section  $\sigma v$ , the solid angle of observation  $\Omega$  and the number of particles emitted by this annihilation  $\frac{dN}{dE}$ , according to Equation 2.19, where the integration is done over the line of sight  $l$  and the solid angle of observation  $\Omega$ .

$$\frac{d\Phi}{d\Omega dE} = \frac{\sigma v}{8\pi m_\chi^2} \cdot \frac{dN}{dE} \iint_{l, \Omega} \rho^2(\vec{r}(s, \Omega)) dl d\Omega \equiv P \cdot J(\Delta\Omega) \quad (2.19)$$

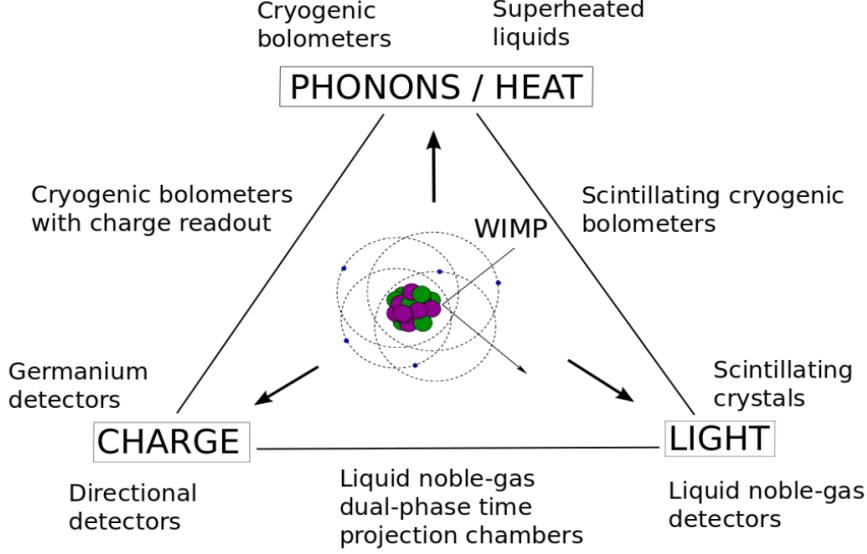


Figure 2.16: Schematic representation of the three main strategies to detect directly the interaction between DM particles and an ordinary nucleus [57].

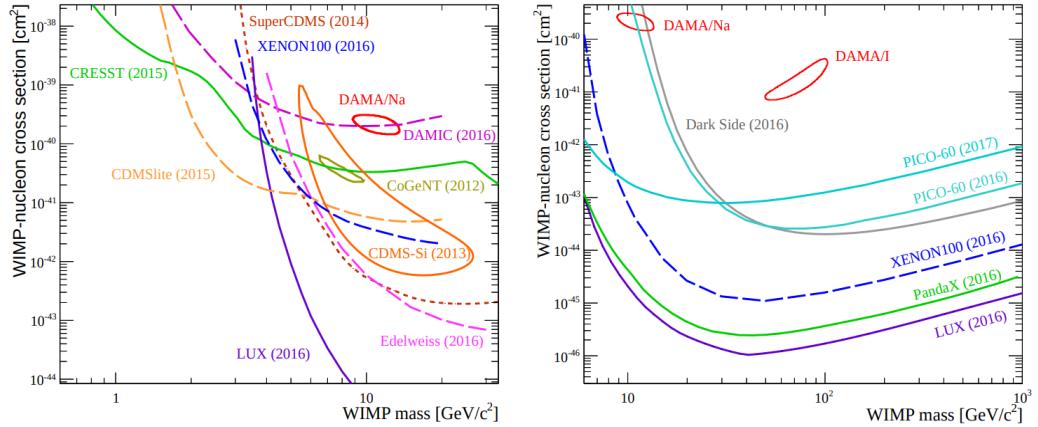


Figure 2.17: Exclusion limits obtained by various direct detection experiments considering a SI interaction cross section for low WIMP (on the left) or high WIMP masses (on the right) [57].

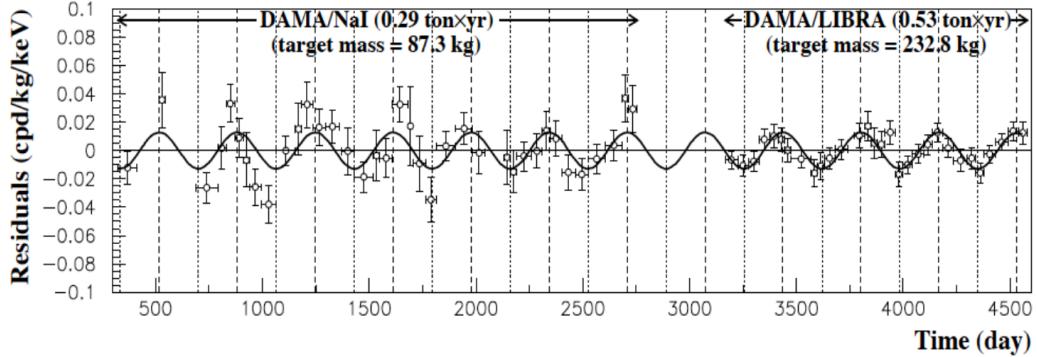


Figure 2.18: Observed and expected annual modulation in single hits events in the 2-6 keV energy range by the DAMA experiment [58].

This equation is extremely important for two reasons. First of all, it shows that, if a signal is found in direct detection, we could use this detection to determine the new object mass and scattering cross section and then use this information in order to obtain the DM density profile this way: the data obtained by the different strategies of detection are actually complementary. The second reason is that as we can see, the flux of incoming particles can actually be divided into two factors:  $P$ , entirely dependent on the physics of the DM particle, and the J-factor  $J(\Delta\Omega)$ , depending only on the distribution of DM within the system considered. This J-factor is in this sense actually a measurement of the quality of an astronomical object for an indirect measurement, since the higher the flux received, the better the measurement will be in general (even though this is not the only factor which matters, since for example the galactic center has a higher J-factor than the best dwarf galaxy observable, but also has a lot of backgrounds affecting the measurement).

As different channels of observations are available for us to analyze the eventual annihilation of DM, several strategies can be used in order to detect DM indirectly, as we will now see, by searching different kind of SM particles. Anyway, all these strategies have one goal in common: try to reduce the background, since the signal searched for is usually quite low while the uncertainties associated to the background in astrophysics are usually quite high.

### Through $\gamma$ -rays detection

The golden channel for such searches are through the production of  $\gamma$ -rays by DM annihilation  $\chi\chi \rightarrow \gamma + X$  or decay  $\chi \rightarrow \gamma + X$ , mainly because the energy scale of the WIMPs implies that most of the annihilation and decay radiation will be emitted in this range of energies and because  $\gamma$ -rays are usually not deflected when traveling to the observer (this means that the exact source of this kind of radiation can be quite easily and precisely pin-pointed). However, they do have one drawback as well: the Earth's atmosphere is usually opaque to this kind of radiation at this level of energy. This means that most of experiments searching for them simply cannot be performed from the ground and have to be sent to space.

One of the most famous detectors of this category is the Fermi Large Telescope (LAT), a pair production detector launched in June 2008 and mostly sensitive to  $\gamma$ -rays between 20 MeV and 300 GeV [61]. This experiment has managed to exclude a large portion of the phase space, as seen in Figure 2.19. The GAMMA-400 experiment, whose launch is scheduled in 2020, will pick up the work of LAT, by studying a similar range but with a much improved angular and energy resolutions [62].

It is much harder for DM to produce high energy  $\gamma$ -rays and therefore, the flux of such particles decreases quite quickly with energy, making it harder to study since telescopes then need a much larger effective area to pick up the same quantity of signal, and have therefore to be put on the ground. Such telescopes do exist, are called Imaging Atmospheric Cherenkov Telescopes (IACT) and have to take into account the atmospheric perturbations to work in an optimal way. They are usually sensitive to a range of energies going from  $\sim 10$  GeV up to  $\sim 100$  TeV, but can usually only study a small portion of the sky (up to a few degrees), forcing these experiments to choose carefully the objects to be studied. The Cherenkov Telescope Array (CTA) is a brand new telescope of this kind whose construction is supposed to start this year, that should improve greatly the sensitivity of high masses indirect DM searches.

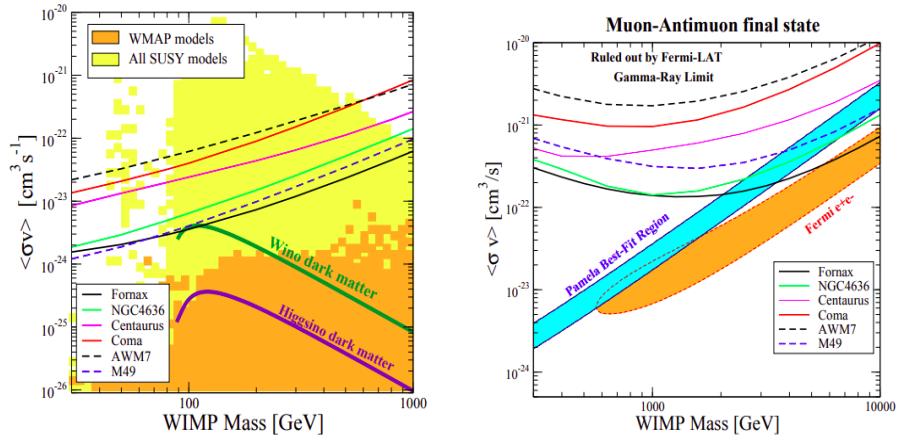


Figure 2.19: Upper limits on the DM annihilation cross section considering  $b\bar{b}$  (on the left) and  $\mu^+\mu^-$  (on the right) final states as a function of the WIMP mass, for different clusters studied [61].

### Through neutrinos detection

Interacting only weakly, neutrinos are another reliable source of data in the Universe since they are not supposed to be altered when traveling large distances as well, even though detecting neutrinos is usually much harder than detecting  $\gamma$ -rays and usually involves huge tanks of water in which neutrinos can produce a Cherenkov effect that will eventually be detected.

The most famous detector of this kind, the IceCube neutrino observatory, actually uses the ice of the South Pole instead of water to detect these particles with photo-detectors, mainly because of the low interaction cross section of the neutrinos which then require the installation of a huge volume of material to increase the probability of interaction. Super-Kamiokande (Super-K), in Japan, is another large Cherenkov experiment dedicated to the detection of cosmic neutrinos. Both detectors are also largely involved as direct searches experiments, since they are also sensitive to the eventual recoil between DM and ordinary matter nuclei.

The problem with this kind of experiment is the difficulty of actually detecting some neutrinos, along with the background levels from atmospheric neutrinos, as represented in Figure 2.20, typically several orders of magnitude larger than the signal. Several strategies therefore need to be put in place in order to reduce the background level in such experiments, such as the study of the directionality of the source and an appropriate choice of angle of observation, since most of the contamination is coming from tau neutrinos, themselves originating from muon neutrinos oscillation, strongly suppressed around the zenith.

### Through cosmic rays detection

Searching for anti-matter in the cosmic rays presents the advantage of being highly sensitive, because of the low levels of backgrounds this kind of searches implies. However, cosmic rays are affected when traveling through the Universe, and determining the exact location of their emission can be quite challenging.

Among the most famous detectors of this kind, we can quote PAMELA, a spatial telescope dedicated to the study of such cosmic rays since 2006. CERN's Alpha Magnetic Spectrometer (AMS), installed in the International Space Station has also studied such radiation from a range of a few

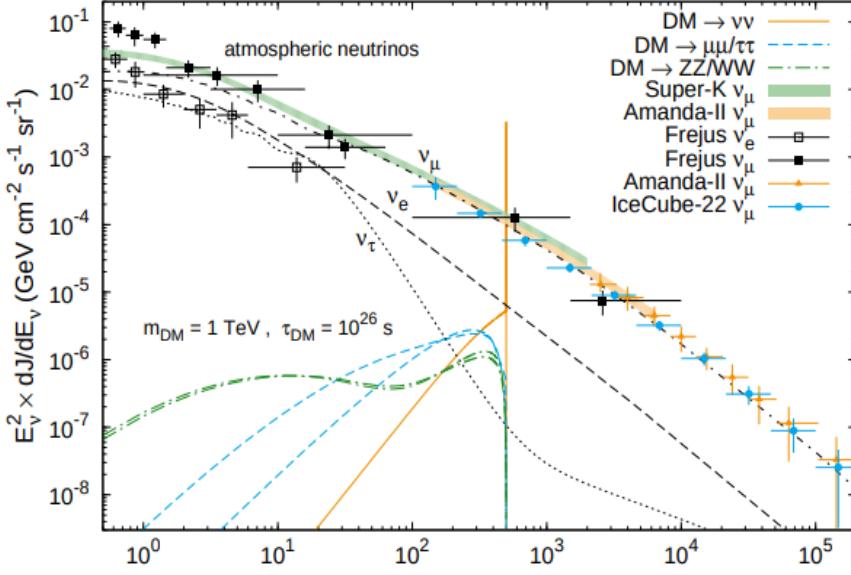


Figure 2.20: Neutrino spectra for a scalar DM candidate of 1 TeV for different indirect detection experiments and the corresponding background level expected [63].

hundred MeV up to 1 TeV. The data collected by this detector is compared to the exclusion limits obtained by the IceCube detector in Figure 2.21.

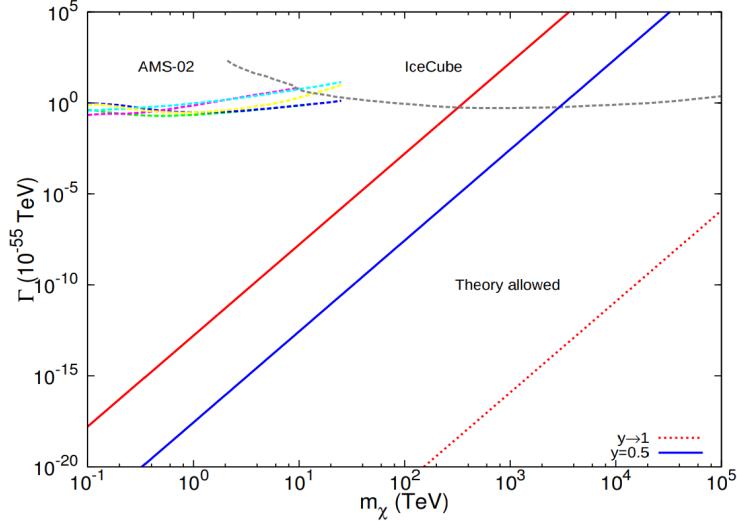


Figure 2.21: Limits of the decay width of the interaction with respect to the DM mass obtained by both IceCube and AMS [64].

### 2.5.3 Collider production

In this particular kind of searches, we are particularly interested in the eventual direct production of DM candidates following the collision between two highly energetic SM particles, a perfectly viable scenario if we keep assuming that DM should at least interact weakly with ordinary matter if we want to be able to produce or detect it in a laboratory.

In this case, two main models for the interaction between SM and DM particles can be considered, each with a different level of complexity and different possible applications:

- The Effective Field Theory (EFT) approach is usually considered to be the easiest and most simplified one, even though it can still give us plenty of information about this kind of interaction. It relies on a strong assumption, assuming that the energy scale of the new exotic physics is much larger than the actual energy accessible with our experiment, since the momentum transfer of this interaction needs to be much smaller than the mediator mass for this simplified approach to be valid.

According to this approach, represented in Figure 2.22, the interaction can be described using simplified operators (the most simple ones being either scalar  $\bar{\chi}\chi\bar{q}q$ , pseudoscalar  $\bar{\chi}\gamma^5\chi\bar{q}\gamma^5q$ , vector  $\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu q$  or axial-vector  $\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu\gamma^5q$ ) since, according to the assumption made, its mediator can usually be integrated out [53].

Despite of this strong assumption, this approach is actually quite useful anyway in the sense that it is able to provide us bounds on the new physics scale  $\Lambda$ , which can be related to the different couplings of the interaction as in Equation 2.20, where  $g_q$  and  $g_\chi$  are the coupling between the mediator and the SM or DM particle, and  $m_{\text{med}}$  is the mass of the mediator.

$$\frac{1}{\Lambda^2} = \frac{g_q g_\chi}{m_{\text{med}}} \quad (2.20)$$

Additionally, direct searches experiments typically also introduce this kind of assumptions to extract constraints from their measurements, making it straightforward to compare the results obtained in both approaches. However, the transfer of momentum in the direct searches experiments is usually of the order of a few keV as seen in Equation 2.21, while in collider experiments such as the LHC, this is of the order of  $\Theta(\text{GeV-TeV})$ .

$$E = \frac{m_\chi v_\chi^2}{2} \simeq 100 \text{ GeV} \cdot 10^{-6} \simeq 50 \text{ keV} \quad (2.21)$$

This means that, especially due to the increasing center of mass energy given by the LHC over the last few years, the basic EFT assumption is usually not respected and gives information about an out of reach phase space region anyway, making its usefulness quite relative in most cases. This is why more complex models need to be developed as well.

- The simplified models attempt to solve the issue related to the approximation made by the EFT approach, by increasing the level of details regarding the interaction between the dark and baryonic sectors. This is usually done by explicitly taking into account the mediator of the interaction, which can be considered of two different types in the context of this work: either scalar  $\phi$  or pseudoscalar  $a$  (both having a spin 0), described by the Lagrangians in Equation 2.22, considering the DM candidate to be a Dirac fermion coupling to the SM through the mediator considered and under the assumption of Minimal Flavour Violation (MFV). In this equation, the sum runs over the three SM families and the parameters  $y_i^f = \sqrt{2} \frac{m_i^f}{v}$  are the Yukawa couplings, much larger for the top quarks because of their mass, which will allow us to simplify the following equations [65].

$$\begin{cases} \mathcal{L}_{\text{fermion},\phi} \propto -g_\chi \phi \bar{\chi} \chi - \frac{\phi}{\sqrt{2}} \sum_i (g_u y_i^u \bar{u}_i u_i + g_d y_i^d \bar{d}_i d_i + g_l y_i^l \bar{l}_i l_i) \\ \mathcal{L}_{\text{fermion},a} \propto -ig_\chi a \bar{\chi} \gamma^5 \chi - \frac{ia}{\sqrt{2}} \sum_i (g_u y_i^u \bar{u}_i \gamma^5 u_i + g_d y_i^d \bar{d}_i \gamma^5 d_i + g_l y_i^l \bar{l}_i \gamma^5 l_i) \end{cases} \quad (2.22)$$

An important parameter in this case is the decay width of this mediator  $\Gamma_{\text{med}}$ , given by Equation 2.23 for either a scalar mediator  $\phi$  or Equation 2.24 for a pseudoscalar mediator  $a$ , where the first term corresponds to the mediator decay to SM particles, the second to its decay to DM particles and the last term its possible decay to gluons.

$$\begin{cases} \Gamma_\phi = \sum_f N_C \frac{y_f^2 g_\nu^2 m_\phi}{16\pi} \left(1 - \frac{4m_f^2}{m_\phi^2}\right)^{3/2} + \frac{g_\chi^2 m_\phi}{8\pi} \left(1 - \frac{4m_f^2}{m_\phi^2}\right)^{3/2} + \frac{\alpha_S^2 g_\nu^2 m_\phi^3}{32\pi^3 \nu^2} \left| f_\phi \left( \frac{4m_t^2}{m_\phi^2} \right) \right|^2 \\ f_\phi(\tau) = \tau \left( 1 + (1 - \tau) \arctan^2 \left( \frac{1}{\sqrt{\tau - 1}} \right) \right) \end{cases} \quad (2.23)$$

$$\begin{cases} \Gamma_a = \sum_f N_C \frac{y_f^2 g_\nu^2 m_a}{16\pi} \left(1 - \frac{4m_f^2}{m_a^2}\right)^{1/2} + \frac{g_\chi^2 m_a}{8\pi} \left(1 - \frac{4m_f^2}{m_a^2}\right)^{1/2} + \frac{\alpha_S^2 g_\nu^2 m_a^3}{32\pi^3 \nu^2} \left| f_a \left( \frac{4m_t^2}{m_a^2} \right) \right|^2 \\ f_a(\tau) = \tau \arctan^2 \left( \frac{1}{\sqrt{\tau - 1}} \right) \end{cases} \quad (2.24)$$

In the case of the simplified models, the minimal set of parameters describing the interaction is therefore  $\{m_\chi, m_{\text{med}}, g_\chi, g_q\}$ .

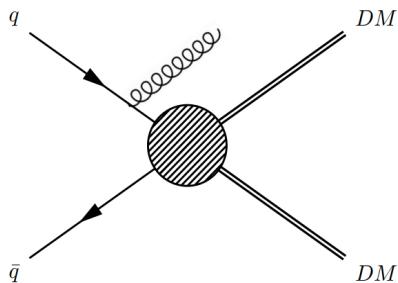


Figure 2.22: Schematic representation of a typical EFT modelization of an LHC event with an Initial State Radiation (ISR) object used to trigger the event [53].

An additional categorization of the DM production models can be done, by separating the so-called s-channel and t-channel models, as shown in Figure 2.23. In the first case, the most common one in collider searches, the mediator between the SM and DM is assumed to be a boson, and usually decays directly into a pair of DM particles. On the other hand, in the t-channel models, the mediator couples to one quark and one DM particle, with a colored exchange particle required.

In this work, simplified models with either scalar or pseudoscalar mediators will always be considered because of their relative simplicity and the lack of strong assumptions behind them. We will now study in particular the production and the search for DM within the LHC.

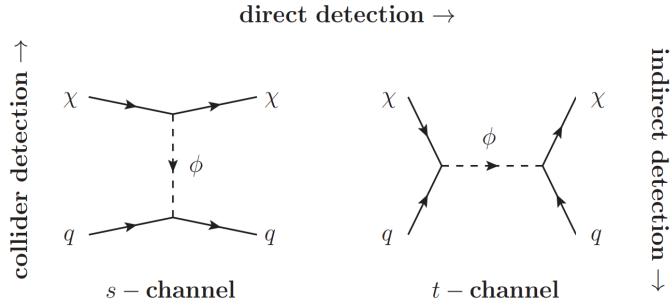


Figure 2.23: Schematic representation of a typical collider DM production through s-channel and t-channel processes [66].

## 2.6 DM production at the LHC

The Large Hadron Collider (LHC), colliding protons at a center of mass energy of 13 TeV, is actually a perfect machine to study such processes, because of the expected range of masses (100 GeV to 1 TeV) for the best DM candidates, as discussed in Section 2.3. However, because of the weak interactions of such particles, they are not expected to interact at all with the detector, which will then basically search for missing transverse energy along with a SM particle triggering the event.

Several major categories of DM searches exist at the LHC, performed mainly by the CMS and ATLAS collaborations, depending on the strategy applied for such searches:

- First of all are the so-called **mono-X** searches, where X stands for a detectable SM particle used to trigger the event while the DM mediator usually decays into a pair of particles escaping the detector, leaving behind some MET, a key variable to all these searches, as shown in Figure 2.24. Depending on the nature of the X particle, several searches can be performed: we can for example mention the mono- $\gamma$  [67, 68] or mono-jet [69, 70] 13 TeV searches performed by the ATLAS and CMS collaborations.

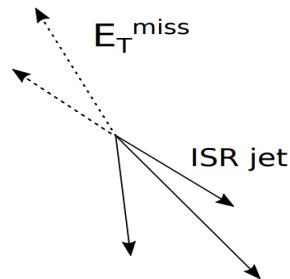


Figure 2.24: Schematic representation of a typical mono-X event, with an ISR jet in this case going back-to-back with some MET [53].

Additionally, if the DM mass is high enough, a coupling to the Higgs boson is also possible, or an eventual decay of the Higgs itself to a couple of DM particles  $H \rightarrow \chi\bar{\chi}$  is kinematically not impossible. This channel, known as the mono-Higgs searches is sensitive to all the decays of the Higgs, even though the searches excluding most of the phase spaces are performed using the  $H \rightarrow b\bar{b}$  and  $H \rightarrow \gamma\gamma$  decays, mainly because of their branching ratios [71, 72].

- Then, we can find the searches for DM production in **association with heavy quark(s)**. This particular search for DM produced in association with one or two quark pairs can be typically quoted within this category and will be detailed in the next section, but other analyses such as the one probing the  $b\bar{b} + \text{DM}$  to its different final states depending on the decay of the bottom quark also belongs to this group, at 8 or 13 TeV [13, 17].

These searches have to combine the discriminant power of several variables to separate the signal from the backgrounds, since the MET distribution on its own is usually not enough, but they present the advantage of being favored by the higher Yukawa coupling to more massive particles implied by the MFV assumption.

- **Dijet searches** are also an important part of the work being performed using the LHC data, and actually provide the best exclusion limits for most the spin-1 mediated models considered (up to a few TeV for typical coupling choices!) [73, 74]. In this case, the sensitivity is obtained by searching for either narrow or large resonances on the exponentially falling QCD background, while the other searches were mostly dedicated in searching for bumps in the MET spectrum.
- **Supersymmetric searches**, such as SUSY searches, can also be performed to search for DM which would solve the hierarchy problem while giving us perfect DM candidates such as the neutralino  $\chi$ , the lightest stable supersymmetric particle, obtained in many of the MSSM theories, having an hypothetical mass below the TeV scale [75].
- The **Higgs portal** to the dark sector is another interesting strategy. In some specific cases, when considering spin-0 mediated interactions between the dark and baryonic sectors, the Higgs could be considered as the mediator of the interaction as well, which only require a minimal modification of the SM Lagrangian. It is then necessary to study the different Higgs production modes, such as the gluon fusion and Vector Boson Fusion (VBF) mechanisms, to search for an eventual invisible decay of the Higgs into a couple of DM particles, assuming that its mass is lower than  $\sim 62.5$  GeV,  $m_H/2$  [76].
- Finally, and this is quite new, **long-lived searches** can also be performed. These are interested because they can extend the current searches performed to also consider the eventual creation of long-lived particles which would decay a few centimeters further than the primary vertex of the  $pp$  collision [77]. Typical SM signatures do not usually include such events, making this channel relatively background-free, even though the reconstruction of the different objects is much harder in this case.

All the results obtained the different searches performed by the CMS collaboration using the full 2016 dataset can be summarized in Figure 2.25 for spin-0 mediators and in Figure 2.26 when considering spin-1 mediators.

These results have also been compared to the results obtained by several direct detection experiments in Figure 2.27, considering both the Spin Dependent (SD) and Spin Independent (SI) cases, as explained in Section 2.5.1.

This particular analysis and our signal of interest, the search for DM production in addition with a single or a pair of top quarks falls into the second category and will now be detailed. It is already important to note that this analysis has been performed considering the DM candidate to be a 1 GeV Dirac fermion with all the couplings equal and set to 1, as recommended by the Dark Matter Working Group (DMWG) [78].

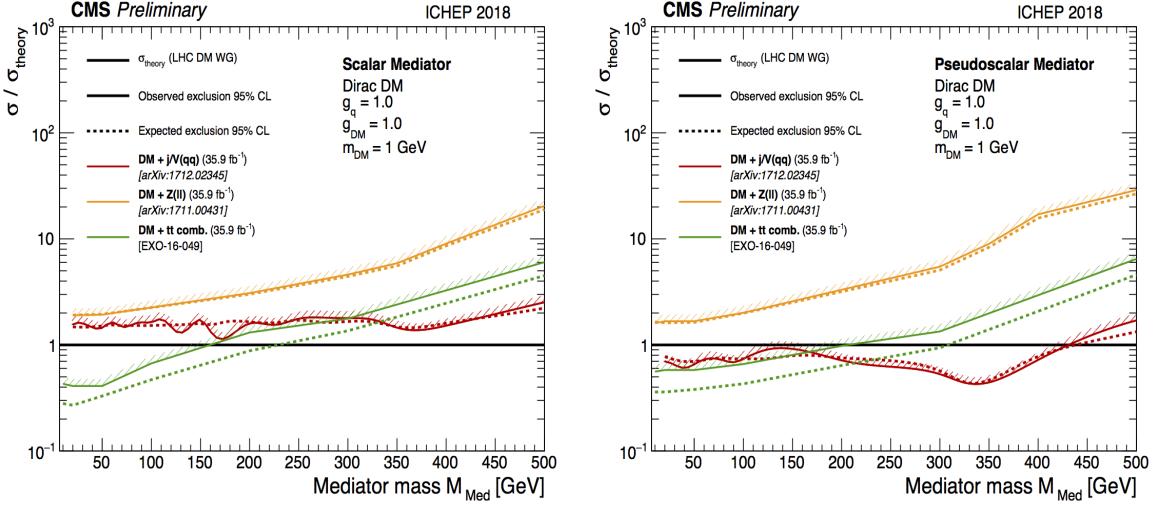


Figure 2.25: Observed and expected 95% exclusion limits obtained by different searches of the CMS collaboration as a function the spin-0 scalar (on the left) or pseudoscalar (on the right) mediator.

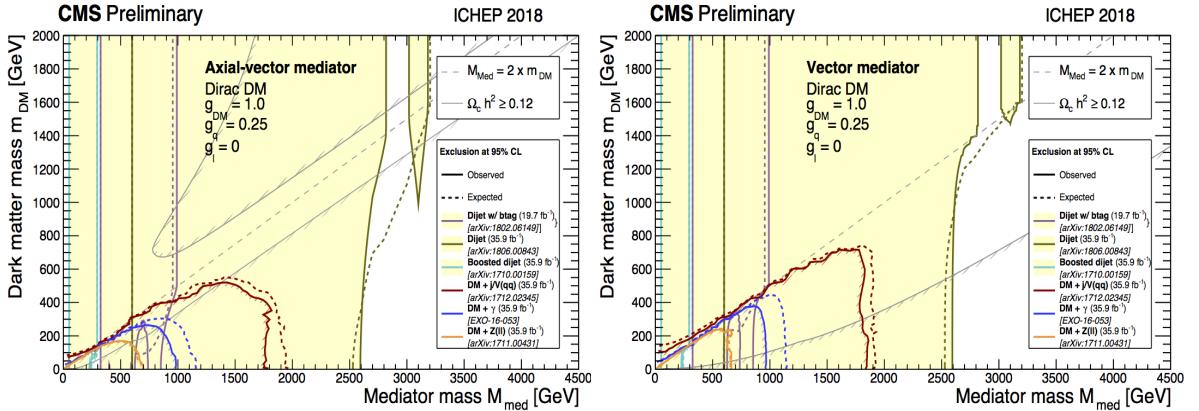


Figure 2.26: Observed and expected 95% exclusion limits obtained by different searches of the CMS collaboration as a function the spin-1 mediator, considering an axial-vector (on the left) and axial (on the right) interaction.

### 2.6.1 The single top production channel

The first kind of signal considered in this analysis is the production of DM in association with a single top quark, known as  $t/\bar{t}$ +DM analysis. This process is expected to be mediated by a spin-0 mediator, either scalar or pseudoscalar, and is associated with a light quark and a W boson (the mono-top analysis is dealing with the case where a single top is created without any additional particle).

Three different Feynman diagrams can be associated to this particular analysis depending on the model considered, as shown in Figure 2.28.

As discussed previously, the top quark will be dynamically favored due to its high mass and therefore high Yukawa coupling, but however this has a negative effect as well, since the lifetime of this quark is extremely low, of the order of  $10^{-15}$  s [79]. This means that this particle usually decays before being able to form hadrons and before reaching the CMS detector, so we can only detect the products of its decay, not the top quark itself. In almost 100% of the cases, the top actually decays into a bottom quark and a W boson, which decays itself before being detected into

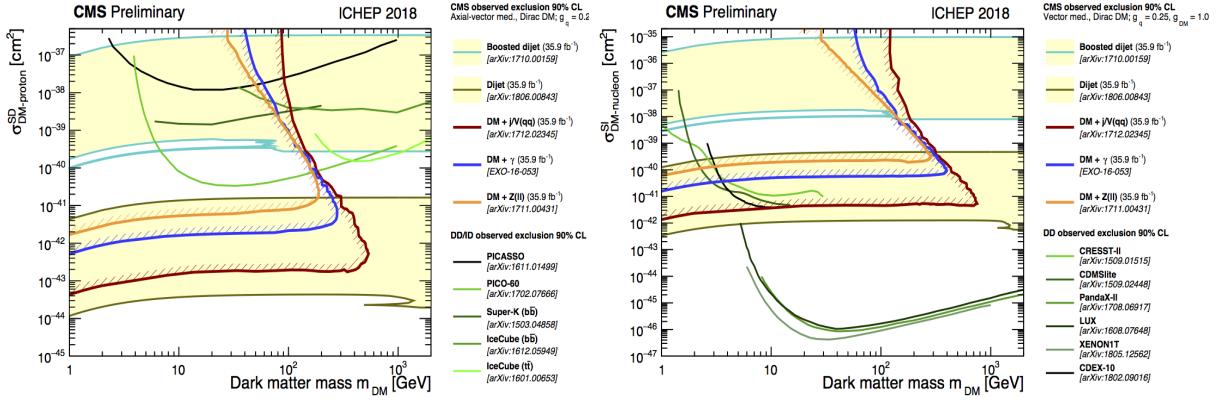


Figure 2.27: CMS 90% exclusion limits compared to the most famous direct detection experiments for the SD (on the left) and SI (on the right) scenarios, obtained using similar couplings.

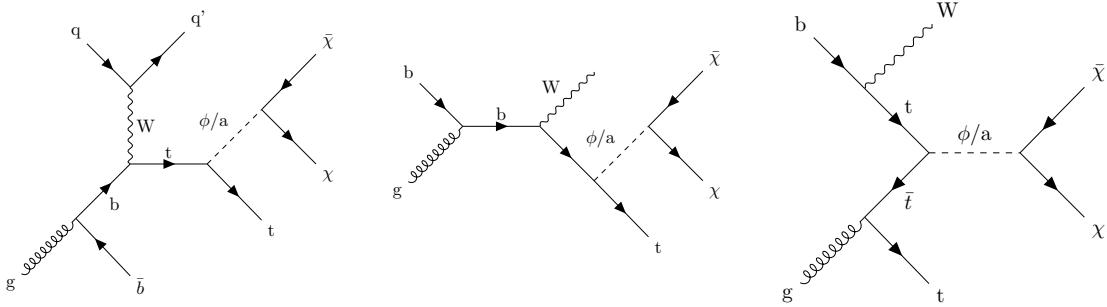


Figure 2.28: Feynman diagrams involving the production of DM with a single top quark with its associated t-channel W boson (on the left), or tW (on the center and on the right) production.

quarks and/or leptons. Even though this will be detailed in Chapter 6, we can already conclude that the typical final state of such signature is therefore made out of MET coming from the DM, one b-tagged jet along with one or two W bosons, seen as a combination of jets and leptons, depending on the channel considered.

## 2.6.2 The $t\bar{t}$ production channel

The  $t\bar{t}+DM$  analysis is really similar, except that in this case, we have two top quarks in the final state, as represented in Figure 2.29.

The final state expected by such an event is therefore made of two b-tagged jets, along with two W bosons and some MET coming from the pair of DM particles as well. In this case as well, both scalar  $\phi$  and pseudoscalar  $a$  mediators will be considered as well, with a mass range from 10 to 500 GeV.

## 2.6.3 The dilepton final state

As previously explained, most of the models considered will produce exactly two W bosons, which are not stable long enough and therefore decay before reaching the detector, meaning that we cannot directly detect such bosons. However, we can detect the results of the decays of these W, since they can decay either to hadrons ( $67.6 \pm 0.27\%$  branching ratio) or to a lepton and a neutrino,

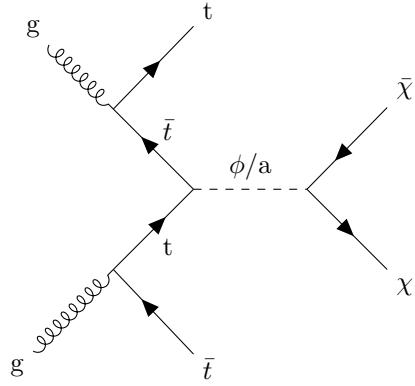


Figure 2.29: Schematic representation of a typical  $t\bar{t}$ +DM event.

giving us an additional contribution of MET ( $10.8 \pm 0.09\%$ ) [79].

This means that this kind of analyses featuring two W bosons in the final state can focus on different channels: either completely hadronic (if both the W decay into quarks), semileptonic or dileptonic (when both the W decay into two neutrinos and two leptons of opposite charge).

Given the Branching Ratio (BR) of the W decay, it is easy to see that the dileptonic channel, which will be the focus of this work, is clearly not favored statistically. However, it is interesting anyway to study this kind of decay because this channel typically features less backgrounds than other channels, resulting in a better signal isolation, and because leptons can usually be reconstructed in a better way than jets (cf. Chapter 4), resulting in lower uncertainties and in improved limits.

## 2.7 Previous relevant results

This analysis being performed at a center of mass energy  $\sqrt{s} = 13$  TeV, only the most relevant results to this particular energy and previously obtained by both the CMS and ATLAS collaborations of CERN will be quoted here.

First of all, the **ATLAS collaboration** published interesting results at this center of mass energy, considering an integrated luminosity of  $13.3 \text{ fb}^{-1}$ , and obtained the corresponding exclusion limits at the 95% CL, considering the  $t\bar{t}$ +DM model and both scalar and pseudoscalar spin-0 mediators for the interaction, as shown in Figure 2.30. In this case, and for the couplings considered, an exclusion up to  $\sim 375$  GeV has been achieved.

Considering the full 2016 dataset of  $36.1 \text{ fb}^{-1}$ , similar results have been obtained, as shown in Figure 2.31. In this case, and for lower coupling values, an exclusion up to around 100 GeV has been obtained considering scalar and pseudoscalar mediators.

The **CMS collaboration** published in 2018 a similar analysis, but using only the  $35.9 \text{ fb}^{-1}$  of data collected during the year 2016. This analysis combined the three different final states possible (hadronic, semileptonic and dileptonic) and computed the limits on the signal strength for different mediator and dark matter masses, considering both scalar and pseudoscalar mediators [19]. The results obtained can be found in Figure 2.32.

Last year, a combination of the  $t/\bar{t}$ +DM and  $t\bar{t}$ +DM analyses has also been published [20], combining this time only the hadronic and semileptonic channels of both analyses. The limits obtained

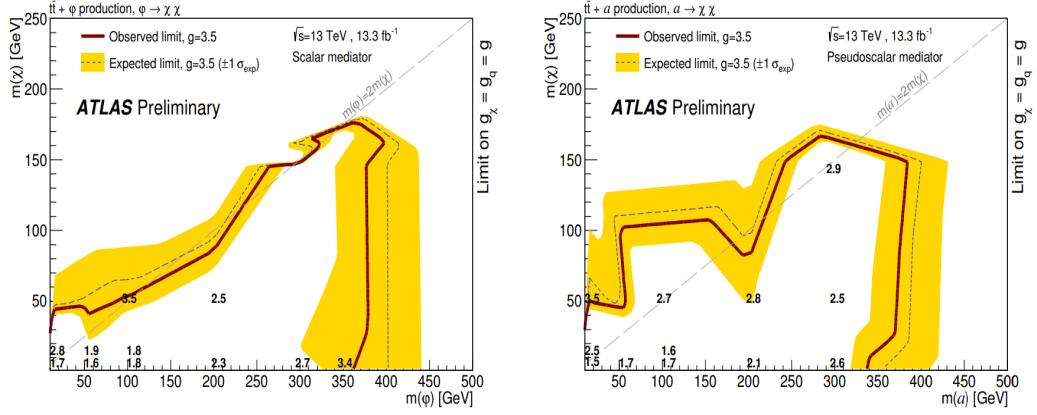


Figure 2.30: Limits on the mass on the DM and mediator masses obtained by ATLAS using  $13.3 \text{ fb}^{-1}$  of  $13 \text{ TeV}$  data, considering scalar (on the left) and pseudoscalar (on the right) mediators [14].

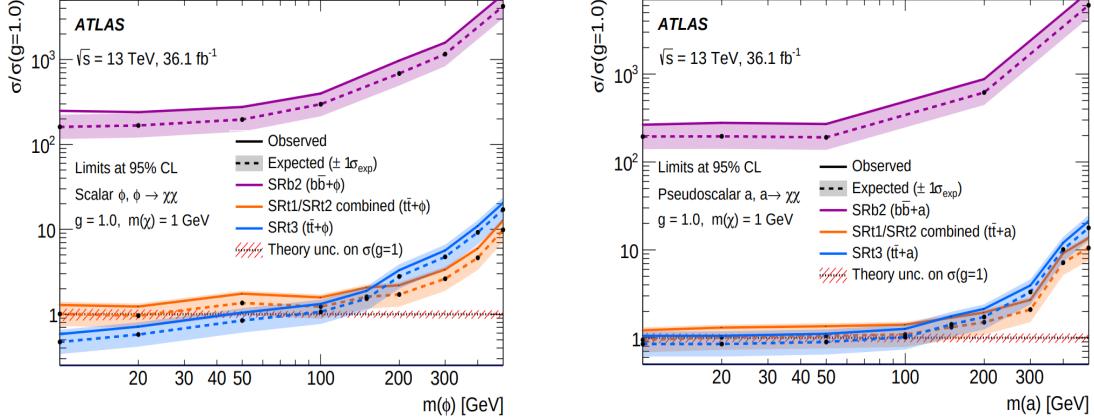


Figure 2.31: Exclusion limits at the 95% CL obtained by ATLAS considering scalar (on the left) and pseudoscalar (on the right) mediators, for a DM mass of  $1 \text{ GeV}$  [17].

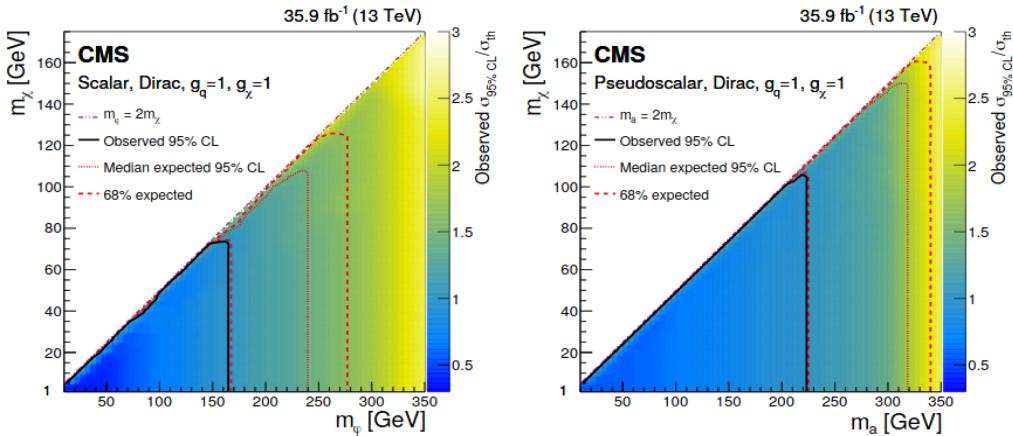


Figure 2.32: 95% CL exclusion plots on the signal strength computed as a function of the mediator and DM masses obtained by CMS considering a scalar (on the left) and a pseudoscalar (on the right) mediator for the interaction [19].

in this case are represented in Figure 2.33, where the limits obtained by each analysis on their own along with the results of the combination, leading to a factor 2 improvement of the limits obtained, have been represented.

This combination managed to exclude the production of scalar mediators up to 290 GeV and pseudoscalar mediators up to 300 GeV, at the 95% CL for the couplings considered. This combined analysis actually lead to the most stringent exclusion limits of the LHC on the production of DM through these categories of spin-0 mediators.

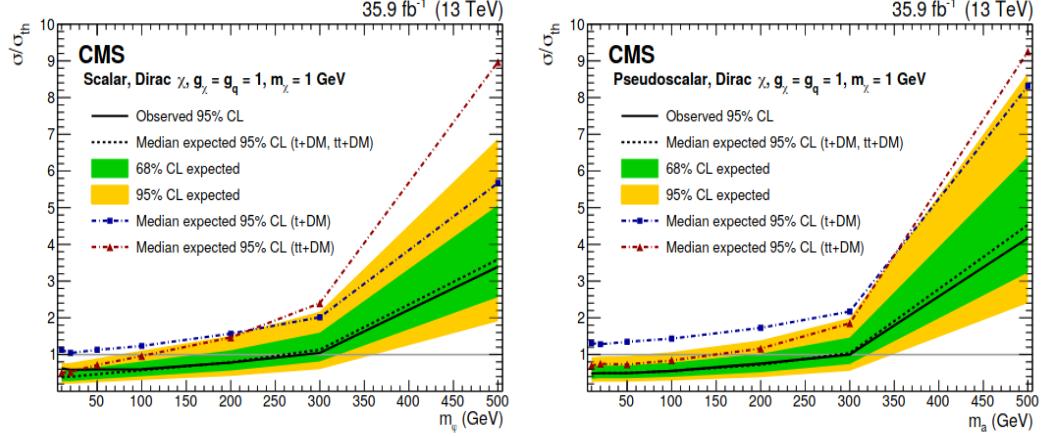


Figure 2.33: Expected and observed 95% CL limits on the DM production cross sections shown considering scalar (on the left) and pseudoscalar (on the right) mediators for the interaction [20].



# Chapter 3

## The experimental setup

The data analyzed throughout this work is the result of the proton-proton collisions at a center of mass energy of 13 TeV provided by the Large Hadron Collider (LHC). It is therefore extremely important to describe this particle accelerator and to detail its main characteristics in order to understand better the origin of our data, as will be done in Section 3.1.

This kind of collisions at high energy typically produces hundreds or even thousands of new particles that we then need to identify, and whose exact properties, such as their momentum or electric charge, need to be measured. This is done by using the Compact Muon Solenoid (CMS) detector, an incredibly sophisticated detector, result of the work of thousands of scientists and made of several different layers, which will be described in Section 3.2 of this chapter.

### 3.1 The LHC accelerator

The Large Hadron Collider (LHC) is a superconducting particle accelerator able to accelerate protons and heavy lead ions up to a velocity close to the speed of light. Planned since the end of the 20th century, this accelerator, a 27 kilometers ring put 100 meters underground to remove part of the contamination due to the cosmic rays and located at CERN, has now been running for 10 years. The LHC is the result of the collaboration between thousands of scientists of more than 100 different nationalities and its main objective was first of all to either infer or confirm the possible existence of the Higgs boson, theoretically predicted in the 1960s [1, 2] but never observed in any experiment. The discovery of the Higgs boson, announced at CERN on the 4th of July 2012 [3, 4], was then an incredible achievement of the accelerator, after only two years of operation.

Now that the Higgs boson has been discovered, the priority of the LHC shifted a bit. Even though many different teams are still studying this particle in order to determine precisely its most fundamental properties such as its mass or its spin, many groups of scientists are now involved in different kind of BSM physics since the LHC, whose center of mass energy has kept increasing over the years, allow us to reach a level of energy never reached before and therefore allows us to probe new parts of the phase space, searching for eventual hints of new physics. These kind of searches of course include the search for eventual DM production as the one performed in this work.

Let's now start with a global description of the general design of this particle accelerator in Section 3.1.1 before describing in more details the key parameters allowing us to evaluate its actual performances in Section 3.1.2.

### 3.1.1 The LHC in a nutshell

The LHC is an underground particle accelerator built in the same 27 kilometers tunnel where the Large Electron Positron collider (LEP) was previously used. This machine is accelerating two beams made out of billions of protons or lead ions in each direction and is mostly made out of more than 4000 superconducting magnets, in majority dipoles and quadrupoles, allowing respectively to curve the beam to maintain a nominal circular trajectory and to focus it by compensating its natural dispersion due to the repulsion of the protons making up these beams. A dedicated small section of the accelerator is then made out of 16 radio-frequency cavities synchronized in such a way that these protons always face a negative electric charge, which is used as the driving process of the actual acceleration of such particles.

Once the nominal center of mass energy  $\sqrt{s} = 13$  TeV reached (this concept will be described in Subsection 3.1.2), the protons are then smashed together in four different places on the LHC, where the four detectors (ATLAS, CMS, A Large Ion Collider Experiment (ALICE) and LHCb) have been placed in order to study the collisions. Both ATLAS and CMS are general detectors able to study exotic processes such as the production of DM and to make precision measurements on SM physics as well (the decision to build two separate detectors was made in order to introduce some redundancy and to check the results). ALICE on the other hand is mostly dedicated to the study of heavy ions collisions that happen  $\sim 10\%$  of the time in order to study in particular a specific state of matter, called the quark-gluon plasma [80]. Finally, LHCb has been designed to study in particular the CP violation phenomena, which could be the sign of some new physics [81].

It is important to note at this point that the LHC is not a standalone accelerator in the sense that protons enter the LHC with a velocity already close to the speed of light. In order to reach such input energies ( $\sim 450$  GeV), previous smaller accelerators of CERN are still used today. A chain of accelerators is then formed: first of all, the protons, extracted from a bottle of ordinary hydrogen, are injected into the LINAC 2, a linear accelerator, before being transferred to the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS) and finally the LHC itself (all this chain of acceleration can be found in Figure 3.1).

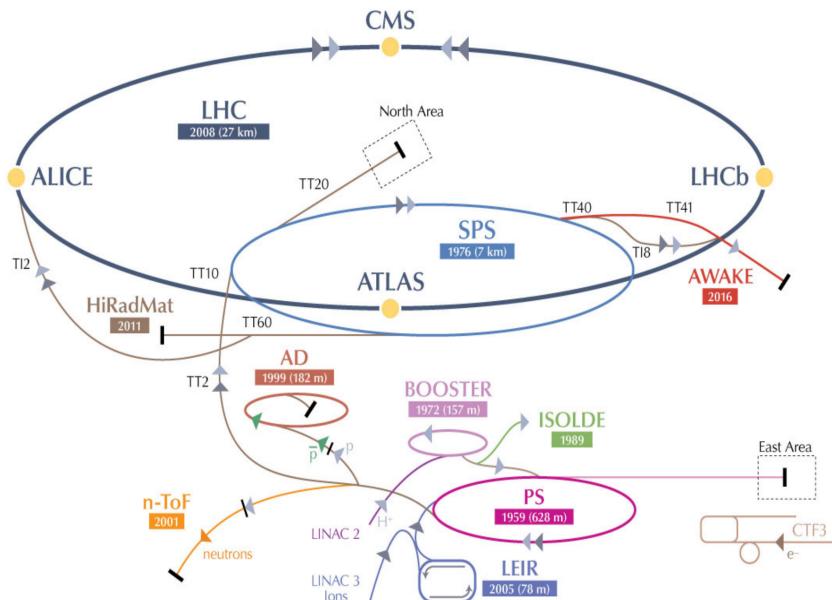


Figure 3.1: LHC injection chain and experiments performed at CERN [83].

Parameter	Run I	Run II	Run III	Design
Energy [TeV]	7 → 8	13	14	14
Bunch spacing [ns]	50	25	25	25
Intensity [ $10^{11}$ ppb]	1.6	1.2	Up to 1.8	1.15
Bunches	1400	2500	2800	2800
Emittance [ $\mu\text{m}$ ]	2.2	2.2	2.5	3.5
$\beta^*$ [cm]	80	30 → 25	30 → 25	55
Crossing angle [ $\mu\text{rad}$ ]	-	300 → 260	300 → 260	285
Peak luminosity [ $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ]	0.8	2.0	2.0	1.0
Peak PU	45	60	55	25

Table 3.1: Expected and observed main parameters of  $pp$  operation of the LHC across the different eras of operation [82].

During this phase of acceleration, the beam is separated in 2808 bunches of protons nominally separated by 25ns (giving a collision rate of 40 MHz), so that the experiments have the time to record the collision, clean the detectors and get ready for the next bunch crossing, coming just a few nanoseconds later. Each time one bunch crossing happens, around 35 collisions of protons happen at once: this phenomena, usually referred to as Pile Up (PU), has to be taken into account as well and will be described in the next section.

As previously stated, the amount of data collected is a crucial parameter for many analyses, meaning that the LHC would ideally need to run 24 hours a day, 365 days per year in order to maximize the data taking. However, this is not possible, as setting up a beam takes time and cannot keep rotating at maximal energy in the machine for a long time, so in ideal conditions around 20 hours of data taking a day are expected. The LHC is then running around 9 months per year, being usually stopped during the winter for maintenance operations, and the data taking periods are defined into Runs of a few years, after which the accelerator is usually stopped for a longer period of time, a Long Shutdown (LS), in order to also have the time to perform upgrade operations of the machine.

The data analyzed in this work corresponds to the second phase of the Run II of operation of the LHC, from 2016 to 2018, while the Run III is now expected to start in the Spring of 2021. The summary of the main parameters of operation of the LHC across the different Runs of operations can be found in Table 3.1.

### 3.1.2 LHC key parameters

#### Center of mass energy

The center of mass energy is defined as a Lorentz invariant quantity under longitudinal boosts resulting of the collisions of two protons (defined as  $E_1, \vec{p}_1, m_1$  and  $E_2, \vec{p}_2, m_2$ ) with a  $\theta$  angle, as developed in Equation 3.1. It is a key variable of the LHC since the phase space of particles that

can be explored directly depend on this value.

$$\sqrt{s} = \sqrt{(m_1)^2 + (m_2)^2 + 2(E_1 E_2 - 2|\vec{p}_1| |\vec{p}_2| \cos(\theta))} = 13 \text{ TeV} \quad (3.1)$$

The LHC started its operation in 2008 running at an energy of 7 TeV, quickly moved to 8 TeV and kept this level of energy during the end of the Run I of operation. In 2015, for the Run II of data taking, the energy was increased until reaching 13 TeV (2 times 6.5 TeV for each beam) and an expected value of 14 TeV, the nominal energy for which the LHC was originally built, is expected to be reached in the near future.

## Luminosity

The luminosity is an another extremely important variable for the operation of the LHC since, as a rule of thumb, it gives an indication on the number of collisions per second given by the accelerator. Increasing this parameter is then crucial to collect as much data as possible, in order to be able to isolate processes having a low production cross section and therefore an extremely low probability of apparition when colliding two protons.

Mathematically, we can first of all define in Equation 3.2 the rate  $R$  of production (in number of events per second) of any given process using its cross section  $\sigma$ , equivalent to its production probability, and the instantaneous luminosity  $\mathcal{L}$ . From this rate can be extracted easily the number of expected interactions  $N$  in a certain amount of time  $T$  as well.

$$\begin{cases} R = \mathcal{L} \cdot \sigma \\ N(T) = \sigma \int_0^T \mathcal{L}(t) dt = \sigma L \end{cases} \quad (3.2)$$

This instantaneous luminosity we just introduced  $\mathcal{L}$  can be defined using Equation 3.3, assuming that the beams have a Gaussian profile, while the integrated luminosity  $L$  can be defined by simply integrating the instantaneous luminosity over time  $L = \int \mathcal{L} dt$  [84].

$$\mathcal{L} = \frac{\gamma f_{\text{rev}} k_B N_p^2}{4\pi \epsilon_n \beta^*} R \text{ where } R = \frac{1}{\sqrt{1 + \frac{\theta_C \sigma_z}{2\sigma}}} \quad (3.3)$$

In this last equation, the following properties of the accelerator have been introduced, giving the LHC a nominal instantaneous luminosity  $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ :

- $\gamma$  is the usual relativistic Lorentz factor (for the LHC, when protons go at their maximum velocity of 99.9999991% of the speed of light,  $\gamma = 7460$ )
- $f_{\text{rev}}$  is the frequency of revolution (11.2 kHz)
- $k_B$  is the number of proton bunches per beam (2808 for a 25 ns bunch spacing)
- $N_p$  is the number of protons per bunch ( $1.15 \cdot 10^{11}$  protons)
- $\epsilon_n$  is the transverse normalized emittance ( $3.75 \mu\text{m}$ )
- $\beta^*$  is the betatron function at the interaction point (0.55 m)

- $R$  is the geometrical factor accounting for the fact that the collisions do not happen exactly head-on, therefore reducing the effective luminosity, expressed from the full crossing angle between colliding beam  $\theta_C$  (285  $\mu$ rads), and  $\sigma, \sigma_z$ , the transverse and longitudinal r.m.s. sizes (respectively 16.7  $\mu$ m and 7.55 cm).

As previously stated, increasing the luminosity of the LHC is always something interesting in order to produce processes having a low production cross-section, and we can then see that many different parameters can be tweaked in order to achieve the highest possible instantaneous luminosity, such as the number of protons per bunch, the number of bunches per beam or the beam crossing angle at the interaction point. New crab cavities will probably be installed during the next LS of the LHC in order to increase the value of the geometrical factor  $R$  and the instantaneous luminosity by a factor  $\sim 10$  (HL-LHC project [85]).

The total integrated luminosity taken by the LHC during its different years of operation has been summarized in Figure 3.2. The final datasets available and analyzed in this work have an integrated luminosity of 35.9  $\text{fb}^{-1}$  (2016), 41.5  $\text{fb}^{-1}$  and 59.7  $\text{fb}^{-1}$  (2018), resulting in a total dataset of 137  $\text{fb}^{-1}$  recorded during the Run II of operation. This means that we roughly expect to produce around 137 events of any process whose cross section of production would be equal to 1  $\text{fb}$ .

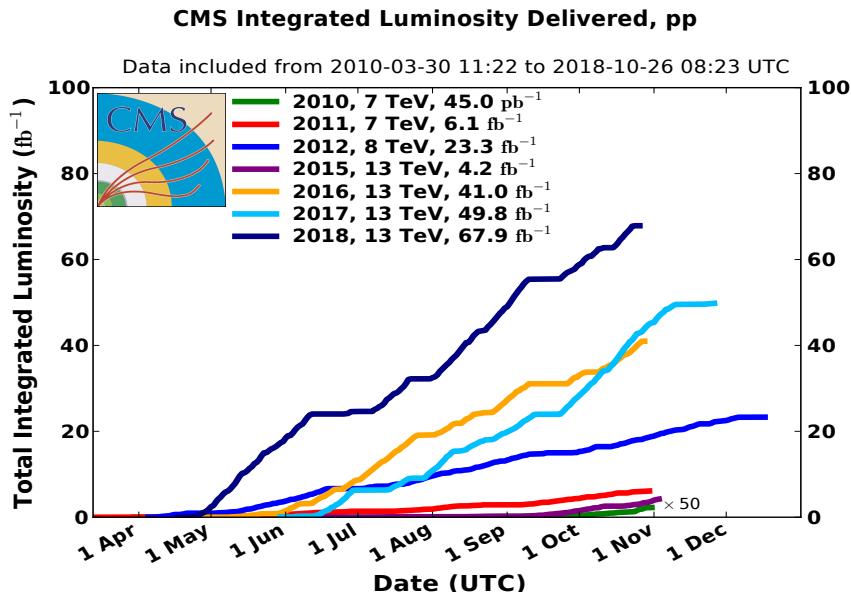


Figure 3.2: Integrated luminosity collected by CMS over its different years of operation so far.

### Pile Up (PU)

The last key parameter of the LHC discussed here is the Pile Up (PU). Usually, because of the high density of protons within the beams, a bunch crossing in a experiment produces around 35 collisions protons collisions, as seen in Figure 3.3, of which only the most interesting one will be kept, defining the PV, while the other vertices are usually referred to as the PU. The tracker of CMS, which will be introduced in Section 3.2.1, therefore needs to be able to reconstruct all these events in order to define the PV of the interaction.

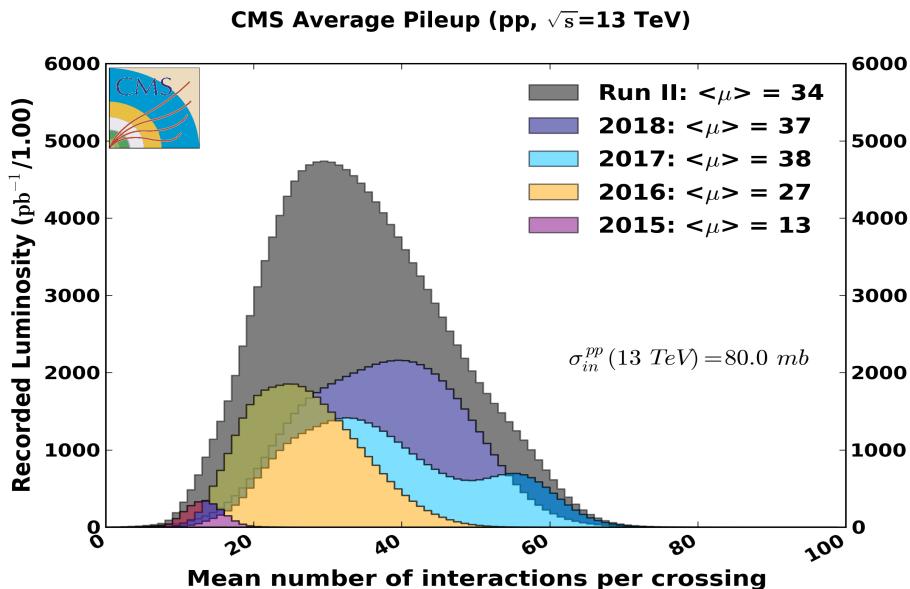


Figure 3.3: Mean PU distribution and luminosity recorded by CMS over the different years of operation of the LHC.

## 3.2 The CMS detector

The Compact Muon Solenoid (CMS) is one of the two general purpose detectors installed at the access point 5 of the LHC. Its main purposes were to discover the Higgs boson, make precision measurement of most of the SM processes and to try and discover BSM physics, such as the possible existence of DM.

It has been carefully designed by hundreds of different physicists and engineers in order to be as hermetic as possible, covering all the possible angles around the beam pipe, the point of origin of all the collisions provided by the LHC, and is therefore made of three main sections: the cylindrical barrel at the center, and the two endcaps, one on each side of the detector. CMS, quite compact with its 14.000 tons, diameter of 15 meters and length of 28.7 meters, was lowered into the experimental cavern after being built on the surface in 14 different moving pieces, a flexible design allowing to access its inner parts, by opening and closing the detector when needed.

### Main goals

The CMS detector has been built in order to meet the goals of the LHC physics program, which have been summarized in [86] as:

- Good muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution ( $\sim 1\%$  at 100 GeV), and the ability to determine unambiguously the charge of muons with  $p < 1$  TeV.
- Good charged-particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of taus and b-jets, requiring pixel detectors close to the interaction region.

- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution ( $\sim 1\%$  at 100 GeV), wide geometric coverage,  $\pi^0$  rejection, and efficient photon and lepton isolation at high luminosities.
- Good missing-transverse-energy and dijet-mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

## CMS subdetectors

As shown in Figure 3.4, CMS is made out of different layers corresponding to different subdetectors, each able to provide different kind of information about the particles created by each collision [86]. These subdetectors will be described in details in the following sections, but they have all been designed in order to make the reconstruction of the different events as efficient, fast and precise as possible and need to be quite cheap but still extremely resistant to the radiation produced.

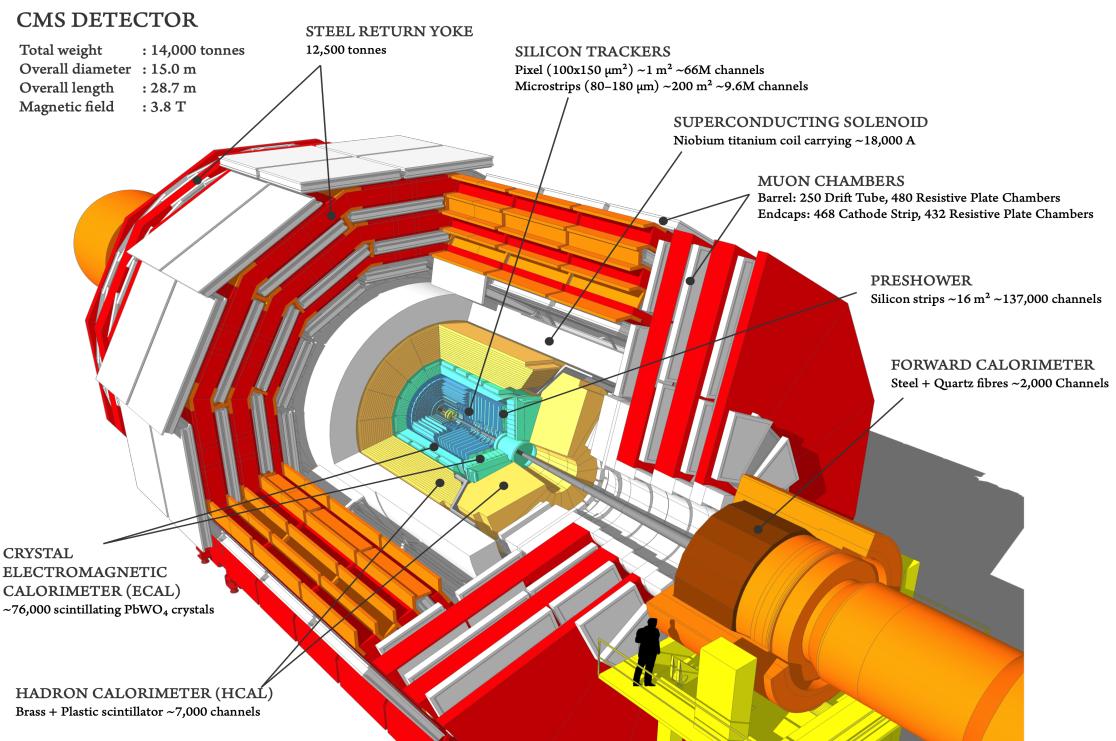


Figure 3.4: Schematic representation of the CMS detector, along with all its subdetectors and main characteristics.

The inner part of the CMS detector is the so-called tracker, a device made out of silicon pixels and strips, described in Section 3.2.1 and responsible for the precise reconstruction of all the charged particles coming from the different interaction vertices. A bit further outside can be found the Electromagnetic Calorimeter (ECAL), made out of thousands of crystals as described in Section 3.2.2 and used to precisely measure the energy of particles able to interact electromagnetically by producing an electromagnetic shower that can be detected. Then comes the Hadronic Calorimeter (HCAL), described in Section 3.2.3 and whose main purpose is to identify and measure the main properties of the hadrons produced in the collisions by producing hadronic showers.

The CMS name partly comes from its central piece, a huge superconducting solenoid described in Section 3.2.4 and able to produce a 3.8 T magnetic field in the detector, with a magnetic flux

density increased even more by the addition of the steel return yoke layers (red parts in Figure 3.4). This magnetic field is essential in the sense that it is able to deflect the charged particles which have been produced via the Lorentz interaction, therefore giving us an additional way to measure their charge and energy from the measurement of the curvature of the induced binding.

Finally, on the outside of the detector can be found the complete muon system and particularly performing in CMS. This subdetector is currently made out of three main sub-systems (the Drift tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs)), as explained in Section 3.2.5, and is responsible for the identification and measurement of the momentum of the muons produced by the collisions.

### Coordinate system

Before starting with the description of all these subdetectors, it is important to detail the coordinate system typically used within the CMS collaboration. As a convention, it has been decided to use a right-handed Cartesian coordinate system with the origin defined as the Primary Vertex (PV), and with an x-axis pointing towards the center of the ring, an y-axis pointing upwards and a z-axis pointing towards the Jura mountains (along the counterclockwise beam direction), as represented in Figure 3.5.

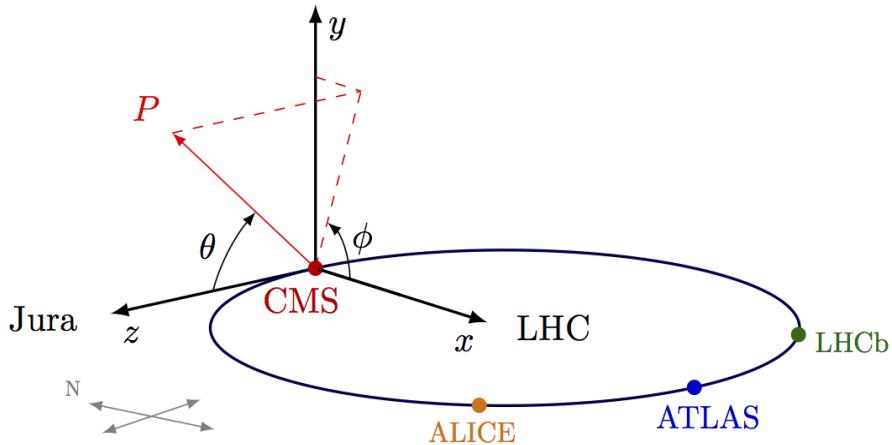


Figure 3.5: Schematic representation of the CMS coordinate system used by convention.

The  $\theta$  and  $\phi$  angles are then defined as the angles between the z and y axes and the x and y axes respectively and the pseudorapidity  $\eta$ , defined in Equation 3.4, a Lorentz invariant quantity under longitudinal boosts quite often used in the different analyses since the multiplicity of high energy particles is roughly constant in  $\eta$ .

$$\eta = -\log \left( \tan \left( \frac{\theta}{2} \right) \right) \quad (3.4)$$

#### 3.2.1 Tracker

The tracker is the innermost part of the CMS detector, is 5.4 meters long and has a diameter of 2.5 meters. Its main purpose is the reconstruction of the trajectories of charged particles issued from the primary and secondary interaction vertices in a quick and precise way in order to identify

them and measure their individual momentum.

Many challenges were faced when designing this system mainly because of the hard conditions provided by the LHC. First of all, at its nominal instantaneous luminosity, an average of 1000 particles are created after each bunch crossing, every 25 nanoseconds. The tracking system then needs to read all its channels extremely fast in order to be ready for the next bunch crossing. However, this fast electronics then needs some cooling to work optimally, which would in return increase the size of the tracker and therefore increase the interactions between the detector and the particles created (by multiple scattering, bremsstrahlung, photon conversion and nuclear interactions). This would affect the trajectory of the particles so a compromise had to be found between the velocity and size of the tracker. Finally, this device needs to be resistant to the extreme radiation environment for its expected lifetime of at least 10 years.

This device is then made out of silicon pixels and strips mainly because of the granularity, reading velocity and radiation hardness offered by such material. It has been set up on several different layers disposed in such a way to make the detector as hermetic as possible, as shown in Figure 3.6. A charged particle crossing the tracker will then leave a hit each time it crosses one of the silicon sensor, from which the track of the particle can be reconstructed using performing reconstruction algorithms that will be detailed in Chapter 4.

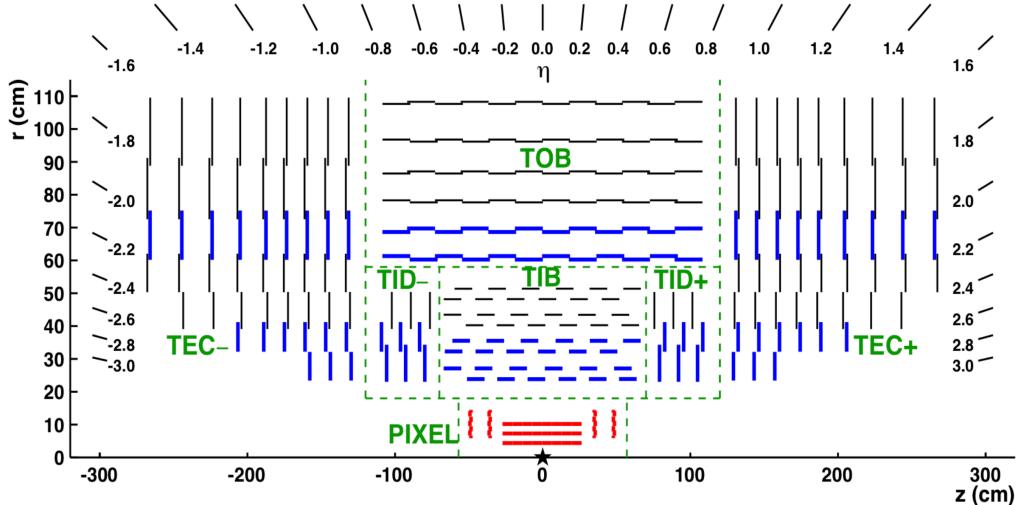


Figure 3.6: Schematic representation of the CMS tracker, for different pseudorapidity values and along the  $z$ -axis [87].

The presence of the magnetic field due to the solenoid can then help us estimate the momentum of the particle, since the Lorentz force applied on such particle will introduce a deviation directly proportional to its momentum. The radius of curvature of particles with a high momentum ( $> 100$  GeV) is really large but the density of pixels and the algorithm can still manage to estimate the momentum of such particles, even though the uncertainty associated will then be greater.

In particular, the tracker is made out of two distinct parts: the smaller and inner pixel detector and the larger silicon strip detector:

- The **pixel detector** is made out of three barrel pixel layers and two endcap disks for hermeticity (located at  $r = 4.4, 7.3$  and  $10.2$  cm of the PV), one on each side. In total, more than 60 millions pixels make up this detector, covering an area corresponding to  $\sim 1$  m $^2$ .
- The **silicon strip detector** on the other hand is composed of three different sub-systems, as

seen in Figure 3.6 and covers a total area of  $\sim 200$  m $^2$ . First of all, the Tracker Inner Barrel and Disks (TIB/TBD) add an additional 4 barrel layers and 3 endcap disks to the tracker system up to a distance  $r = 55$  cm to the PV of the interaction, using silicon micro-strip sensors parallel in the barrel and perpendicular to the beam axis in both endcaps. Then, the Tracker Outer Barrel (TOB) surrounds this first layer; having an outer radius of 166 cm and going up to 118 cm along the z-axis, it adds 6 layers to the inner tracking system. Finally, the Tracker EndCaps (TECs) are made out of 9 disks and complete the measurement of particles emitted along the z-axis and having a high pseudorapidity  $\eta$ .

The CMS tracker is extremely performing: we can see in Figure 3.7 that muons are indeed reconstructed with an efficiency higher than 99% for most of the pseudorapidity spectrum, even though we can observe that this efficiency drops at high  $\eta$  values mainly because of the reduced coverage provided by the pixel forward disks. The interactions between the hadrons and the tracking system is also a bit higher, which result in a lower reconstruction efficiency for such particles.

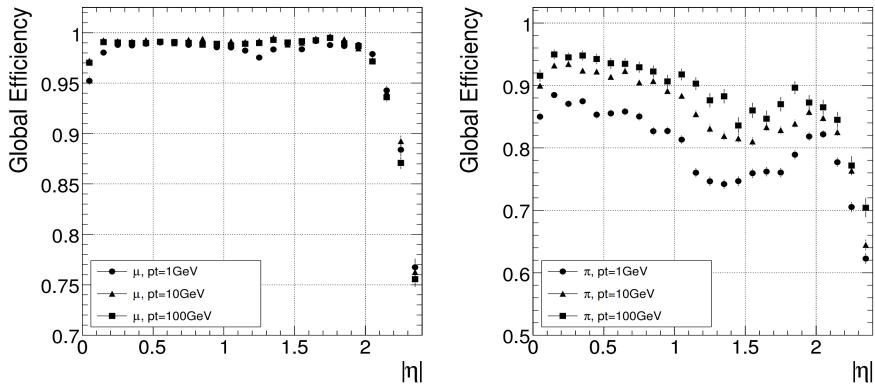


Figure 3.7: Tracker reconstruction efficiency of muons (on the left) and pions (on the right) for different pseudorapidities and particle momenta (1, 10 and 100 GeV) [86].

### 3.2.2 Electromagnetic Calorimeter (ECAL)

The Electromagnetic Calorimeter (ECAL) of CMS is a mostly homogeneous detector inside the solenoid but enclosing the tracker system that gives information about the energy of electrons and photons, both able to interact electromagnetically with its crystals.

The ECAL can also be divided into several sections: first of all, at pseudorapidities  $|\eta| < 1.479$  is found the barrel part of the ECAL (EB), made out of 61 200 lead tungstate ( $\text{PbWO}_4$ ) crystals, located at a radius of 1.29 meters of the beam pipe. Then, two endcaps, each made out of 7 324 of those crystals, increase the coverage of the detector up to  $|\eta| < 3$ , as shown in Figure 3.8, and the preshower completes the ECAL..

Its principle of action is simple and based on electromagnetic showers: when a particle such as an electron or a photon enters the ECAL, it will start to interact in different ways, depending on its nature. Photons will mainly produce pairs of electrons and anti-electrons, while the electrons themselves tend to emit additional photons by bremsstrahlung effect. This results in a chain reaction during which the incident particle will give most of its energy to the detector itself, energy measurable using photodetectors and photomultipliers. This effect is known as an electromagnetic shower, is represented in Figure 3.9 and is usually characterized using the so-called radiation

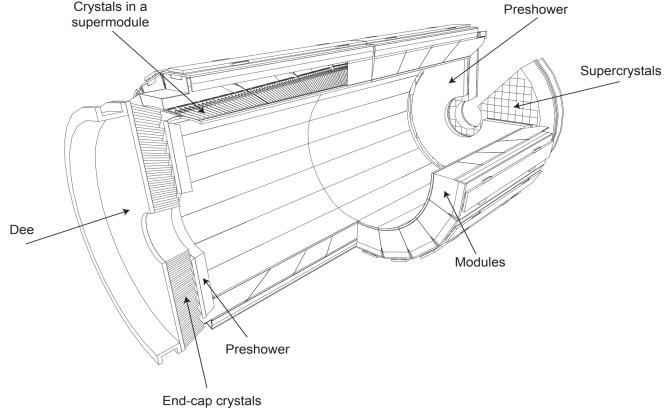


Figure 3.8: Schematic representation of the sub-systems of the CMS ECAL [86].

length  $X_0$ , the mean distance over which a high-energy electron loses all but  $1/e$  of its energy, then determining the total length of interaction of a particle in the ECAL.

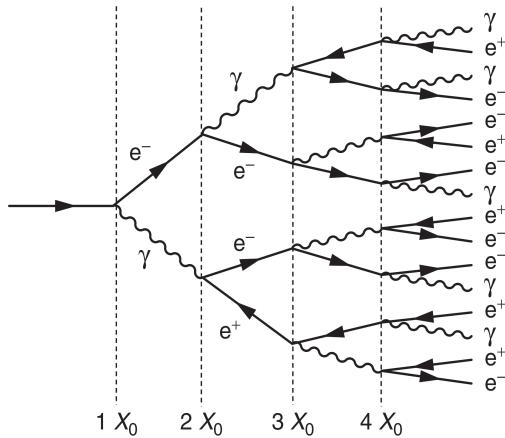


Figure 3.9: Schematic representation of a typical electromagnetic shower and the radiation length  $X_0$  concept [86].

The high density, short radiation length and low scintillation decay time (smaller than the bunch spacing of 25ns) of the PbWO<sub>4</sub> crystals make them perfect candidates towards a compact ECAL in CMS. These crystals do have some drawbacks as well, mainly their relative fragility when it comes to radiation, and the dependence on the temperature of their response. Indeed, a cooling system had to be built in order to keep the huge detector under temperature variations lower than 0.1° to avoid eventual fluctuations in the response of the crystals.

These crystals, which had to be grown individually in laboratories, measure 2.2x2.2x23cm in the barrel and 3x3x22cm in the endcaps, cover a solid angle equal to  $(\Delta\eta, \Delta\phi) = (0.0174, 0.0174)$ , and have therefore a length corresponding to around 26 radiation lengths, more than enough to stop even the most energetic particles. Since the light output of such crystals is quite low (only around 4.5 photoelectrons per MeV of energy), they have to be connected to both avalanche photodiodes and vacuum phototriodes in order to multiply the signal measured. Finally, they have been mounted using a specific installation, slightly tilted with respect to both  $\phi$  and  $\eta$  in order to remove any possible gap between two adjacent crystals.

The typical energy resolution of the ECAL installed at CMS is given by Equation 3.5 [86], accounting for several different effects, such as the stochastic nature of the observe scintillation, the electronics and PU noise and the calibration and detector non-uniformity uncertainty.

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12\%}{E}\right)^2 + (0.30\%)^2} \sim \frac{3 - 10\%}{\sqrt{E/\text{GeV}}} \quad (3.5)$$

Finally, a preshower layer has been set up in the fiducial region ( $1.653 < |\eta| < 2.6$ ) of the endcaps, where the angle between the two photons coming from the decay of neutral pions  $\pi^0 \rightarrow \gamma\gamma$  is small enough to be misidentified as individual photons. This detector has then been installed in order to reduce the possible misidentification of such events and to help with the identification of electrons against minimum ionizing particles. It is made of a lead layer able to initiate the electromagnetic shower process, followed by two layers of silicon strips for the actual measurement.

### 3.2.3 Hadronic Calorimeter (HCAL)

We know that charged hadrons do lose energy in a continuous when they traverse matter due to the ionization process and that all the hadrons strongly interact with the nuclei of any given medium. These principles are actually used in order to measure the energy of the hadrons produced by the LHC collisions using the Hadronic Calorimeter (HCAL) sub-system of CMS.

In this case as well, showers of particles due to a chain reaction are expected since the primary hadronic interaction will produce several additional hadrons, themselves interacting even more with the medium while losing energy. This kind of hadronic showers is characterized by the  $\lambda$  parameter, the nuclear interaction length, defined as the mean distance between two interactions of relativistic hadrons. The nuclear interaction length  $\lambda$  is usually much larger than the radiation length  $X_0$ , resulting in a HCAL typically much larger in size than the ECAL.

A typical HCAL setup consists in alternating thick and high-density layers of absorber material, in which the showers can develop, and thin layers of active material used for the actual detection by sampling the energy deposition. This measurement is usually much less precise than the measurement provided by the ECAL, mostly since  $\pi^0$  decaying into photons can appear in these showers, leading to an electromagnetic component of the shower that cannot be measured, and because around 30% of the incident energy is usually lost due to nuclear excitation and break-up effects [84]. In this case, the energy resolution can therefore be expressed using Equation 3.6.

$$\frac{\sigma_E}{E} > \frac{50\%}{\sqrt{E/\text{GeV}}} \quad (3.6)$$

In CMS, the HCAL, represented in Figure 3.10, is also divided into a barrel (HB), radially constrained between radii values of 1.77 meters (outer radius of the ECAL) and 2.95 meters (inner radius of the solenoid), and two symmetrical forward regions (HF) extending the pseudorapidities coverage from  $|\eta| = 3$  to  $|\eta| = 5.2$  and located at a distance of 11.2 meters to the PV. A final part composing the HCAL is the so-called Hadron Outer (HO), which has to be put outside of the solenoid in order to increase the amount of shower absorber material of the HCAL and therefore the effective nuclear radiation length  $\lambda$ .

The HCAL barrel uses 36 identical azimuthal wedges as absorber, placed along the beam axis in such a way that the eventual cracks between them is smaller than 2 mm. The total absorber

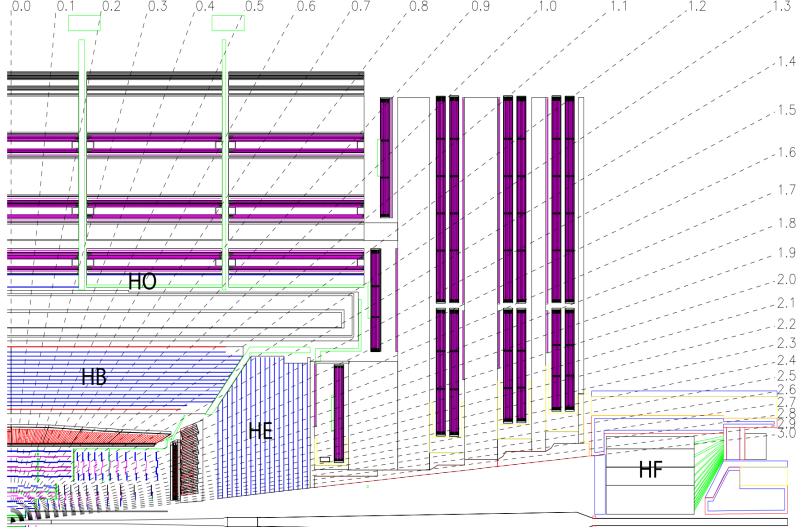


Figure 3.10: Schematic representation of the HCAL sub-system in CMS [86].

thickness at a  $90^\circ$  incidence angle is equal to only  $5.8\lambda$ , which explains why the HO had to be added in order to increase this value to make sure to slow down and completely stop even the most energetic hadrons. The active medium of the HB is made out of 70 000 tiles able to collect the scintillation light, using the wavelength shifting fibre concept to measure the energy of the hadrons. The HF on the other hand are using a Cerenkov-based radiation-hard technology to make the measurements required.

The barrel covers  $|\eta|$  regions up to 1.3, while the coverage up to  $|\eta| < 5.2$  is given by two endcaps on each side of the detector, placed in such a way to minimize the eventual cracks between the HB and the two HF. Finally, the HO, built in order to ensure adequate sampling depth at low pseudorapidities values, actually uses the solenoid itself as additional absorber material, since it is placed a bit outside of this coil. Its shape is constrained by the muon system and the mean fraction of recovered energy from the HO has been estimated to be equal to 0.38% for 10 GeV pions and up to 4.3% for 300 GeV pions, as shown in Figure 3.11.

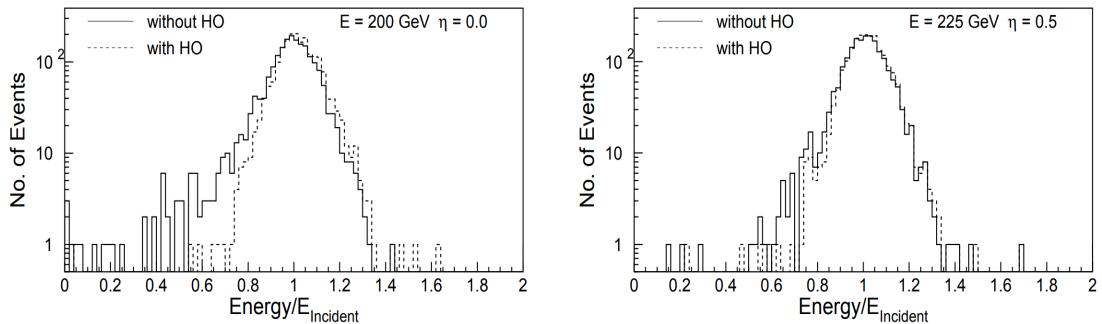


Figure 3.11: Distribution of the measured energy scaled to the incident energy for pions with incident energies of 200 GeV at  $\eta = 0$  (on the left) and 225 GeV at  $\eta = 0.5$  (on the right), with and without the inclusion of the HO in the HCAL system [86].

### 3.2.4 Solenoid

The central piece of CMS is its extremely large (12.5 meters of length and 6 meters of diameter) and heavy (220 tons) superconducting solenoid able to produce a 3.8T magnetic field, storing when active a huge energy of 2.6 GJ. It is the largest magnet of its type ever constructed, therefore allowing the tracker, ECAL and HCAL calorimeters to be placed inside the coil, resulting in a detector that is, overall, quite compact compared to detectors of similar weight.

The magnetic field produced by this coil is extremely useful since it allows to measure quite precisely the charge and the momentum of the different charged particles interacting with the detector just by measuring the curvature of their track, according to the Lorentz equation 3.7. This solenoid has been designed to reach a momentum resolution  $\Delta p/p \sim 10\%$  at  $p = 1$  TeV.

$$\vec{F} = \frac{m \vec{v}^2}{R} = q \vec{E} + q \vec{v} \times \vec{B} = q \vec{v} \times \vec{B} \quad (3.7)$$

The hoop strain  $\epsilon = 130$  MPa applied throughout this solenoid is quite large compared to the strain applied on other previous detectors and this had to be taken into account during the conception of this magnet. It has then been designed in such a way that a large fraction of the CMS coil has a structural function, dividing the strain between the layers of the magnet and the support of the coil itself ( $\sim 30\%$ ). At the end, the conductor of this solenoid, made from a Rutherford-type cable combined with aluminium, is mechanically reinforced with an aluminium alloy.

The coil of the magnet is then completed with a huge steel yoke return system, as seen in Figure 3.12, made out of 6 endcap disks and 5 barrel wheels, weighting in total more than 12 000 tons and therefore accounting for most of the weight of CMS. This system is composed of many steel blocks up to 620 mm thick combined and actually also serves as the absorber plates of the HO and muon detection system that will be described in the next section.



Figure 3.12: Picture of the solenoid system of CMS being setup in the assembly hall.

Finally, a two pumping stations system has been put in place in order to setup a vacuum as strong as possible inside the  $40 \text{ m}^3$  volume of the coil cryostat and an helium refrigeration plant has been installed near the site of the detector, able to cool down the solenoid up to 4.5 K, giving a 2 K security margin with respect to the critical field of the superconducting coil. All these systems were extensively tested on the surface during the summer of 2006, before lowering down the complete solenoid in the experimental cavern where it now stands.

### 3.2.5 Muon system

The muon detection in a detector is extremely useful since many interesting processes are expected to produce such particles. Their detection and the correct measurement of their main properties such as their position and momentum is therefore crucial in most of the analyses performed. Detecting muons is at the end of the day quite easy, as we will see, and the data extracted from them is usually more reliable than the one obtained from electrons since muons are less likely to be affected by the inner parts of the detector, such as the tracker, because of their low interaction cross section with usual matter.

The muon system of CMS is actually made out of three different gaseous sub-subsystems combined in order to perform a reconstruction as precise as possible of such particles over the entire kinematic range of the LHC. These different muon chambers systems do share some characteristics: they mostly have to be cylindrical, because of the geometrical shape of the inner systems of CMS and they have to be reliable and cheap, since they cover a total area of more than 25 000m<sup>2</sup>.

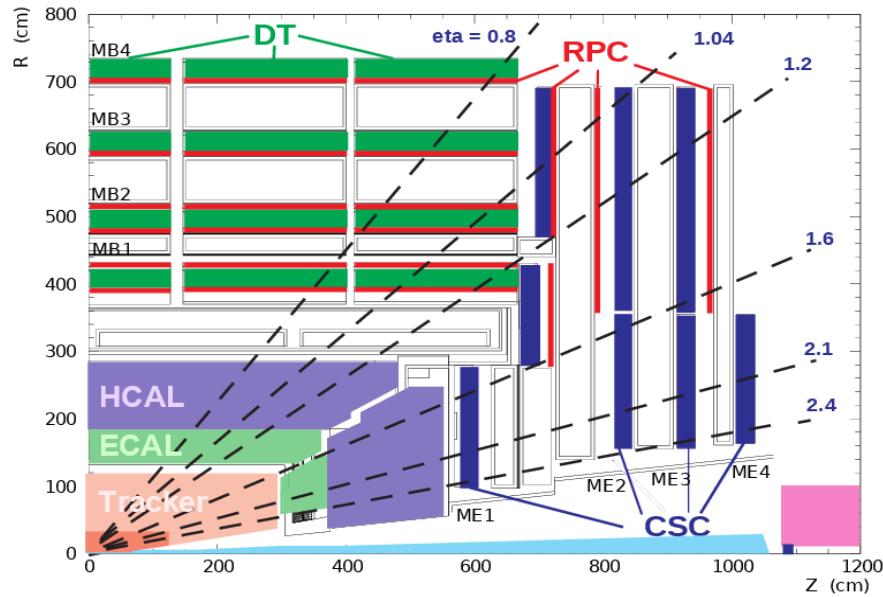


Figure 3.13: Geometrical repartition along the z-axis of the different muon chambers in CMS [88].

Let's now describe in details each category of muon chambers used in the different pseudorapidities areas, as shown in Figure 3.13, in order to form as muon system as hermetic as possible.

#### Drift tubes (DTs)

First of all, in the barrel region, where the flux of muons is low and where the magnetic field is mostly uniform and low as well, the **Drift tubes (DTs)** have been installed. This system covers the  $|\eta| < 1.2$  area and has been divided into 4 different layers, each containing a number of stations optimized in order to provide a full coverage of the  $\theta$  angles, a good efficiency for the muon hits reconstruction into a single track and a good rejection of eventual background hits. This distribution of the DTs is represented in Figure 3.14.

The DT system is made out of 172 000 sensitive wires able to collect the residuals charges left by the ionization tracks of muons through the 250 chambers installed. The system has been set up

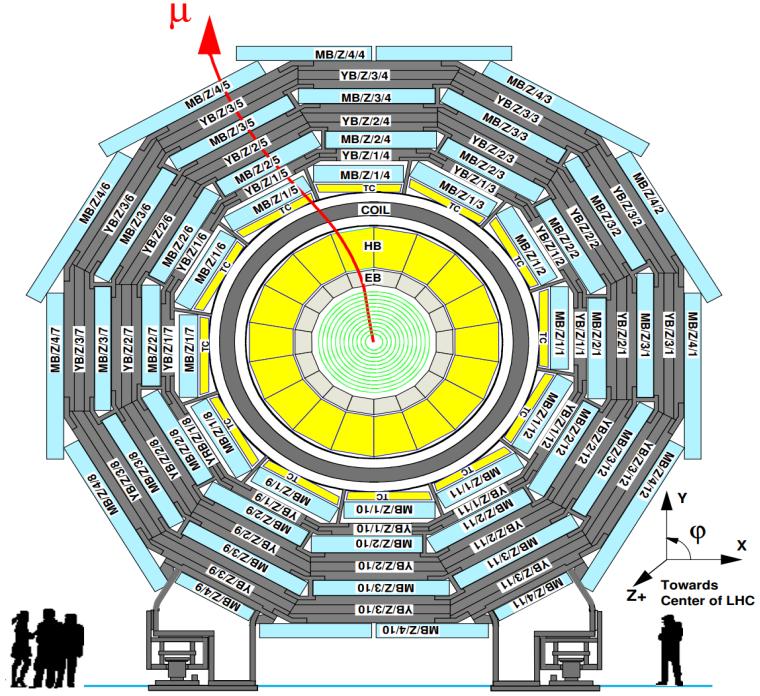


Figure 3.14: Lateral geometrical division of the different DT chambers in one of the 5 wheels of CMS [86].

in such a way that the maximal drift of any charge is lower than 21mm, corresponding to a drift time of 380ns in the gaseous chambers made out of 85% of Ar and 15% of CO<sub>2</sub>, a value small enough to produce negligible occupancy in the different wires and to avoid the need of multi-hits electronics. Redundancy of the DTs provided by the installation of multiple layers is extremely important, mainly to reduce the backgrounds coming from eventual neutrons or photons, whose rate is actually much larger than the one obtained from prompt muons.

### Cathode Strip Chambers (CSCs)

In the two endcap regions, where the muon rates and the background levels are much larger and where the magnetic field is large and non-uniform, a different system had to be installed. First of all, the **Cathode Strip Chambers (CSCs)**, multiwire proportional chambers providing a fast response while being resistant to the radiation, are able to identify muons in a  $0.9 < |\eta| < 2.4$  region (in the  $0.9 < |\eta| < 1.2$  region, muons cross both DTs and CSCs while in the  $1.2 < |\eta| < 2.4$  area, muons cross between 3 and 4 CSCs only).

This sub-system is made out of 540 different chambers in total, all perpendicular to the beam pipe. The sensitive plates of this sub-system are made out of 2 million wires, cover about 5000m<sup>2</sup> and the total gas volume included in such chambers is equal to about 50m<sup>3</sup>.

### Resistive Plate Chambers (RPCs)

Finally, some **Resistive Plate Chambers (RPCs)** have been added to the barrel and to the endcap regions in order to cope with the uncertainty associated with the eventual background rates

and with the (in)ability of the previous muon system to identify unequivocally the correct bunch crossing when the LHC is running at full luminosity. Indeed, the time resolution of the DTs of 380ns is way larger than the bunch spacing in the LHC, while a RPC is capable of tagging an ionizing event in less than 25ns, make it an ideal candidate to trigger the event.

The RPCs are double-gap chambers operated in avalanche mode to ensure good operation at high rates and they are able to produce a fast response with good time resolution, even though its position resolution is worse than the one obtained with DTs or CSCs. The RPCs are also useful in the sense that they can help to resolve ambiguities when attempting to construct tracks from multiple hits in a chamber.

Finally, the different features of the three muon sub-systems used by the CMS detector and just described are summarized in Table 3.2.

Muon sub-system	DT	CSC	RPC
$ \eta $ coverage	0.0-1.2	0.9-2.4	0.0-1.9
Stations	4	4	4
Chambers	250	540	480 (barrel) 576 (endcaps)
Readout channels	172 000	266 112 (strips) 210 816 (anode channels)	68 136 (barrel) 55 296 (endcaps)

Table 3.2: Comparison of the three main sub-systems currently used by CMS in order to identify muons [89].

Another advantage of the muon system such as the one built in CMS is that it can also directly be used by the trigger system, which will be described in Section 3.2.6, independently of the rest of the detector and in addition of being able to detect, identify and measure several properties of muons crossing it.

The muon reconstruction efficiency obtained by the muon system strongly depends on the pseudorapidity value of the muon considered, along with its transverse momentum, as shown in Figure 3.15. In this figure, we can also see that several different kinds of muons can be defined, such as the **standalone muons**, defined using only the data coming the muon system and the **global muons**, defined using both the information coming from the muon system and the tracker. This distinction will be detailed when discussing about the muons reconstruction in Section 4.3.1).

Taking advantage of the Long Shutdown (LS) 2, a new muon system is currently being installed in the experimental cavern: the so-called **Gas Electron Multipliers (GEMs)**, placed in the endcaps, where the radiation and event rates are the highest. This new subdetector will provide additional redundancy and measurement points to the current system, therefore allowing a better muon track identification and reconstruction and a wider coverage in the very forward region.

The first 144 chambers of the GEM sub-system, filled with a mixture of Ar and CO<sub>2</sub> and where the primary ionization due to incident muon is expected to happen, are currently being installed in the first disk of both endcaps (cf. Figure 3.16), while the rest will be set up during the next LS expected in 2024, before the phase II of operation of the LHC.

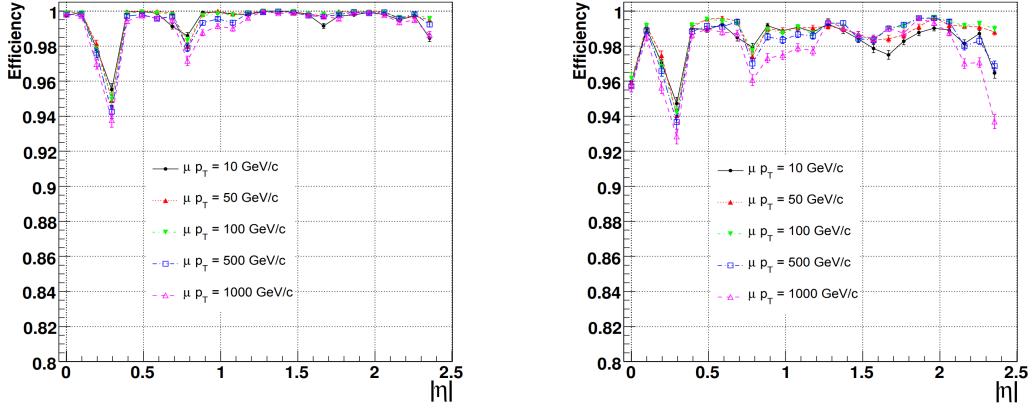


Figure 3.15: Muon reconstruction efficiency with different  $p_T$  and  $\eta$  values, considering only the muon system (on the left), and the combined information from the muon system and the tracker (on the right) [86].

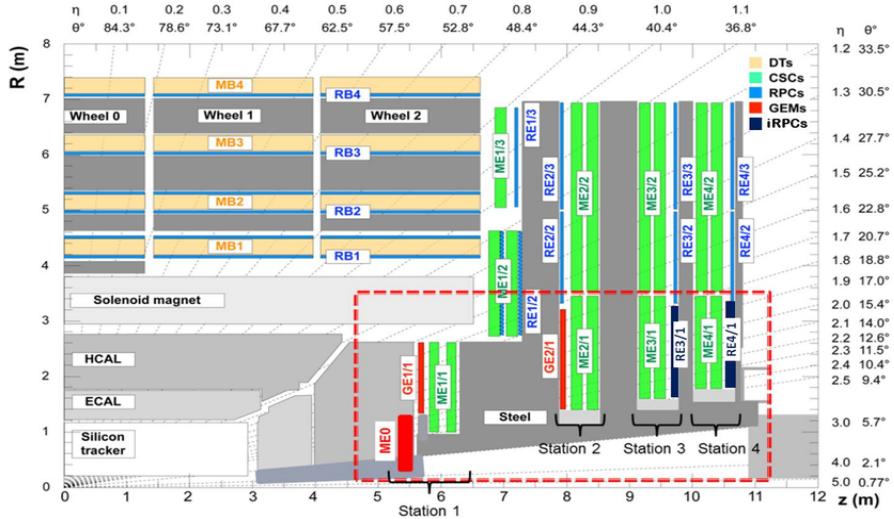


Figure 3.16: Location of the new GEM muon subsystem currently being installed in the very-forward region of CMS [86].

### 3.2.6 Trigger system

The CMS experiment is facing a data acquisition problem since the collision rate delivered by the LHC (one bunch crossing each 25ns, leading to an impressive rate of collisions of 40 MHz) is much larger than the data acquisition rate currently achievable by nowadays' electronics (around 1kHz only, more or less equivalent to 1Gb of data per second). It is therefore impossible to store and process all the collisions provided by the LHC ; instead, a selection needs to be made in order to select and record only the most interesting events.

A system, called the trigger system, has therefore been put in place in order to select extremely quickly 1 kHz of interesting events out the 40MHz. This system is based on two different levels: first of all, the Level-1 Trigger (L1), a hardware set of electronics selecting around 100kHz of data, followed by the High-Level Trigger (HLT), a software layer improving the selection even more. Let's now give a bit more details about these two levels of trigger.

## Level-1 Trigger (L1)

The Level-1 Trigger (L1) is the first level of trigger, based directly on hardware. In order to maximize its efficiency, it is implemented in different subsystems of the detector (on the calorimeters and muons system) and in the service cavern, just next to the detector, so that the electric signals do not have to travel large distances, therefore saving a few precious nanoseconds of decision time.

This trigger, whose architecture is represented in Figure 3.17 gets new data each 25ns and today's electronics is not fast enough in order to deal with such a massive input of data, so an ingenious systems of buffers had to be put in place in order to put in line several events before analyzing all of them at once, using only basic segmented data provided by the detector.

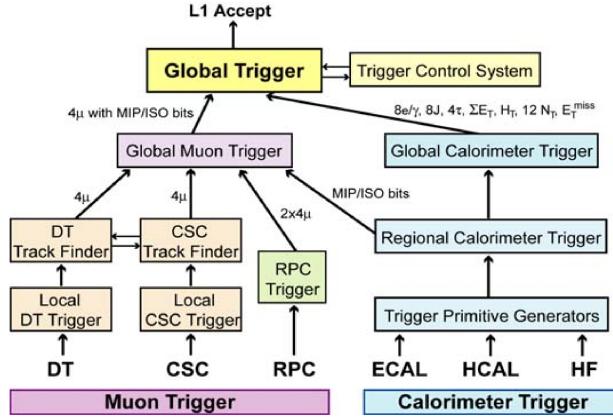


Figure 3.17: Architecture of Level-1 Trigger (L1) of CMS [86].

## High-Level Trigger (HLT)

On the other hand, the High-Level Trigger (HLT) does get access to the complete read-out data of the detector, since the rate has already been strongly reduced by the L1 Trigger, allowing it to perform complete calculations such as the ones that will be later performed offline. Since this is a software based layer and the decision time is not as critic any more, it runs on a farm of computers on the surface and is in constant evolution, getting constantly improvements and updates in order to select and reconstruct in a better and more efficient way interesting data from the collisions.

The HLT trigger is divided in several so-called **trigger paths**, especially developed in order to select a target object or physics process, usually with some  $p_T$  requirement in order to reduce the statistics collected by any given trigger, to avoid passing the bandwidth limit of 1kHz. These triggers paths can then for example require the presence of one or two leptons, a jet whose  $p_T$  is higher to a certain threshold or the presence of Missing Transverse Energy (MET) in the event. If the selection implied by a trigger is too large, then it needs to be **prescaled**, meaning that its statistics is artificially reduced by forcing it to select one event every N, so that it does not use too much bandwidth: this is typically the case for trigger paths with low  $p_T$  requirements, as the ones used for the fake rate calculation explained in Section 5.6.4.

### 3.2.7 Data acquisition system

Finally, the CMS Data Acquisition System (DAQ), whose architecture is represented in Figure 3.18, has been designed to collect and analyze the data information at the nominal collision rate of 40MHz and is feeded directly from the Level-1 Trigger (L1). This means that it has to be able to read a flux of data of the order of 100kHz ( $\sim 100\text{Gb/s}$ ) coming from approximately 650 different sources at once, while providing enough computing power for the HLT to be able to reduce this rate by a factor  $\sim 100$ , while keeping some resources available for other tasks.

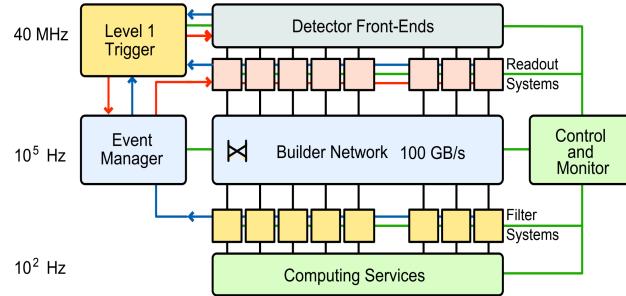


Figure 3.18: Architecture of CMS Data Acquisition System (DAQ) system [86].

The DAQ is indeed also in charge of performing additional tasks, such as the generation of the Data Quality Monitoring (DQM) information resulting from online event processing in the HLT, the transfer of the data from local storage at the CMS site to mass storage in the CERN data centre at the Meyrin site and the operation of the Detector Control System (DCS) system, ensuring the correct operation of the detector and a high quality data taking at all times.

# Chapter 4

## Objects reconstruction

As we just saw, the CMS detector is made out of different layers, each able to convert the interaction of the particle with this detector into electronic signals that can be measured and stored. However, this signal collected is made out of raw information coming from different kind of subsystems and an algorithmic strategy then needs to be put in place in order to read all these separate signals and to combine them to extract some useful data, such as the number of particles produced by the collision along with their energy, charge and direction. Producing this kind of data is essential for all the offline analyses which usually rely on these high-level physics objects to make precision measurements or search for new physics.

The algorithm able to combine this raw data and to produce useful objects and variables is the so-called Particle Flow (PF) algorithm [90], which will be described in Section 4.1. Then, a particular focus will be given to the definition and reconstruction of different objects of our particular analysis, such as the electrons and muons (Section 4.3), the jets (Section 4.4), the Missing Transverse Energy (MET) (Section 4.5) and the top reconstruction (Section 4.6) of the different  $pp$  collisions recorded.

### 4.1 Particle Flow (PF) algorithm

The Particle Flow (PF) is an algorithm aiming to combine in the best way possible all the information coming from the different parts of the CMS detector in order to identify and reconstruct the hundreds of new particles produced by each  $pp$  collision provided by the LHC. This reconstruction can be divided into two main steps: first of all, the data coming from the different subsystems of the detector is read in order to identify and measure the properties of some basic stable objects, such as leptons, photons and hadrons. Then, more complex calculations are performed to identify eventual unstable particles, jets from the hadronization of quarks and to compute complex variables such as the leptons isolation and the Missing Transverse Energy (MET).

The most basic elements used by this algorithm for the reconstruction of high-level physics objects are the tracks of charged particles, the clusters of energy left in the two calorimeters and the hits recorded in the muon chambers. For this algorithm to be as efficient as possible, the detector has been carefully designed: a magnetic field as large as possible and a small calorimeter granularity are indeed crucial in order to separate efficiently charged and neutral particles, and the tracker was designed to be as efficient and small as possible to have the smallest material budget possible in front of the calorimeters.

The way the different particles produced in each collision are identified is quite easy to summarize

and is represented in Figure 4.1. Basically, the different kind of particles produced are going to interact with different parts of the detector and combining the information given by the tracker and the rest of the subsystems then allows to unequivocally identify each particle. This is usually done in a specific order in order to be as efficient as possible:

1. First of all, the most energetic **Primary Vertex (PV)** is identified by taking into account the Pile Up (PU) and by assigning the different tracks to the different  $pp$  collisions happening during a single bunch-crossing.  
All the particles originating from less energetic PV or from a secondary vertex of interaction are typically ignored and the corresponding hits in the tracker left by such particles can therefore be removed, leaving less hits available for the clustering algorithm later on, allowing for a more efficient reconstruction of the following objects.
2. Then, **muons** are the easiest particles to identify since they are at first order the only particles leaving many hits in the muon chambers placed on the outside of the detector. Each muon identified is associated to its track in the tracker, where all the hits matching a muon can therefore be removed.
3. **Electrons** do have a charge, so they are visible by the tracker and by the ECAL, where they are going to produce an electromagnetic shower. Identifying electrons is a bit more challenging than muons because of their associated bremsstrahlung emission of photons that need to be attached to the original electrons to avoid any double counting. All the tracker hits corresponding to electrons are also removed after identification.
4. **Charged hadrons** also leave hits in the tracker and some energy deposits in the ECAL, but mostly in the HCAL, so they are easy to identify as well as a third step, using the latest hits available in the tracker.
5. **Photons** are on the other hand neutral particles, so they do not leave any hits in the tracker. They then appear as some energy deposits in the ECAL for which no corresponding tracker track can be associated.
6. Finally, **neutral hadrons** can be identified as particles leaving some energy mostly in the HCAL for which no corresponding tracker track has been found as well.

We will now study in a bit more details the reconstruction method applied in order to reconstruct the main objects of this analysis, i.e. leptons, jets, MET and top quarks.

## 4.2 Primary Vertex (PV) definition

Different kind of vertices originating from a single  $pp$  collision can usually be defined, as shown in Figure 4.2: the **Pile Up (PU) vertices**, corresponding to the different simultaneous collisions of a single bunch-crossing of the LHC, the **primary vertex**, usually assumed to be the most energetic PU vertex and the only vertex considered in most of the physics analyses, and the **secondary vertices**, due to the eventual presence of long lived particles, decay chains or jets.

The first task of the PF algorithm is to identify all these vertices of the interaction. This is done by considering all the tracker tracks, by clustering them together and by performing fits to determine the likelihood these tracks originated from a common vertex. The reconstructed vertex with the largest total  $p_T$  is then assumed to be the PV, as it is considered to be the origin of the most interesting  $pp$  collision from which many different tracks are emitted.

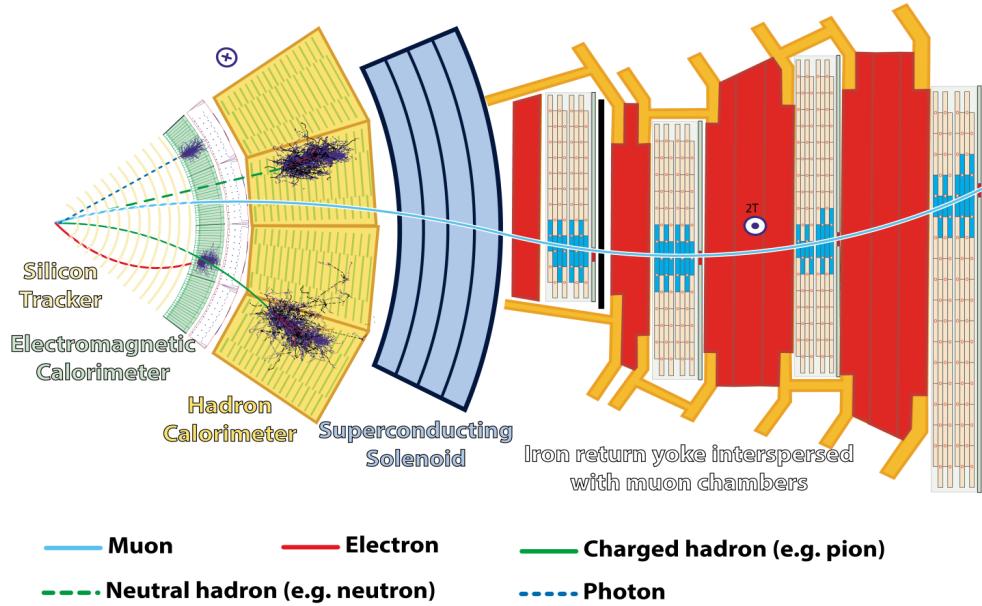


Figure 4.1: Transverse section of CMS showing the different tracks expected by different kind of particles in the detector.

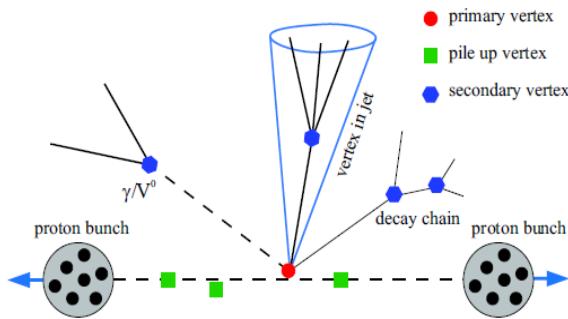


Figure 4.2: Different kind of vertices typically observed in a  $pp$  collision in the LHC.

## 4.3 Leptons reconstruction

Different kind of leptons are typically produced by a  $pp$  collision. The muons and the electrons can be quite easily identified, mainly because their lifetime and velocity is high enough, meaning that they are not expected to decay inside the CMS detector, so they can be directly identified. Taus on the other hand are a bit trickier to deal with because they usually decay inside of the beam pipe itself,  $\sim 35\%$  of the time to electrons and muons and  $\sim 65\%$  of the time to hadrons. However, since our analysis does not consider taus directly but only the leptons originating from the their decays, the details of their reconstruction will not be explained in this section.

### 4.3.1 Muons

Muons are the first leptons to be reconstructed by the PF algorithm since, at first order, they are the only particles expected to reach the muon chambers, resulting in their easy identification.

The typical signature of a muon consists in several hits in the silicon tracker forming a track since muons do have an electric charge, associated with several hits in the muon chambers, electronic signals coming from the wires and strips of these chambers due to the gas ionization induced by the passage of these charged particles. Muons only deposit a negligible amount of energy within the two calorimeters since their interaction cross section is quite low for their full range of energies, going from a few hundreds MeV up to a few TeV.

The data coming from the different subsystems of CMS are then combined and fed into three different PF algorithms, able to reconstruct different kind of muons.

### **Standalone muons**

These are muons reconstructed using only the hits observed in the muon system without trying to relate this data to the tracker hits. Basically, the PF algorithm looks in this case at the eventual hits left in the DTs, CSCs and RPCs and tries to reconstruct a vector of trajectory in each case using a Kalman Filter (KF) filter [91]. These segments are then combined in the best statistical way possible in order to form a candidate track for each muon of the event, by extrapolating the innermost vectors are propagated to the layer surface of the next chamber and by comparing it with the local track segment. The trajectory parameters are then computed and the process continues until reaching the outermost chamber, before being reversed in order to estimate the innermost track parameters as well.

To limit the possibility of misidentification due to showering of cosmic rays, the tracks need to pass some quality criteria in order to be considered valid: for example, at least two hits need to be measured for the fit to be performed, one of them coming from either the DTs or CSCs, in order to remove fake segments contamination due to combinatorics. Additional constraints by checking the extrapolation of the trajectory to the point of closest approach to the beam line also allow to reduce this contamination.

Candidates reconstructed as standalone muons typically have a worse momentum reconstruction and are more sensitive to cosmic muons contamination.

### **Tracker muons**

The algorithm able to reconstruct such muons on the other hand is able to propagate tracks identified in the inner silicon tracker (having a momentum  $p > 2.5$  GeV and  $p_T > 0.5$  GeV) to the muon system itself in order to try and find corresponding segments in the different muon chambers (these tracks are therefore said to be built *inside-out*).

These muons are particularly efficient for less instrumented regions of the detector and for the low  $p_T$  end of the energy spectrum but they are also quite contaminated with fake muons tracks, since a single hit in any of the muon chambers is enough for the candidate to be considered a valid tracker muon, even though hadron shower remnants can for example quite easily reach the innermost muon station. The momentum assigned to such muons is the same as the one measured by the silicon tracker track itself.

## Global muons

Finally, these muons are built *outside-in* since they are obtained matching standalone muon tracks with independently reconstructed tracks coming from the tracker itself (of course, in order to avoid any double counting, global muons and tracker muons that share the same tracker track are actually merged into a single candidate).

This category of muons presents the advantage of being less sensitive to the muon misidentification rate than tracker muons since it uses the information from more than one muon chamber. The  $p_T$  measurement in this case is also improved (especially at high  $p_T$ ,  $> 200$  GeV) by exploiting the information from both the inner tracker and the muon system, while at low momentum, the best momentum resolution for muons is obtained from the inner silicon tracker directly.

Using this strategy, about 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed either as global or tracker muons [89], as seen in Figure 6.3.

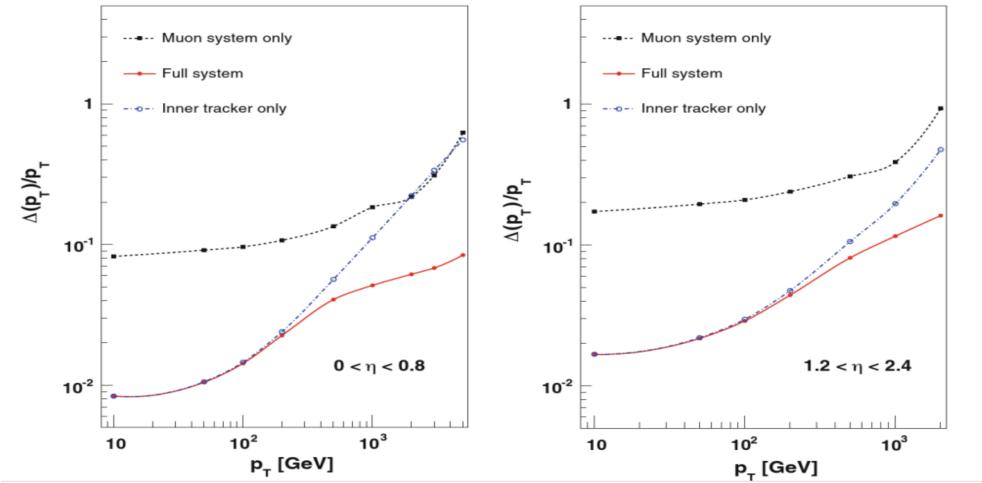


Figure 4.3: Muon  $p_T$  resolution obtained in simulation in the barrel (on the left) and endcap (on the right) for different kind of reconstructed muons [92].

Once reconstructed, candidates are required to pass some selection criteria and are then fed to the actual PF algorithm itself to start the global reconstruction of the event. This selection consists mainly in applying identification and isolation (evaluated relative to its  $p_T$  by summing up the energy in geometrical cones of radius  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$  surrounding the muon, as shown in Figure 4.4) criteria in order to enhance the purity of the reconstructed muons by rejecting muons coming from W+jets and QCD processes.

Different Working Points (WPs) can then be defined for the offline analyses, from loose to tight, in order to reject more or less contamination from misidentified leptons, keeping in mind that a tighter selection will also have an impact on the efficiency of the selection. The actual selection applied to the global muons used in this particular analysis will be detailed in Chapter 6.

### 4.3.2 Electrons

Electrons are reconstructed by combining the tracker tracks and the several clusters of energy deposited in the ECAL by the electromagnetic showers appearing due to the interaction between the electron and the crystals composing this subdetector.

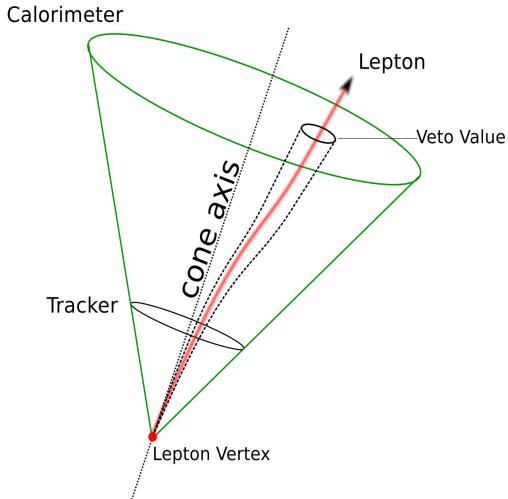


Figure 4.4: Lepton isolation cone typically used to reduce the background levels from SM processes.

It is usually a bit harder to reconstruct electrons than muons mainly because electrons do interact with the tracker and this interaction therefore needs to be modeled to understand the exact behaviour of such particles: this interaction is for example responsible for the emission of secondary bremsstrahlung photons crashing into the ECAL but not coming from the PV. In fact, it is estimated that in CMS between 33% and 86% of the energy of an electron is actually radiated before it reaches the ECAL, depending mostly on its pseudorapidity [93]. In order to measure precisely the energy of an electron, all the photons emitted by bremsstrahlung before reaching the ECAL (usually, along the  $\phi$  axis because of the deviation implied by the solenoid) then need to be collected as well and associated to the correct electron of the event.

The actual PF reconstruction of electrons is performed in different steps:

1. A **clustering algorithm** is first of all defining the so-called **Super Cluster (SC)**. Its goal is to reconstruct the particle showers individually by identifying a seed crystal for the cluster, defined as the crystal collecting the most energy, since the energy deposited in the ECAL is usually spread into several different crystals because of the electromagnetic shower effect discussed in Section 3.2.2 and because of the bremsstrahlung emission of photons due to the interaction with the tracker. The algorithm therefore searches for eventual crystals around this seed whose energy detected would be superior to  $2\sigma$  of the electronic noise and matching some quality criteria ( $E_{\text{seed}} > 230$  MeV in the barrel,  $E_{\text{seed}} > 600$  MeV and  $E_{\text{seed}}^T > 150$  MeV in the endcaps).

The excited contiguous crystals found are then grouped into clusters, themselves considered candidates for the final global cluster, the SC, if their energy is higher than another given threshold ( $E_{\text{cluster}} > 350$  MeV in the barrel,  $E_{\text{cluster}}^T > 1$  GeV in the endcaps) [93]. The SC energy is then given by the sum of the energies of all its constituent clusters, while its position is calculated as the energy-weighted mean position of the different clusters.

2. Once the SC identified, **electron tracker tracks** are reconstructed using a procedure a bit different than the usual KF reconstruction method for all the tracks of the silicon tracker [91] because of the large radiative losses for electrons in the tracker material.

This reconstruction is known to be very time consuming, so a good identification of potential electron seeds has to be performed as the method efficiency greatly relies on this first

identification. Two different strategies can be used to perform this seeding (even though the electron seeds found using the two algorithms are usually combined afterwards):

- The **ECAL-based seeding** relies on the information obtained for the SC energy and position in order to estimate the electron trajectory to find compatible hits in the tracker. This can be done knowing that the electron or positron is moving according to an helix in the magnetic field of the detector. This seeding is mostly optimized for isolated electron in the  $p_T$  range relevant for the Z and W decays.
- The other way to proceed is the **tracker-based seeding**, based on tracks reconstructed using the usual KF algorithm and looking for matches within the possible reconstructed SC. This seeding is mostly suitable for low  $p_T$  electrons and also performs quite well with electrons inside jets.

Once the seeds identified, the identification of tracks can begin. First of all, the gathering of compatible of hits from the different seeds is done using a dedicated modeling of the electron energy loss and a combinatorial KF algorithm allowing to construct possible tracks when compatible hits are found. The compatibility matching between the predicted and found hits is usually chosen to be quite loose in order to maintain a good efficiency even in case of bremsstrahlung emission.

Finally, once the hits are collected, a Gaussian Sum Filter (GSF) fit is performed to estimate the different track parameters by reconstructing the layer-to-layer propagation of electrons in the tracker. A mix of Gaussian distributions is used in this case to approximate the loss in each layer, associating a different weight and  $\chi^2$  penalty to each distribution, depending for example on the number of missing hits. This fit is also able to take into account sudden changes in the curvature radius caused by an eventual bremsstrahlung photon emission.

3. The final step consists in identifying the clusters left in the ECAL by the photons emitted by extrapolation of the GSF track and in **merging this GSF track and the ECAL SCs** previously built. This step is also designed to preserve the highest efficiency possible while keeping the misidentification probability low and ambiguities related to single electron seeds which can often lead to several reconstructed tracks are also resolved at this stage.

Finally, a loose preselection is applied to the electron candidates in order to reject fake electrons and the variables related to the energy and geometrical matching between the GSF track and the ECAL cluster(s) are combined into a Multi-Variate Analysis (MVA) estimator allowing to define several electron Working Point (WP) as well.

This complete electron workflow explained here in a simplified way can be summarized in Figure 4.5.

## 4.4 Jets reconstruction

Eventual jets and gluons produced by a  $pp$  collision of the LHC usually manifest themselves as hadronic jets in the detector because of the colour confinement principle stating that coloured particles, such as the quarks, cannot be isolated and therefore be observed on their own.

This practically means that once a single quark is produced, it will start losing energy by forming new  $q\bar{q}$  pairs, themselves forming additional  $q\bar{q}$  pairs. This chain continues until the resulting pairs of quarks have such a low energy that they can start combining into colourless hadrons. This is called the *hadronization* process and the actual result of the apparition of a quark is a shower

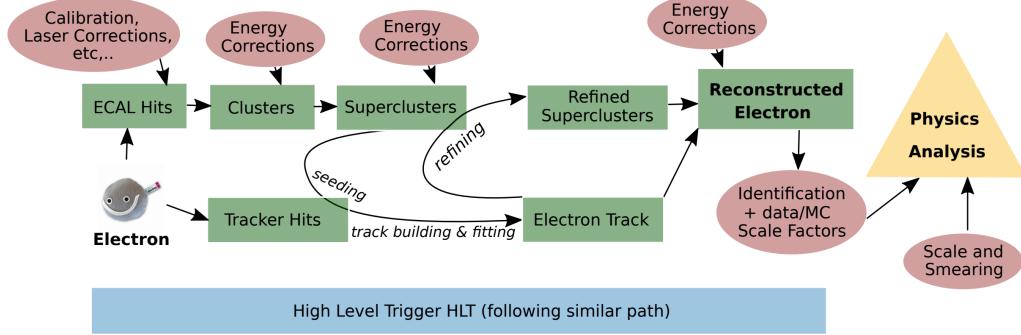


Figure 4.5: Schematic representation of the full electron reconstruction workflow in CMS [94].

of collimated particles, usually called jet, and seen by the detector as a set of tracks and energy deposits in the calorimeters, as shown in Figure 4.6.

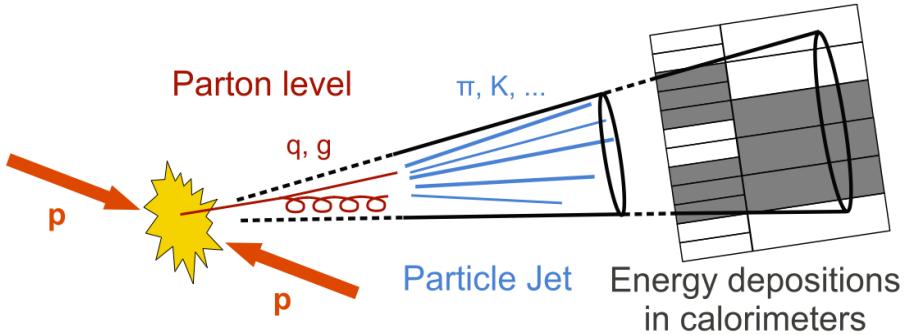


Figure 4.6: Schematic representation of the typical development of a jet within the CMS detector.

Several algorithms can be used to reconstruct the jets by linking the information coming from the tracker and the calorimeters, but the most used tool in CMS is the so-called anti- $k_T$  algorithm, able to cluster all the charged and neutral hadrons along with the eventual non-isolated photons or lepton produced and merge them into a single jet [95]. Its main objective is to compute the energy and direction of the original quark as precisely as possible. This is actually the best algorithm developed so far to resolve jets, but the worst for studying jet substructure due to its clustering preference.

To perform such a job, sequential clustering algorithm such as this one rely on the value of two distances:  $d_{ij}$ , the distance between two particles  $i$  and  $j$  that need to be clustered and  $d_{iB}$ , the distance between the particle  $i$  and the beam axis  $B$ . As seen in Equation 4.1, these distances can be computed using different variables such as  $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i + \phi_j)^2$ , the distance between  $i$  and  $j$  in the  $(\eta - \phi)$  space, the  $p_T^2$  of each particle and  $R^2$ , the algorithm radius parameters determining the final jet size and usually set to 0.4 by the CMS collaboration.

$$\begin{cases} d_{ij} = \min \left( \frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \Delta R_{ij}^2 \\ d_{iB} = \frac{1}{p_{T,i}^2} \end{cases} \quad (4.1)$$

The algorithm works by looking at all the  $i, j$  combinations, comparing the distances  $d_{ij}$  and  $d_{iB}$

until only jets are present in the event:

- If  $d_{ij}$  is smaller, then  $i$  and  $j$  are combined into a single particle ( $ij$ ) by summing their 4-vectors and both are removed from the list of particles to be clustered.
- If  $d_{iB}$  is smaller, then  $i$  is considered to be the final jet and is therefore removed from the list of jet candidates as well.

Several corrections are then usually applied to the jets constructed using this algorithm, in order to take into account several parameters, such as the non-linearity of the response of the calorimeter, the electronic noise, the Pile Up (PU) effects and the dependance of the reconstruction on the jet flavor and distance parameter  $R$ . This typically introduces a source of uncertainty that will be taken into account and discussed in Section 7.1.

The efficiency of the PF algorithm for jet identification and reconstruction has been checked using simulation, as shown in Figure 4.7. This study clearly shows that between 95 and 97% of the energy of the PF jet candidates can be reconstructed, compared to a 40-60% reconstruction efficiency using only the calorimeters data, and that this algorithm also leads to a gain in resolution up to a factor 3, depending on the jet  $p_T$ .

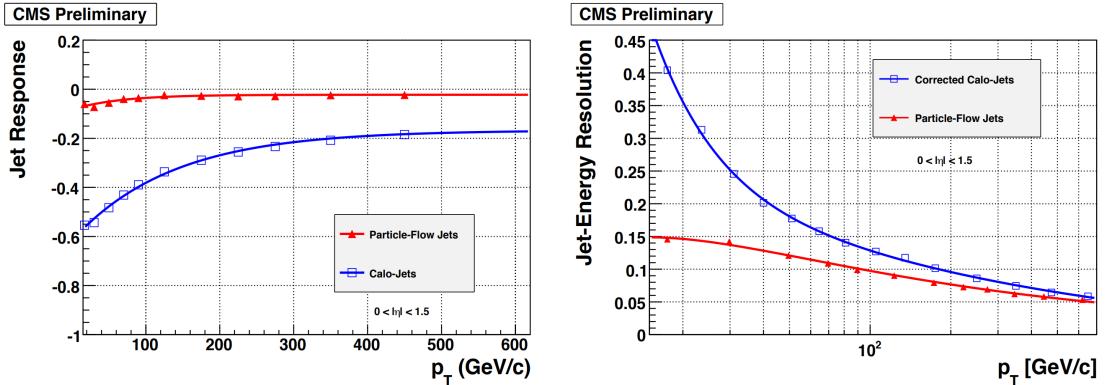


Figure 4.7: Comparison of the jet response (on the left) and jet energy resolution (on the right) for dijets simulated events in the barrel for jets reconstructed using only the calorimeters (in blue) and jet candidates from the PF algorithm (in red) [97].

#### 4.4.1 B-tagging

Jets coming from bottom quarks are usually quite interesting to study, especially in this analysis which relies heavily on the number of b-jets produced to define the control and signal regions, as will be discussed in Chapter 6.

This specific kind of jets can be distinguished from other jets because of the relatively long lifetime of the bottom quark ( $\tau \sim 1.5$  ps) that produces in the detector a secondary vertex displaced by a few millimeters with respect to the PV, as shown in Figure 4.8; and this gives a perfect way to discriminate jets coming from light quarks from b-jets. Another consequence of the large mass of the bottom quark is that a large number of particles is typically present inside this particular kind of jets and that the decay of the bottom quark even leads to the apparition of soft leptons in the decay chain in around 20% of the cases.

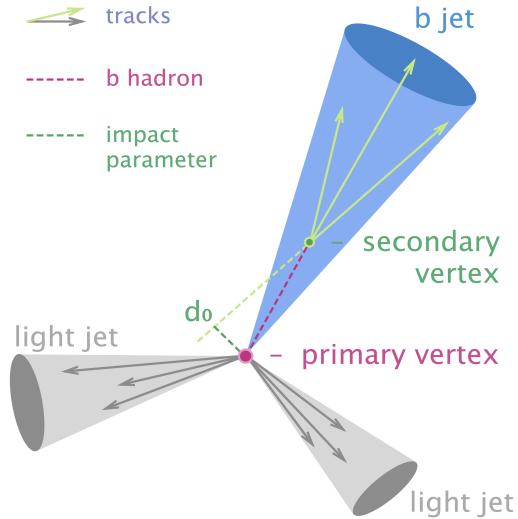


Figure 4.8: Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex.

Because of these specific properties, an algorithm can quite easily distinguish between jets coming from a bottom quark or from a lighter quark, and this will be a key point in this analysis. In our case, this discrimination is additionally optimized by using a multivariate technique able to combine all the discriminating power of the previous typical characteristics of any heavy flavour jet in the best way possible after reconstruction of all the vertices of the event. The main objective of the algorithm is to be able to identify b-jets as efficiently as possible while reducing the risk of possible misidentification of a jet.

In this analysis, the typical deep Combined Secondary Vertex (CSV) algorithm able to combine the information on the secondary vertex with the one on the track impact parameters and based on a Deep Neural Network (DNN) previously trained, has been used to identify such b-jets. The performance of this method can be observed in Figure 4.9, where we can see that this deep CSV algorithm is one of the best algorithms able to identify b-jets, depending on the phase space, while keeping a relatively low misidentification rate for light-flavor jets (u, d, s and gluons).

Different Working Points (WPs) are then made available for all the offline analyses in order to tweak the combo b-jet identification efficiency/misidentification rate. The loose, medium and tight WPs have been developed in such a way to limit this misidentification rate of a light jet as a b-jet to 10%, 1% and 0.1% respectively.

## 4.5 Missing Transverse Energy (MET)

Since the  $pp$  collisions happen mostly head-on, we know that the initial total transverse momentum of the event is exactly equal to 0 before the collision and we expect that it stays 0 afterwards because of the momentum conservation.

However, this statement is not totally true since we are aware of several effects that could induce an imbalance in this transverse momentum, as shown in Figure 4.10:

- Even though the detector has been carefully designed, some particles could be created outside

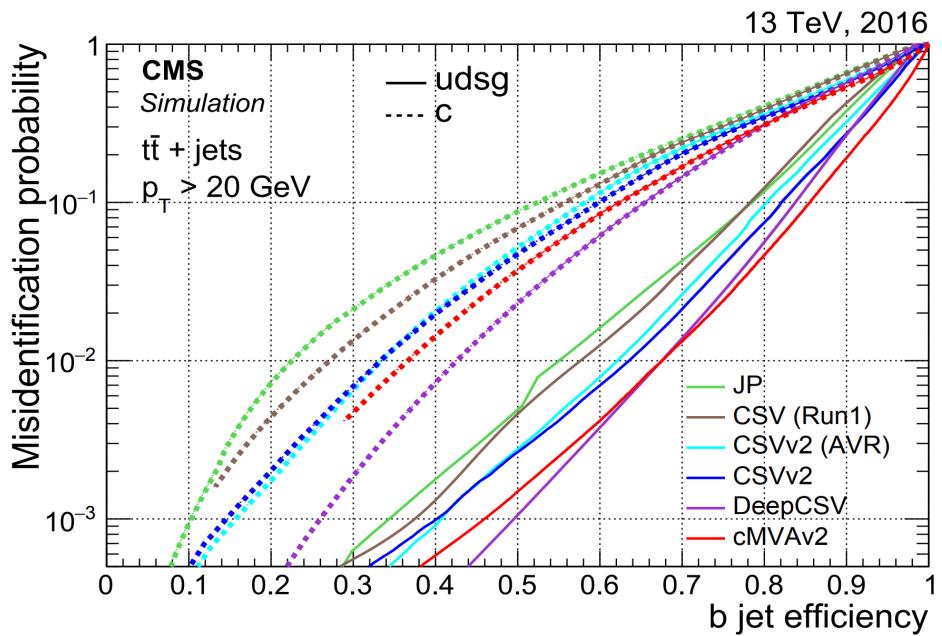


Figure 4.9: B-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [98].

of the acceptance of the detector and therefore escape the detection (a particle can for example be created with such a boost that it could be emitted back to the beam pipe itself, making it impossible to detect it).

- Because of their extremely low interaction cross-section, SM neutrinos are expected to escape the detector with some energy staying completely undetected.
- The finite momentum resolution of the detector can also lead to some inaccuracies in the measurement of the transverse momentum of all the particles created, leading to an instrumental MET in some cases.
- Finally, the eventual exotic weakly interacting particles produced, such as Dark Matter (DM), is typically expected to leave some MET in the detector as well.

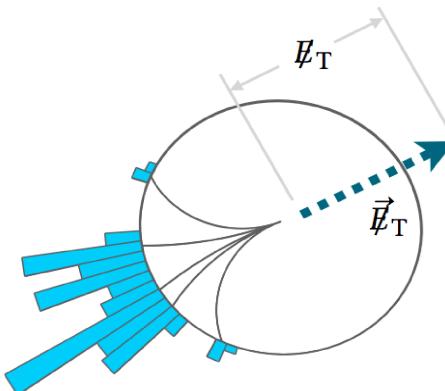


Figure 4.10: Schematic representation of the Missing Transverse Energy (MET).

The Missing Transverse Energy (MET) variable, defined in Equation 4.2 as the negative sum of the transverse momentum of all the particles  $j$  of the event, is able to take into account this eventual

imbalance in the transverse momentum and is therefore a key variable in most of the analyses searching for new BSM physics, which is not expected to interact with the detector.

$$\vec{p}_T^{\text{miss}} = \cancel{E}_T = - \sum_j \vec{p}_{T,j} \quad (4.2)$$

Different algorithms can be used in order to reconstruct this variable, the most famous being [99]:

- The **particle flow MET** (PFMET), including all the information of the detector (as opposed to the **calorimeter or tracker MET**, for example) and only the PF reconstructed objects to estimate the MET value. This is the typical variable used in most of the analyses today, because of its simple, robust, yet very performant estimate of the MET spectrum.
- The **Pileup Per Particle Identification (PUPPI) MET**, actually used in this analysis, has been developed on top of the PFMET in order to further reduce the dependence on the pileup of this variable by using local shape information around each PF candidate in the event along with event PU properties and tracking information.

Several corrections also need to be applied to this spectrum to filter anomalous high MET events arising because of a variety of reconstruction failures induced by the detector, because of several effects, such as the the electronic noise and eventual dead cells in the calorimeters or the presence of eventual beam halo particles from the LHC itself, leading to a global miscalculation of the final energy of the event. These filters are extremely important, especially in the end of the MET spectrum, as observed in Figure 4.11.

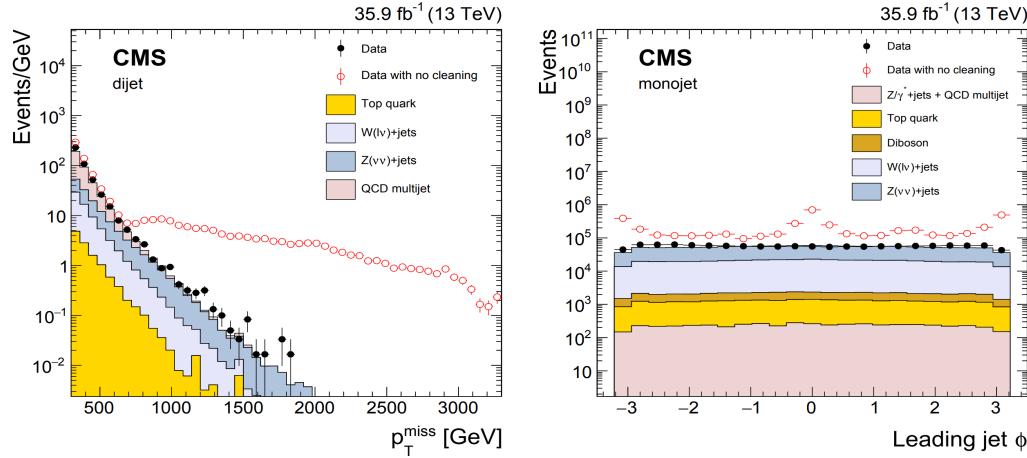


Figure 4.11: MET and jet  $\phi$  distributions with and without MET filters applied [99].

## 4.6 Top reconstruction

Although not formerly a part of the PF algorithm and done off-line, the kinematic reconstruction of the SM  $t\bar{t}$  process detailed in Section 5.6.1 is still an extremely important part of this analysis. As previously explained, because of the high mass of the top quark, it is expected to decay before reaching the detector, almost 100% of the time into a  $b\bar{b}$  pair of quarks, which can be identified using the b-tagging method explained in Section 4.4.1, and a W boson, unstable as well (in our

particular case, only the W decay to a lepton and a neutrino is considered for reasons explained in Section ??, even though this decay has a low Branching Ratio (BR)).

The ideal kinematics of such an event can be expressed using the four-momenta of the different particles involved, as expressed in the Equations 4.3a to 4.3c.

$$\begin{cases} p_x^{\text{miss}} = p_{\nu_x} + p_{\bar{\nu}_x} \\ p_y^{\text{miss}} = p_{\nu_y} + p_{\bar{\nu}_y} \end{cases} \quad (4.3a)$$

$$\begin{cases} m_{W^+}^2 = (E_{l^+} + E_\nu)^2 - (\vec{p}_{l^+} + \vec{p}_\nu)^2 \\ m_{W^-}^2 = (E_{l^-} + E_{\bar{\nu}})^2 - (\vec{p}_{l^-} + \vec{p}_{\bar{\nu}})^2 \end{cases} \quad (4.3b)$$

$$\begin{cases} m_t^2 = (E_b + E_{l^+} + E_\nu)^2 - (\vec{p}_b + \vec{p}_{l^+} + \vec{p}_\nu)^2 \\ m_{\bar{t}}^2 = (E_{\bar{b}} + E_{l^-} + E_{\bar{\nu}})^2 - (\vec{p}_{\bar{b}} + \vec{p}_{l^-} + \vec{p}_{\bar{\nu}})^2 \end{cases} \quad (4.3c)$$

In all these previous equations, the quantities related to the leptons and the b-jets can be measured using the PF reconstruction, while the energy of the neutrinos can be considered equal to their momentum because of their extremely low mass. It is also important to note that the different masses appearing in these equations will be treated as exactly known even though this is only true at first approximation since they are actually Breit-Wigner (BW) distributions.

We therefore have 6 equations to solve and exactly 6 unknowns corresponding to the three momentum components of each neutrino produced, a problem that can in principle be solved, leading to a quartic equation in one of the unknowns analytically solvable but quite ambiguous given the variable number of solutions of such equation [100].

#### 4.6.1 Top reconstruction in practice

Even though the equations and the problem seems quite well defined and solvable, in practice, several complications quickly appears. First of all, when considering real data, the pairing between the leptons and the b-jets observed is not obvious to perform. Typically, three cases will then be defined in order to take this effect into account:

- If exactly 0 b-jets are observed, the event will not be considered in this analysis, according to the selection of the signal regions performed and explained later in Chapter 6.
- If more than one b-jet is observed, then only two different permutations considering the two b-jets and the two leptons having the highest momentum  $\{l^+ \leftrightarrow b_1, l^- \leftrightarrow b_2\}$  and  $\{l^- \leftrightarrow b_1, l^+ \leftrightarrow b_2\}$  are considered.
- Finally, if exactly one b-jet is observed, then it will be kept and all the non b-tagged jets will be considered as the second b-jet candidate. In this case, several combinations between the two leptons and all these jets are then considered, resulting in  $2n$  different combinations, where  $n$  is the number of jets in the event.

The combination leading to the lowest invariant mass for the  $t\bar{t}$  system is then chosen between all

the different possibilities for the reconstruction of the system.

Section to be updated once the analysis performed and the details about the top reco performed known, and add the results obtained.

# Chapter 5

## Data, signals and backgrounds

In order to find a possible hint of the production of DM in the LHC collisions considering our signal models of interest, briefly described in Section 2.6, the data collected needs to be compared with Monte Carlo (MC) simulations produced in a central way for each SM process. Indeed, any deviation of the data observed with respect to what we expect to see, obtained from these MC simulations, might be the sign of some BSM physics. All of the steps needed to mathematically simulate the  $pp$  collisions of the LHC and to take into account the effect of the detector on the particles produced will first of all be introduced in Section 5.1.

Then, the different formats of files available to perform the analysis and the code used will be briefly introduced in Sections 5.2 and 5.3 and the different data samples collected during the Run II of operation of the LHC will be then detailed in Section 5.4, while the signal models and samples considered in this particular analysis along with the MC samples used for the simulation of the different backgrounds will be introduced in Sections 5.5 and 5.6 respectively.

### 5.1 The Monte Carlo (MC) simulation method

As previously explained, the generation of MC simulations for the most common SM processes is a crucial step of any analysis because they are considered to be the reference to which the data collected is compared in order to try and find some discrepancies, which could be the sign of the existence of BSM physics. Searches for exotic physics therefore heavily depend on these simulations, which need to be generated with great care and to which a large uncertainty is typically associated since the collision between the partons of two protons and the interaction between the particles produced and the detector itself are extremely complex by nature.

The basic idea of the MC simulation consists in using a random number generator to simulate the randomness of nature and produce as many events as computationally possible for all the SM processes, taking into account the probability density functions of these processes. This is performed by specific softwares called **event generators** and it is important to note that since we usually don't know everything about the SM or BSM process being generated, the perfect event generator does not exist.

To make the generation of such simulations a bit easier, the description of a typical  $pp$  collision can usually be divided into several steps that will now be described, as shown with the color code used in Figure 5.1. The typical approximations used to make this kind of simulation possible from the computational point of view will also be briefly introduced at this point.

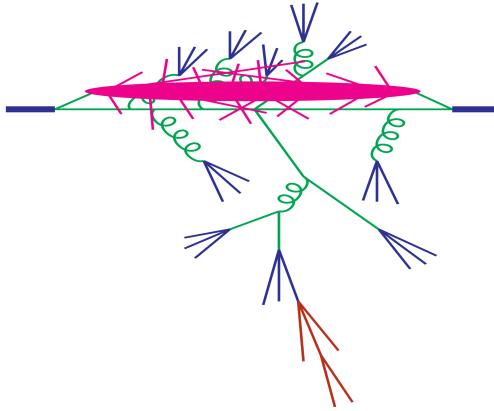


Figure 5.1: Structure of a  $pp$  collision and different steps of the MC simulation used by the event generators, such as the parton shower (in green), the Underlying Event (UE) (in pink), the hadronization (in blue) and the decay of unstable particles (in red) [101].

### Hard scattering

A typical  $pp$  collision at a center of mass energy  $\sqrt{s}$  is usually described by an event generator as the interaction between a parton  $i$  coming from one proton with a parton  $j$  coming from the other, leading to the production of a final state  $A$ , made out of  $n$  different particles. The total cross section of such process can be expressed with Equation 5.1 [102].

$$\sigma_A(s) = \sum_{i,j} \int \int dx_1 dx_2 f_i(x_1, \mu^2) f_j(x_2, \mu^2) \hat{\sigma}_{ij \rightarrow A}(\hat{s}, \mu^2) \quad (5.1)$$

In this equation, several variables have been introduced, such as:

- The artificial parameter  $\mu^2$  used as the delimitation between short and long range physics.
- The Parton Density Functions (PDFs)  $f_i(x, \mu^2)$  of both partons involved in the collision, giving the probability of finding in the proton a parton of flavor  $i$  (quark or gluon) carrying a fraction  $x$  of the proton momentum.
- The integrated parton-level cross section  $\hat{\sigma}_{ij \rightarrow A}$  describing the short range physics between the partons, taking into account the phase space and the matrix element obtained considering all the Feynman diagrams of a given process.
- The square invariant mass of the two partons  $\hat{s} = (p_i + p_j)^2$ .

Many algorithms have been developed in order to select a hard process  $ij \rightarrow A$  and determine its kinematics by solving this equation using different methods. The samples used in this work have actually been produced at different orders and by different hard scattering generators, such as MADGRAPH [104] (at LO) and POWHEG [105] and MC@NLO [106] (at NLO).

### Parton showers

The parton shower phase is then used to describe what happens to the incoming and outgoing partons after the initial collision that has just been described. The hard process induce by definition

a large acceleration to the partons involved, which then tend to emit QCD radiation under the forms of gluons, just like accelerated electric charges do by emitting photons. However, the gluons emitted do have a color charge and can therefore emit further radiation until reaching such a low energy that they are able to form colourless hadrons, as discussed in Section 4.4. This process typically leads to the creation of the so-called **parton showers**, approximate higher-order real-emission corrections to the hard scattering, that need to be simulated by the event generators as well since they are an important part of the kinematics of the collision.

The parton showering then consists in simulating these showers for not only the final state particles produced by the hard scattering, but also for the particles in the initial state and for the remnants of the colliding protons, since gluons can actually be emitted by Initial State Radiation (ISR) and by these remnants themselves.

### Underlying Event (UE)

Once the hard scattering and all the possible gluon emissions simulated, the next step consists in considering the so-called **Underlying Event (UE)** arising from the parton showers just described and from the secondary collisions between partons not involved in the primary hard process, the so-called Multiple Parton Interactions (MPIs). The UE is usually responsible for the production of particles at low transverse momenta  $p_T$  that cannot be experimentally distinguished from particles produced from initial or final state radiation but still need to be simulated.

These secondary collisions typically lead to the production of extra hadrons and therefore need to be simulated as well by events generators, usually by distributing the partons of the incoming protons in an area of  $1\text{fm}^2$ : an increased UE will be obtained when the so-called impact parameter, the distance between the parton and the centre of this area, is decreased, making the collision mostly central and almost head-on [107]. The UE is typically well simulated using softwares such as Herwig [108] and PYTHIA [109]. The spectrum for the generation of some variables in a top enriched sample can be found in Figure 5.2.

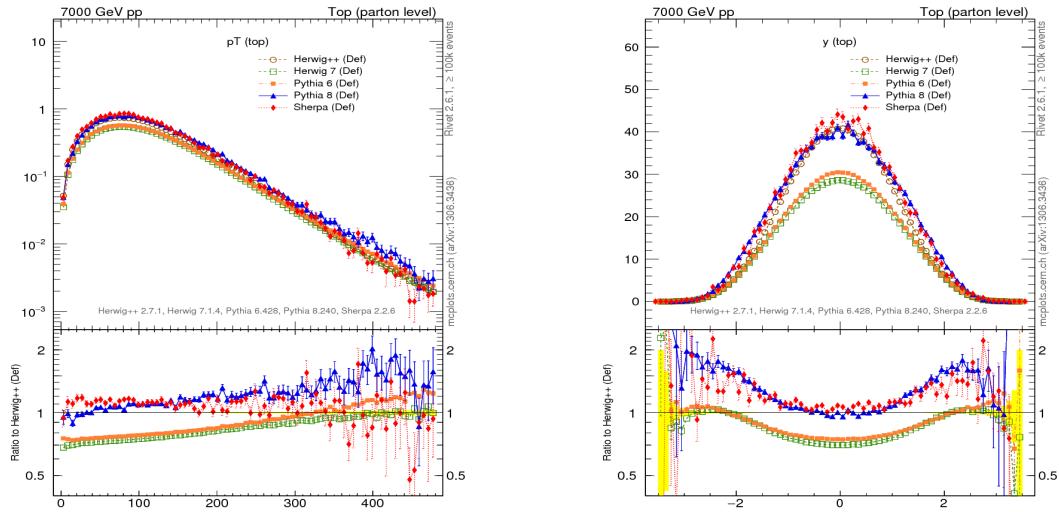


Figure 5.2: Top  $p_T$  (on the left) and rapidity (on the right) distributions obtained using different MC generators [110].

## Hadronization

Once all the primary and secondary collisions simulated, it is time for the event generators to simulate the **hadronization** and binding processes of the different coloured partons emitted into colourless hadrons, as explained in Section 4.4. This hadronization process happen at low energies, when the perturbation theory becomes invalid and the dynamics enter a non-perturbative phase, which leads to the formation of the observed final-state hadrons. Non-perturbative calculations then have to be used by the event generators in order to simulate this effect.

## Unstable particle decays

The last step of the MC generation consists in finding a model allowing the unstable hadrons created in the hadronization process to decay, and to study these decays. This is extremely important because experimental data clearly shows that a large fraction of the observed final state particles come from the decays of such excited hadronic states.

## Detector simulation

Once the event completely simulated using the event generators and the PU taken into account by reproducing the hard scattering process several times, another step is required: simulating the interaction between the "perfect" particles previously created and the "imperfect" CMS detector.

This is typically done by the GEANT4 software [111], able to model different effects, such as:

- Modeling of the interaction region
- Modeling of the particle passage through the volumes that compose CMS detector and of the accompanying physics processes
- Modeling of the effect of multiple interactions per beam crossing and/or the effect of events overlay (PU simulation)
- Modeling of the detector's electronics response

This modeling accounts for all the cracks and for the disposition of the subsystems inside of the CMS detector. This software is for example able to model the interaction of the electrons with the tracker, responsible for the emission of bremsstrahlung photons, as explained in Section 4.3.2.

The results of the comparison between the output of two different versions of the GEANT4 software and prototypes of the CMS calorimeter in the test beam facility at CERN lead to comparable results, as shown in Figure 5.3.

However, the modeling of the detector is not perfect and not all the inefficiencies can be accounted for. In some cases, Scale Factors (SF) are then used to correct the MC simulations and correct some expected discrepancies between data and MC. This will be detailed later on.

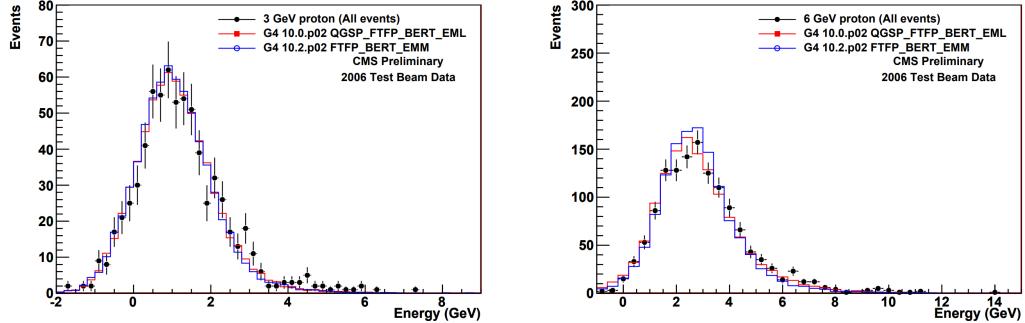


Figure 5.3: Proton energy distribution at 3 (on the left) and 6 (on the right) GeV compared for the test beam data (in black) and two different GEANT4 versions [112].

## 5.2 Files format

Once recorded (or simulated), the data (or MC) still needs to go under a complete post-processing in order to change its format and reduce the total size of the samples to be considered in the analyses. Different types of analysis are expected to need different levels of data reduction, so the data is usually accessible at different levels [113]:

- Virgin-RAW: used only in low rate runs with heavy ions collisions (10-15Mb/event)
- RAW : standard raw data event content (1Mb/event)
- RECO: detailed information on reconstructed physics objects (3Mb/event)
- Analysis Object Data (AOD): physics objects used in analysis (400-500kB/event)

Two additional formats were introduced since the end of the Run I. First of all the MiniAOD was introduced to reduce the size of the AOD by a factor 10 while retaining most of the information about all the particles that were created, without applying any further selection.

Because of the increased integrated luminosity collected by CMS over the last few years, a brand new file format featuring another reduction of the file size of a factor  $\sim 50$  was recently introduced: the nanoAOD, able to retain most of the information of each collision in around 1kB of data per event only. This reduction in size was achieved by optimizing the floating point of the variables, by not storing quantities that can be recomputed from the available information and by limiting the number of physics objects available, for example. This means that some low-level analyses cannot use this format to work, but it has been estimated that around 50-70% of the analyses performed at CMS can rely on such files in order to work.

In this particular case, the 6th version of the nanoAOD, introducing a series of bug fixes and the latest jet energy corrections, was used for both the data and the MC samples (signal and backgrounds) that will now be listed in the next sections.

### 5.3 Analysis code

The code used for the event generation, simulation and reconstruction is the version 10.2.X of the official software of the CMS collaboration, called CMSSW [115]. This software contains the CMS Event Data Model (EDM) which is able to describe every event as a C++ object containing all the RAW and reconstructed information related to the collision. These objects are stored using the ROOT file format [116], an analysis package written in C++.

Once all the different samples produced centrally up to the nanoAOD stage, another framework was put in place in order to do a post-processing of such samples, by selecting objects interesting for different dileptonic analyses, reducing therefore even more the size of the samples to be considered by selecting only events having 2 tight leptons. This selection will be detailed in Chapter 6. This *Latino* framework, written in python, is common to several different analyses and has been developed by tens of different people over the past few years, providing several tools to produce samples, read the files, apply different corrections to the MC samples and produce the histograms needed to perform a search such as this one.

### 5.4 Data samples

As already explained in Section 3.1.2, the data analyzed in this work has been taken at a center of mass energy of 13 TeV during the second part of the Run II of operation of the LHC.

During this period, an integrated luminosity of  $35.9 \pm 0.9 \text{ fb}^{-1}$  (2016) [117],  $41.5 \pm 1.0 \text{ fb}^{-1}$  (2017) [118] and  $59.7 \pm 1.5 \text{ fb}^{-1}$  (2018) [119] has been collected, resulting in a total dataset of  $137.1 \pm 2.0 \text{ fb}^{-1}$  recorded by the CMS detector and ready to be analyzed. This data has been obtained by combining a set of single and double lepton triggers that will be described in Section 6.1.1 by taking care of avoiding any eventual double counting due to events present in different triggers. All the data samples considered for this analysis are listed in Section A.1.

### 5.5 Signal samples

To be completed once the files are actually available Listed in Section A.2.

### 5.6 Backgrounds prediction

Several different SM background processes have been considered for this analysis, all listed in Section A.3 and mostly estimated directly from MC. In this section, the main backgrounds to consider for this particular analysis will be reviewed, such as:

- The major background for the  $t\bar{t}$ +DM analysis, the SM  $t\bar{t}$ , kinematically really close to the signal searched for (Section 5.6.1).
- On the other hand, the major background for the  $t/\bar{t}$ +DM analysis is the single top production, which has an even higher cross section than the  $t\bar{t}$  (Section 5.6.1).

- Then, mainly because of its huge cross section at 13 TeV, as shown in Figure 5.4, the DY process is usually important to consider. Even though quite reduced in the Signal Regions (SRs) because of the cuts applied, a specific Control Region (CR) will be dedicated to check this background, taken directly from MC (Section 5.6.2).
- The  $t\bar{t} + V$  ( $t\bar{t} + Z$  and  $t\bar{t} + W$ ) may have a kinematics even closer to our  $t\bar{t}$ +DM signal than the  $t\bar{t}$  process and is therefore extremely important in our signal regions, even though its low cross section does limit its impact (Section 5.6.3).
- Finally, the non-prompt background is another important piece of this analysis mainly because of the particular data-driven method used to compute them (Section 5.6.4).

Add percentage of each background once known

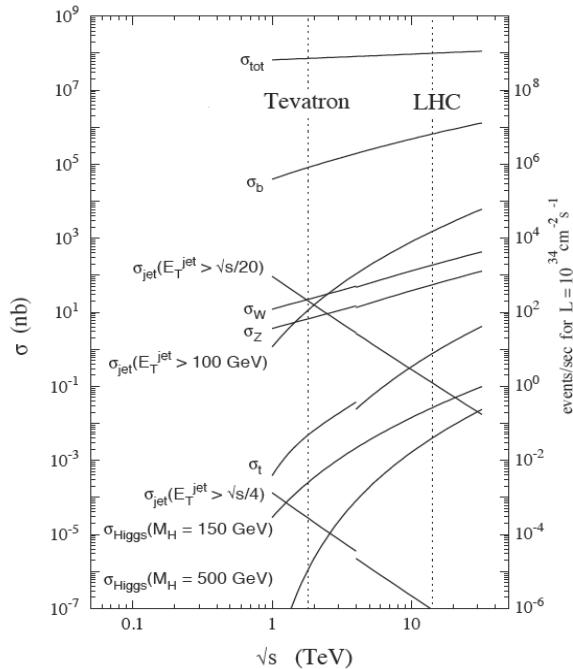


Figure 5.4: Production cross section of the most common SM processes considering different center of mass energies, such as the 13 TeV of the LHC.

Finally, some smaller backgrounds will be introduced in Section 5.6.5, such as the diboson and triboson production, and the weights and corrections applied to all these MC samples will be detailed in Section 5.6.6.

### 5.6.1 Top production

Because of the relatively high production cross section of top quarks at 13 TeV, the production of a single or a pair of top quarks, but without the production of associated DM, is obviously the dominant background in both searches.

They also have a kinematics quite close to the one expected for our signal: some additional MET is expected because of the production of a pair of DM particles, which might lead to a distribution  $\phi$  distribution as well, but other than that achieving some discrimination between these processes

is quite complex. This process mostly relies on the use of Machine Learning (ML) techniques and will be fully detailed in Section 6.4.

Typically, given the importance of such processes, a specific CR to check for the validity of the CR of such processes is defined, as explained in Section 6.3.

### The main background: $t\bar{t}$

Different Feynman diagrams contribute to this process at Leading Order (LO) in a hadron collider, as shown in Figure 5.5. This background is estimated directly from MC and, due to its importance, dedicated CRs have been designed in which to check for the data/MC agreement.

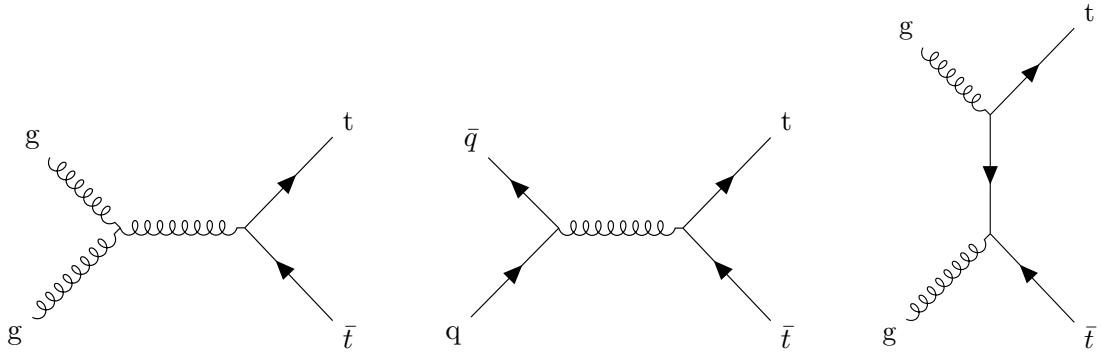


Figure 5.5: Main feynman diagrams for the production of the SM  $t\bar{t}$  process.

### Single top

Different Feynman diagrams also account for this process in the s-channel (Figure 5.6), t-channel and tW production mode (Figure 5.7).

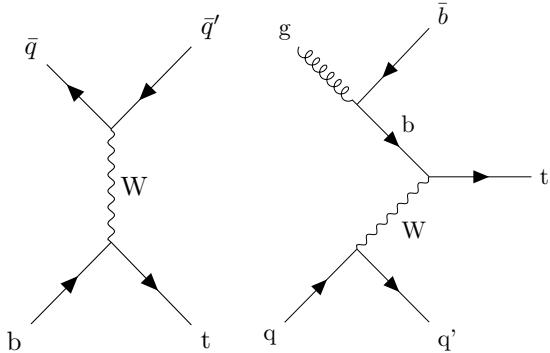


Figure 5.6: Feynman diagrams for s-channel production mode of a single top quarks.

### Top decay

As previously mentioned, the top is the heaviest particle of the SM and is expected to decay inside of the beam pipe itself, usually into a bottom quark, giving us a b-jet, and a W boson; this boson can decay itself into different channels even though only its leptonic decay is consider in this particular case. The decay considered of the top/antitop produced is represented in Figure 5.8.

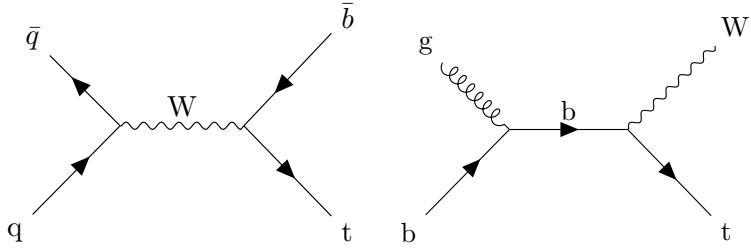


Figure 5.7: Feynman diagrams for t-channel (on the left) and tW (on the right) production modes of a single top quarks.

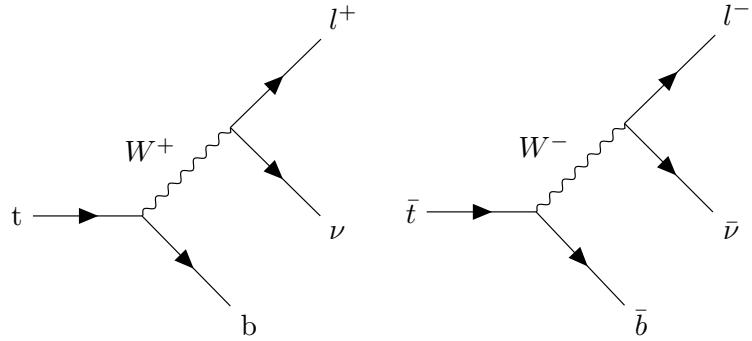


Figure 5.8: Feynman diagrams for the leptonic decay of the top (on the left) and antitop (on the right) quarks.

### 5.6.2 Drell-Yan estimation

As previously mentioned, most of the DY, produced through the Feynman diagram represented in Figure 5.9, is not expected to survive the selection applied to the analysis but because of the huge cross section of this process, two to three orders of magnitude larger than the production of top, the contribution of this process in the SRs is still quite important.

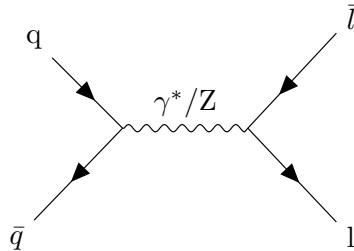


Figure 5.9: Feynman diagram for the DY process involving a virtual  $\gamma^*$  or  $Z$  boson.

This background is also estimated from MC as well and is being checked in a specific CR as well, as described in Section 6.3.

### 5.6.3 $t\bar{t} + V$

This background is coming from a usual  $t\bar{t}$  production along with an ISR or Final State Radiation (FSR) production of a W or Z boson, as shown in Figure 5.10. The contribution of this background is also taken directly from MC.

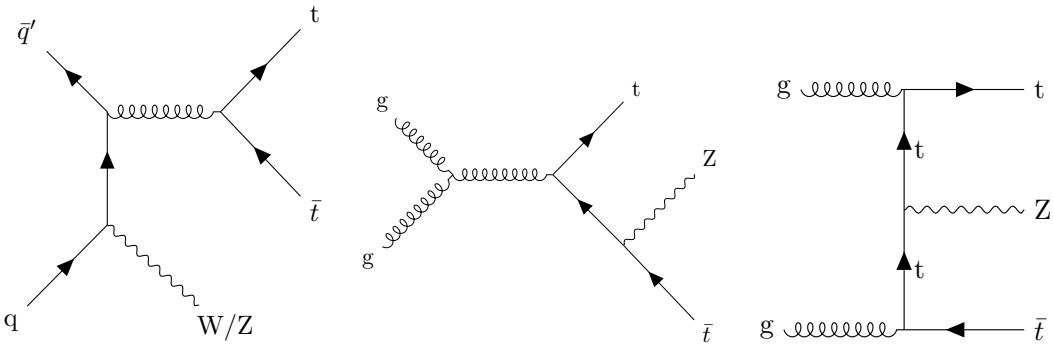


Figure 5.10: Possible Feynman diagrams for the Initial State Radiation (ISR)  $t\bar{t}$  with a W/Z boson (on the left) and for the production of an Final State Radiation (FSR)  $ttZ$  (on the center and right).

The resulting cross section of such process is a bit lower than the production of the SM  $t\bar{t}$  on its own but the kinematics of this background can be extremely close to our signal, since the W or Z boson produced can give a SM neutrino, leading to some actual MET.

#### 5.6.4 Non prompt contamination

Even though not extremely important in the sense that its kinematics allow us to remove most of its contributions in the SRs, this background is interesting in the sense that it can be estimated using a data-driven method that will now be described instead of being taken directly from MC.

A few definitions are first of all needed to explain the method used to compute the importance of this background in the different regions of the analysis:

- First of all, a **prompt lepton** is defined as a real lepton, in the sense that the lepton is originating from the PV of a  $pp$  collision.
- The **Prompt Rate (PR)** is defined as the number of prompt leptons passing the tight selection criteria of the analysis over the number of leptons passing the loose selection criteria.
- On the other hand, by **fake** or **non-prompt lepton**, we usually refer to truly **fake leptons**, such as jets misidentified by the detector as leptons, as shown in Figure 5.11, and real leptons coming from eventual heavy flavor decays.
- The **Fake Rate (FR)** is then defined similarly to the prompt rate but considering this time fake leptons only for the tight-to-loose ratio. This ratio therefore corresponds to the probability for a fake lepton to be considered as a real lepton in the analysis.

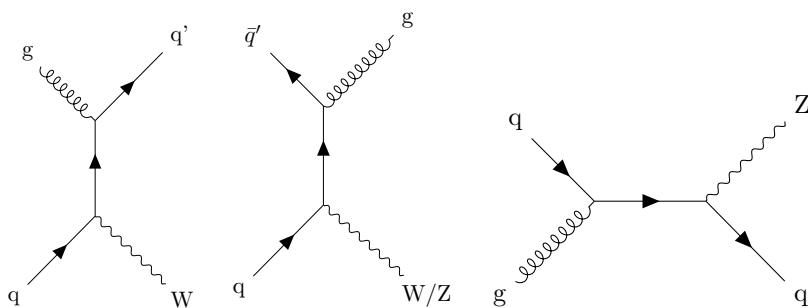


Figure 5.11: Possible Feynman diagrams for the production of a W/Z boson with a jet.

This background is particularly important at low  $p_T$ , where the misidentification rate is higher, and is not expected to be modeled correctly by MC because of its complexity: a general **tight-to-loose datadriven method** is then used to compute its kinematics and final contribution in the different regions of the analysis.

In general, this method contains three main steps: the computation of the FR and PR, the extension of these rates in a region kinematically close to the SR of the analysis and the definition of a same sign CR enriched in fakes in order to perform a closure test of the yields and kinematics of this background. All these steps will now be detailed.

### Fake Rate (FR) computation

Because of its definition, the FR is computed in a prompt lepton-free region, typically in a loose QCD enriched region, defined with the following cuts:

- Exactly 1 lepton
- $mtw1 < 20$  GeV
- $p_T > 13$  (10) GeV for  $e$  ( $\mu$ )
- PuppiMET  $< 20$  GeV
- $|\eta| < 2.5$  (2.4) for  $e$  ( $\mu$ )
- PassJets

All the previous cuts have been designed to define a loose QCD region as pure as possible by removing most of the W+jets and Z+jets contribution. The PassJets cut is a boolean obtained by looping over all the jets of the event trying to find a jet having an  $E_T$  higher than a given threshold in order to control the average  $p_T$  of the jet that fakes the lepton (actually, different FR have been computed for different  $E_T$  thresholds, from 10 to 50 GeV).

Using this method, the jet that fakes a lepton is actually the one recoiling against ( $\Delta R > 1.0$ ) the jet used for the systematics, as shown in Figure 5.12.

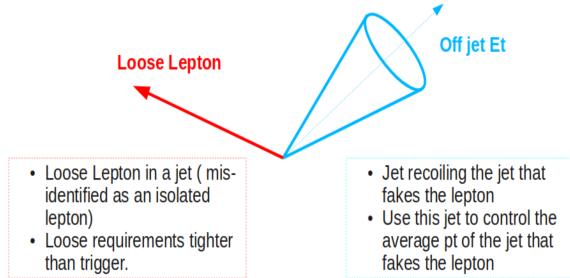


Figure 5.12: Schematic representation of the two jets used for the systematics and for the jet faking a lepton in the tight-to-loose datadriven method.

Events passing the following prescaled triggers are then selected in this region:

$$\text{Muon triggers} = \begin{cases} \text{HLT\_Mu8\_TrkIsoVVL (if } p_T < 20 \text{ GeV)} \\ \text{HLT\_Mu17\_TrkIsoVVL (if } p_T \geq 20 \text{ GeV)} \end{cases} \quad (5.2a)$$

$$\text{Electron triggers} = \begin{cases} \text{HLT\_Ele8\_CaloIdL\_TrackIdL\_IsoVL\_PFJet30 (if } p_T < 25 \text{ GeV)} \\ \text{HLT\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\_PFJet30 (if } p_T \geq 25 \text{ GeV)} \end{cases} \quad (5.2b)$$

The remaining of the Electroweak (EWK) processes ( $W+jets$ ,  $Z+jets$ ) able to pass the previous cuts of the QCD region, are then simply subtracted: this is the so-called **EWK subtraction**.

Since both the FR and PR heavily depend on the kinematics of the event and on the Working Point (WP) chosen for the leptons of the analysis, they are computed separately depending on the flavor of the lepton and 2D histograms (accounting for the  $p_T$  and  $\eta$  of the event) need to be created at this stage to calculate this factor, for a given input jet  $E_T$  threshold; 1D histograms corresponding to the projections of these 2D histograms along both their axes are also defined at this point, as shown in Figure 5.13.

### Prompt Rate (PR) computation

The PR, taking into account the real lepton contamination in the CR defined, is also important to calculate, even though the objects WP are usually chosen in such a way that this ratio is quite close to 1 and can therefore be ignored.

In our case, this rate has been calculated as well using a general tag and probe method in a  $Z+jets$  enriched sample. The main objective is to reconstruct  $Z \rightarrow ll$  events in this region and to select all the events for which the first lepton can be characterized as tight. Then, we search for the second lepton coming from the decay of  $Z$  within all the leptons detected by calculating the reconstructed mass of all the possible leptons combinations and selecting the one which is closer to the expected mass of the  $Z$  boson. We can then simply count how many times this second lepton, expected to be tight, has actually been measured as a tight lepton to estimate this PR.

The results obtained in this case have been represented in Figure 5.14

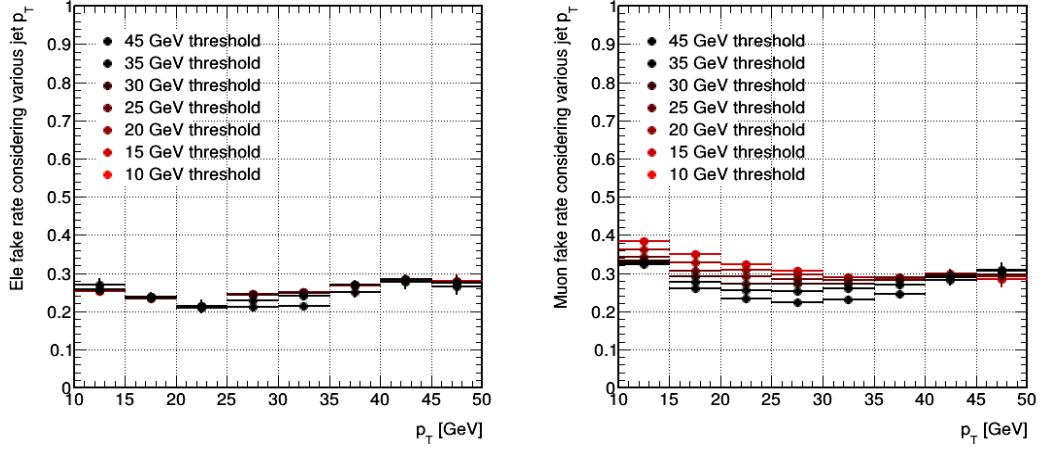
### Fake weight calculation

Once the fake and prompt rates computed in their specific region, it is still necessary to apply them to a fake-lepton region kinematically close to the SRs of the analysis (usually, a l2loose region). For this, a simple set of equations can be used. We start by defining the following quantities:

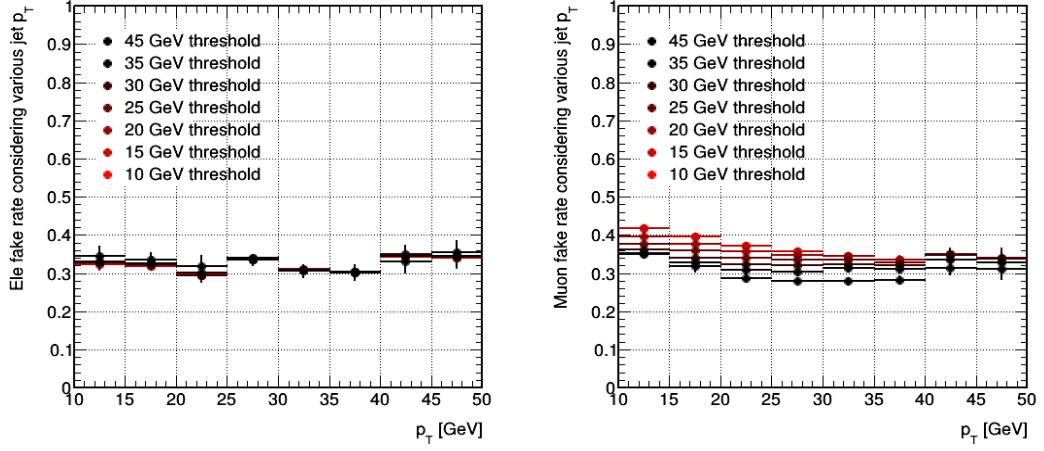
- $N_{pp}$  events where both leptons are prompt
- $N_{fp}$  events where one lepton is prompt and the other is fake
- $N_{ff}$  events where both leptons are fake
- $N_{tx}$  ( $x = 0, 1, 2$ ) events with 0, 1 or 2 leptons passing the right cuts, the **only quantity directly measurable** by the detector

It is then possible to see in Equation 5.3 that, if  $p$  is the PR and  $f$  the FR previously calculated.

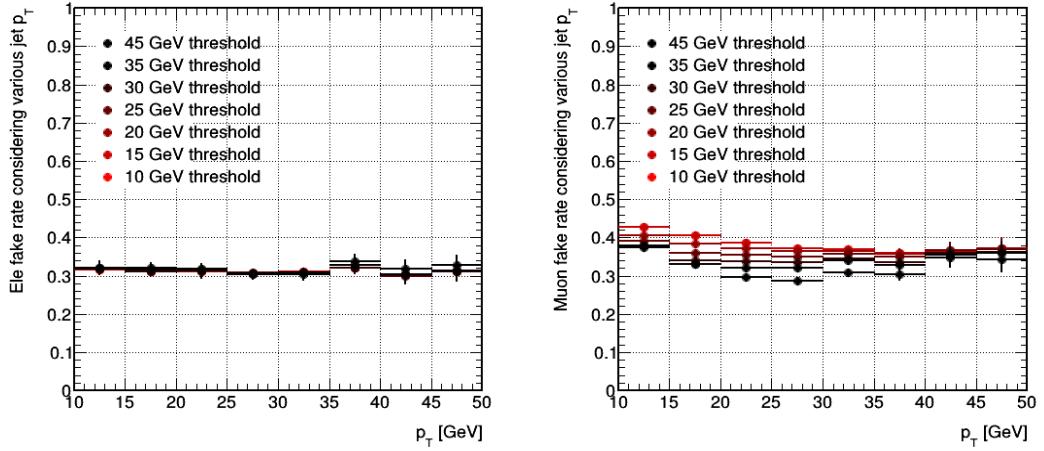
$$\left\{ \begin{array}{l} N_l = N_{pp} + N_{fp} + N_{ff} = N_{t2} + N_{t1} + N_{t0} \\ N_{t0} = (1-p)^2 N_{pp} + (1-p)(1-f)N_{fp} + (1-f)^2 N_{ff} \\ N_{t1} = 2p(1-p)N_{pp} + (f(1-p) + p(1-f))N_{fp} + 2f(1-f)^2 N_{ff} \\ N_{t2} = p^2 N_{pp} + pfN_{fp} + f^2 N_{ff} \end{array} \right. \quad (5.3)$$



(a) 2016 electron and muon fake rates

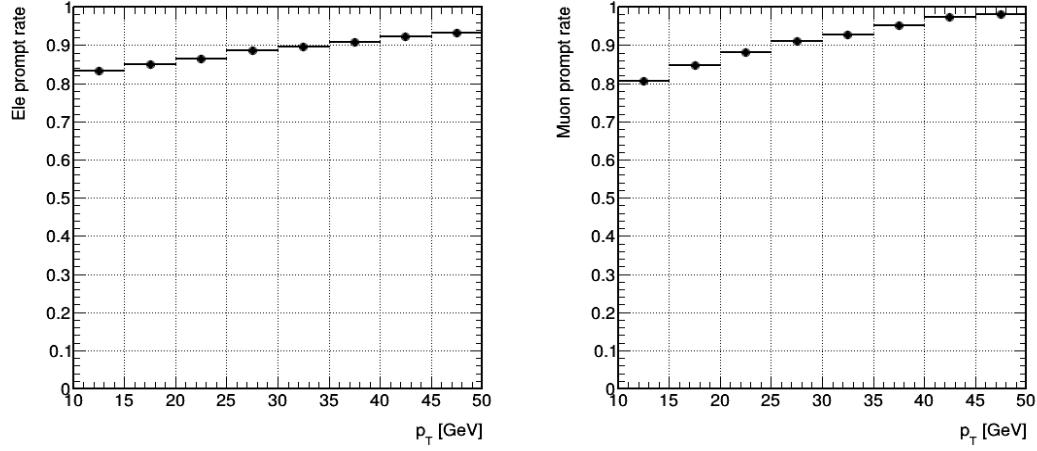


(b) 2017 electron and muon fake rates

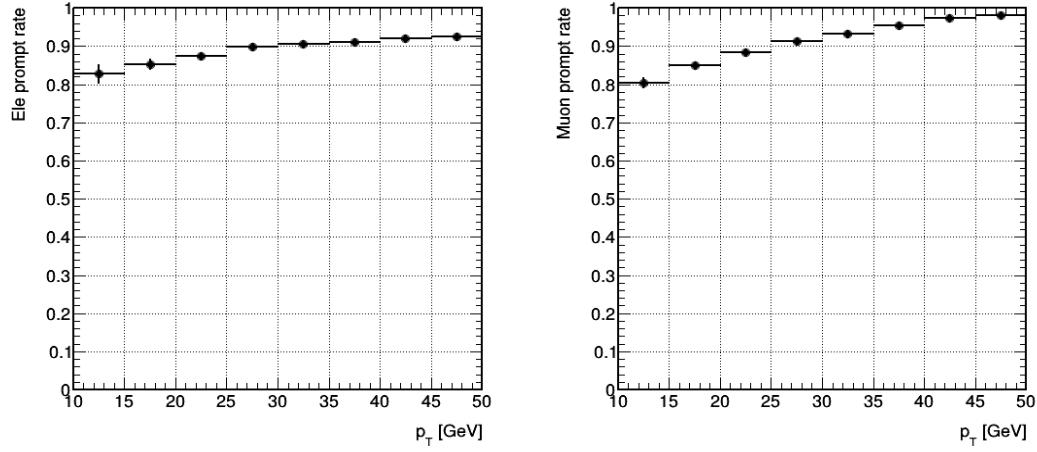


(c) 2018 electron and muon fake rates

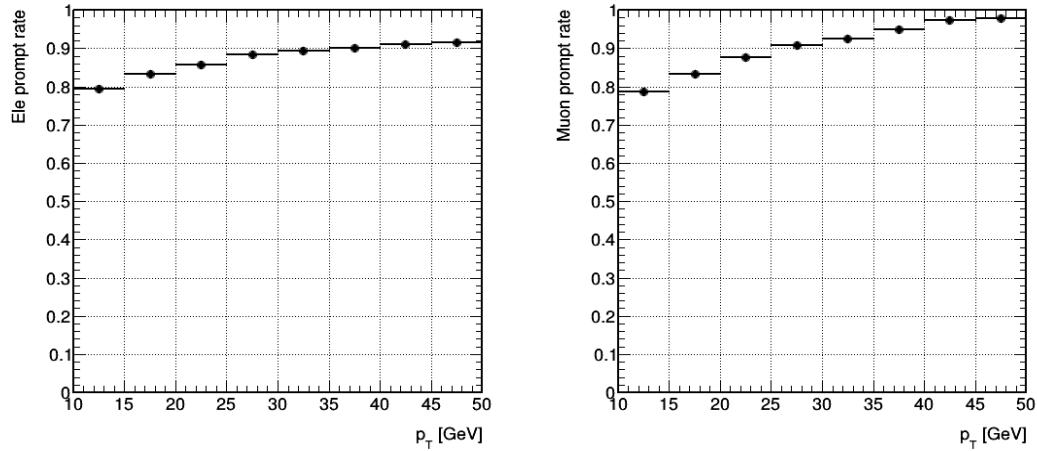
Figure 5.13: Electron (on the left) and muon (on the right) FR obtained in a QCD enriched region for different jet  $E_T$  thresholds for 2016, 2017 and 2018 with respect to the  $p_T$  of the lepton.



(a) 2016 electron and muon prompt rates



(b) 2017 electron and muon prompt rates



(c) 2018 electron and muon prompt rates

Figure 5.14: Electron (on the left) and muon (on the right) PR obtained in a Z+jets enriched region by a general tag and probe method for 2016, 2017 and 2018 with respect to the  $p_T$  of the lepton.

These equations can be inverted in order to represent the unknowns with respect to the known variables, as shown in Equation 5.4, giving us a way to apply the weights previously calculated to this particular l2loose region.

$$\begin{pmatrix} N_{pp} \\ N_{fp} \\ N_{ff} \end{pmatrix} = \frac{f-p}{-(p-f)^3} \cdot \begin{pmatrix} f^2 & -f(1-f) & (1-f)^2 \\ -2fp & p(1-f) + f(1-p) & -2(1-p)(1-f) \\ p^2 & -p(1-p) & (1-p)^2 \end{pmatrix} \cdot \begin{pmatrix} N_{t0} \\ N_{t1} \\ N_{t2} \end{pmatrix} \quad (5.4)$$

### Same sign control region

Now that we have a way to estimate the kinematics and yields of the non-prompt background in a different region, a same sign CR enriched in fakes is usually defined in order to check this background. This will be done and detailed in Section 6.3.4 of this work.

### 5.6.5 Smaller backgrounds

Even though quite negligible, some additional backgrounds still need to be considered, such as the dibosons (WW, WZ and ZZ) and tribosons (WWW, WWZ, WZZ, ZZZ) productions, as shown in Figure 5.15. Backgrounds such as the  $W\gamma$  and Higgs related productions are also considered at this stage. All these smaller backgrounds are taken directly from MC.

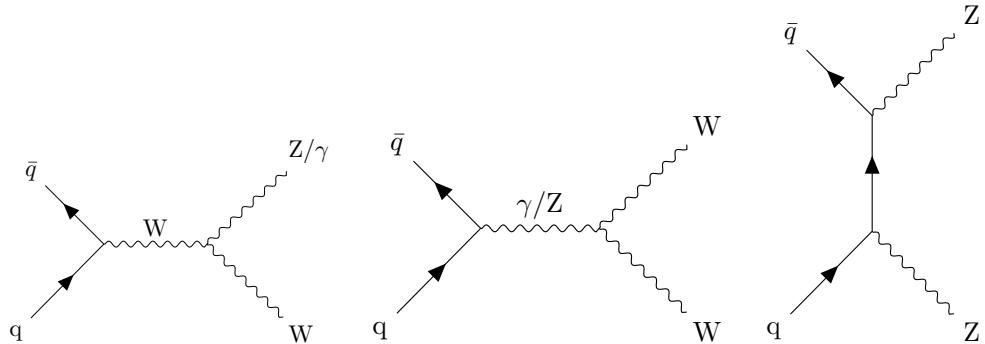


Figure 5.15: Possible Feynman diagrams for smaller backgrounds of this analysis: WW (on the left),  $W\gamma$  and WZ (on the center) and ZZ (on the right).

### 5.6.6 Weights and corrections applied

Several different weights and Scale Factors (SF) usually need to be applied to the different MC processes in order to account for several effects observed in data but not accounted for during the generation of the MC simulation, such as the efficiency of the selection of the different objects, which will be discussed in Chapter 6, usually directly provided by the different Physics Object Groups (POGs) for their own default objects definitions.

In this section, several weights that are applied to the MC samples will be detailed.

## Prefiring corrections

In 2016 and 2017 an issue, known as the **prefiring**, causing highly energetic readout from jets, photons and electrons in the ECAL endcap to be assigned by the Level-1 Trigger (L1) to the previous bunch crossing was discovered. To make up for this difference, a weight  $(1 - x)$  is usually applied to all MC events, where  $x$  is the probability of an event to be prefire [120].

The effect this correction has on the data/MC agreement in a 2017 top enriched control region is shown in Figure 5.16.

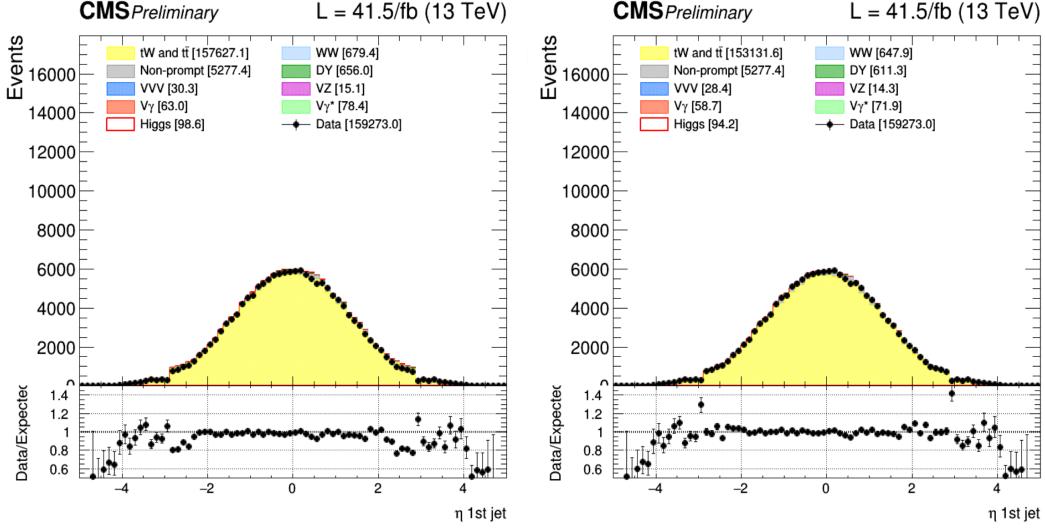


Figure 5.16: Data/MC agreement with (on the right) and without (on the left) prefiring corrections in a 2017 top control region.

## Top $p_T$ reweighting

Previous studies of the  $t\bar{t}$  generator typically considered in the physics analysis predict a harder top quark  $p_T$  spectrum than the one observed in data [121].

Section to be completed with plots showing the effect of the weights

# Chapter 6

## Event selection

This Chapter will be dedicated to the analysis itself, by defining first of all the different objects actually used in this case, along with the actual selection that has been applied to enhance the quality of such objects in this particular search in Section 6.1. Then, the different Signal Regions (SRs) defined in which a high purity of signal is expected are defined in Section 6.2 while all the different Control Regions (CRs) defined in order to check the behavior of the MC simulation performed for the major backgrounds on this analysis, such as the single top or SM  $t\bar{t}$  production, will be introduced in Section 6.3.

Finally will come a description about the different variables expected to naturally introduce some discrimination of the  $t/\bar{t}$  and  $t\bar{t}+{\rm DM}$  signals with respect to the different backgrounds in Section 6.4, along with a global description of the Machine Learning (ML) techniques employed in order to optimize the discriminating power of these variables in the best way possible.

### 6.1 Objects selection

We already described what to expect from a typical  $t/\bar{t}$  or  $t\bar{t}+{\rm DM}$  signal: the typical signature of such signals is made out of a certain number of b tagged jets along with two leptons (electrons and/or muons) and some Missing Transverse Energy (MET) coming from the two DM particles created along the way. It is therefore extremely important to describe the Working Point (WP) chosen and the selection applied in order to select the objects of the analysis, such as the leptons and the jets used, in such a way to optimize the lepton reconstruction efficiency while reducing as much as possible the possible misidentification rates of such objects.

First of all, the different triggers used to collect the data will be detailed in Section 6.1.1. Then, the leptons used in this analysis will be introduced in Sections 6.1.2 (for electrons) and 6.1.3 (for muons). Finally, given the nature of the DM signal searched for, a complete description of the jets selected in the analysis will be necessary and performed in Section 6.1.4.

#### 6.1.1 Triggers selection

The triggers, described in Section 3.2.6, and particularly the trigger paths chosen are an important part of each analysis since they will describe the kind of data that can be collected and therefore analyzed. The triggers used in this analysis for the years 2016, 2017 and 2018 can be found in Tables 6.1, 6.2 and 6.1 respectively.

Dataset	Run range	HLT trigger path
SingleMu	[273158,284044]	HLT_IsoMu24_v* HLT_IsoTkMu24_v*
SingleEle	[273158,284044]	HLT_Ele27_WPTight_Gsf_v* HLT_Ele25_eta2p1_WPTight_Gsf_v*
DoubleEG	[273158,284044]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
DoubleMu	[273158,281612] [281613,284044]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v* HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
MuonEG	[273158,278272] [278273,284044]	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 6.1: 2016 trigger paths considered for this analysis.

Our analysis relying on the dilepton final state, the single lepton trigger are only considered in order to recover some of the efficiency lost in some cases when one lepton passes the tight identification criteria while the second one does not, and does therefore not trigger the event. The logical *or* of all the trigger paths are usually considered. Eventual events passing several triggers is taken into account as well to make sure to avoid any double counting due to this effect.

These triggers have been studied in order to make sure that they are efficient enough in the  $p_T$  region of the leptons of the analysis to avoid any undesired effect due to the turn-on of any trigger. These trigger efficiencies, calculated using a general tag and probe method and found for example for different runs of the 2017 data taking period in Figure 6.1 for a DoubleEG trigger, are then used to reweight the simulated samples.

Dataset	Run range	HLT trigger path
SingleMu	[297046,306462]	HLT_IsoMu27_v*
EGamma	[297046,306462]	HLT_Ele35_WPTight_Gsf_v* HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[297046,299329] [299368,306462]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*
MuonEG	[297046,306462] [297046,299329] [299368,306462]	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 6.2: 2017 trigger paths considered for this analysis.

Dataset	Run range	HLT trigger path
SingleMu	[315252,325172]	HLT_IsoMu24_v*
		HLT_Mu5_v*
	[314859,325175]	HLT_IsoMu27_v*
EGamma	[315252,325172]	HLT_Ele32_WPTight_Gsf_v*
		HLT_Ele35_WPTight_Gsf_v*
		HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[315252,325172]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*
		HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*
MuonEG	[315252,325172]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
		HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 6.3: 2018 trigger paths considered for this analysis.

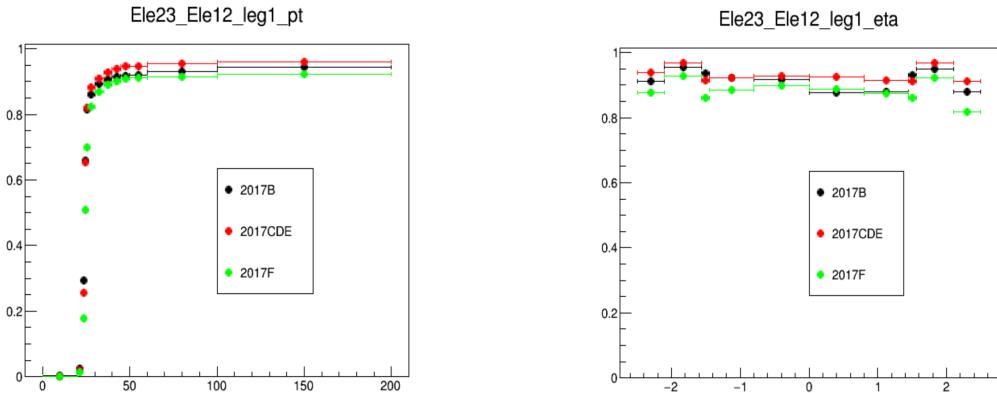


Figure 6.1: DoubleEG trigger efficiencies with respect to the  $p_T$  (on the left) and  $\eta$  (on the right), computed using a tag and probe method, for the 2017 data taking period.

### 6.1.2 Electrons selection

Several strategies are used in CMS in order to be able to identify prompt electrons and isolate this signal over background sources coming mainly from photon conversions, misidentification of jets or electrons coming from the semileptonic decay of the bottom and charm quarks. Several variables, which can be divided in the several following categories, allow to introduce some discrimination between these prompt and fake electrons:

- The **calorimetric observables** use the transverse shape of electromagnetic showers in the ECAL, the fact that these electromagnetic showers should be narrower than hadronic showers and the fraction of energy deposited in the HCAL and in the preshower/endcaps of the ECAL itself for the discrimination. Many different variables belong to this category, such as:
  - **hOverE** ( $\frac{H}{E}$ ), where  $H$  corresponds to the energy deposited in the HCAL and  $E$  the total energy deposited in the ECAL.
  - **ooEmooP** ( $\frac{1}{E_{SC}} - \frac{1}{p}$ ), where  $E_{SC}$  is the Super Cluster (SC) energy and  $p$  the momentum of the track at the point of closest approach to the PV.

- **dEtaInSee**  $\Delta\eta$  (**dPhiInSee**  $\Delta\phi$ ), the  $\eta$  ( $\phi$ ) difference between the SC and the inner track extrapolated from the interaction vertex.
- **sigmaIetaIeta** ( $\sigma_{\eta\eta}$ ), the weighted cluster Root Mean Square (RMS) inside 5x5 regions of SCs along  $\eta$ .
- The **isolation variables**, requiring the electron candidates to be quite isolated with respect to nearby energetic activity since most of the non-prompt electrons, such as electrons within a jet, are emitted with a large amount of surrounding energy.
  - The **relIsoWithEA** is the main variable that belongs to this category, corresponding to the PF isolation defined in a cone of size  $\Delta R = 0.3$  around the electron direction and relative to the electron  $p_T$ , and taking into account the PU contamination in this cone.
- The **tracking quality variables**, such as:
  - The **expected inner hits**, the number of pixels without corresponding hits in the trajectory of a reconstructed gsfTrack.
  - The **matched gsfTracks hits**, the  $\chi^2$  value calculated from the reconstructed gskTrack and its corresponding hits.
- The **conversion rejection variables**, mostly used to reject most of the photon conversion contamination when defining electrons, using variable such as:
  - The **transverse  $d_0$**  (or  $d_{xy}$ ) and **longitudinal  $d_z$  impact parameters**.
  - The **conversion veto**, checking if an electron candidate also matches at least one conversion candidate which also passes the selection cuts.

In this analysis, instead of relying on the four basic POG official Working Points (WPs) (veto, loose, medium and tight) that can be defined with some quality cuts, we rely on the MVA approach that consists in using a single discriminator variable to perform the discrimination between genuine and misidentified electrons, combining the information coming from more than 20 variables at once using Boosted Decision Trees (BDT).

Two WPs are then given directly by the CMS EGamma POG [122], corresponding to an electron selection efficiency of 80 and 90% respectively. For this analysis, the tight POG *mva\_90p\_Iso2016* (for 2016) and *mvaFall17V1Iso\_WP90* (for 2017 and 2018) WPs have been chosen for the electron definition, along with additional quality cuts defined in Tables 6.4 and 6.5 (as previously mentioned, these cuts sometimes differ quite a bit depending on whether the electron interacts with the ECAL barrel ( $|\eta| < 1.479$ ) or one of the endcaps ( $|\eta| \geq 1.479$ )).

The electron efficiencies computed for this particular selection can be seen in Figure 6.2.

### 6.1.3 Muons selection

The selection applied to muons is mostly based on the Muon POG, providing references efficiencies for standard selection, recommendations for the tight selection [123], with a few additional cuts on both the impact parameters.

At the end of the day, a muon can be labeled as **Tight POG** if:

- The PF muon reconstructed is a global muon.

Variable		Barrel cut ( $ \eta  \leq 1.479$ )	Endcap cut ( $ \eta  > 1.479$ )
<b>Basic selection</b>			
$ \eta $	<	-	2.5
<b>HLT safe selection</b>			
hOverE	<	0.060	0.065
ooEmooP	<	0.013	0.013
$ \text{dEtaInSee}(\Delta\eta) $	<	0.004	-
$ \text{dPhiInSee}(\Delta\phi) $	<	0.020	-
$\sigma_{\eta\eta}$	<	0.011	0.031
ecalPFClusterIso	<	0.160	0.120
hcalPFClusterISO	<	0.120	0.120
trackIso	<	0.08	0.08
GsfTrack $\chi^2/\text{NDOF}$	<	-	3.0
<b>Additional selection</b>			
lostHits	<	1	1
$d_{xy}$	<	0.05	0.1
$d_z$	<	0.1	0.2
pfRelIso03	<	0.0588	0.0571

Table 6.4: Quality cuts applied to define a 2016 tight electron in this analysis.

- The  $\chi^2/\text{NDOF}$  of the global muon track fit is less than 10 and at least one muon chamber hit is included in the global muon track fit
- Muon segments in at least two muon stations have been observed.
- Its tracker track has a transverse impact parameter  $d_{xy} < 0.2$  cm and a longitudinal impact parameter  $d_z < 0.5$  cm with respect to the PV .
- The number of pixel hits is larger than 0.
- At least 5 tracker layers with hits have been observed.

The selection applied to muons of this particular analysis is a bit tighter though, since the following cuts are applied on top of this selection:

- $p_T > 10$  GeV,  $|\eta| < 2.4$  and  $|d_z| < 0.1$  cm
- $|d_{xy}| < 0.01$  cm (if  $p_T < 20$  GeV) or  $|d_{xy}| < 0.02$  cm (if  $p_T \geq 20$  GeV)
- Tight muon isolation requirement ( $< 0.15$ ) with  $\Delta\beta$  correction and in a cone size  $\Delta R < 0.4$ , as defined in Equation 6.1, in order to reduce the number of muons coming from the hadronization process of bottom and charm quarks.

$$\text{ISO} = \frac{\sum p_T^{\text{ch. had. (PV)}} + \max(0, \sum E_T^{\text{neut. had.}} + \sum E_T^\gamma - 0.5 \times \sum p_T^{\text{ch. had. (PU)}})}{p_T(\mu)} \quad (6.1)$$

Variable		Barrel cut ( $ \eta  \leq 1.479$ )	Endcap cut ( $ \eta  > 1.479$ )
<b>Basic selection</b>			
$ \eta $	<	-	2.5
convVeto	=	1	1
$d_{xy}$	<	0.05	0.1
$d_z$	<	0.1	0.2
pfRelIso03	<	0.06	0.06

Table 6.5: Quality cuts applied to define a 2017/2018 tight electron in this analysis.

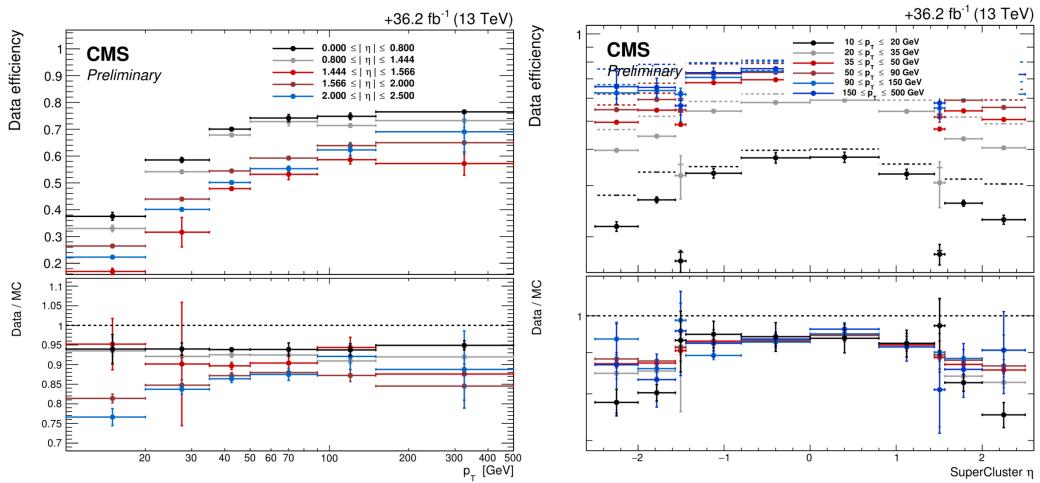


Figure 6.2: Tight electron efficiencies for this analysis, based on the Egamma POG `mva_90p_Iso2016` WP, for the data taking period of 2016.

This selection leads to efficiencies higher than 80% for all the muon momenta and pseudorapidities range, as shown in Figure 6.3.

#### 6.1.4 Jet selection

Jets are an important part of this analysis as well given the final state searched for in this case. As explained in Section 4.4, the jets are clustered from the PF candidates using the anti-kT algorithm (with a typical distance parameter  $R = 0.4$ ).

A selection is then applied on these jets in order to separate noise from real jets. In this analysis, this selection follows the tight WP definition given by the CMS JET/MET POG [124], whose selection depends on the pseudorapidity of the jet and on the year of data taking, as shown in Tables 6.6 and 6.7. The tight WP has been chosen since it offers an efficiency higher than 98-99% for all the jets and a background rejection higher than 98% for jets having  $|\eta| < 3.0$ .

A slightly different selection (loose PU jet ID) is applied to jets having a  $p_T > 50$  GeV in order to reject jets coming from PU interactions in 2017 and 2018, as shown in Table 6.8.

The b-jets are then also select using the recommendations given by the B-Tagging and Vertexing POG [125]. The loose deepCSV b-tagging WP is used in this case, as explained in Section 4.4.1, for which the rate for misidentifying a light jet as a b-jet is around 10%. A jet is considered a b-jet

$p_T/\eta$	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.0	0.0:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4
10:15	0.9995	1.0016	0.9985	0.9907	0.9844	1.0183	1.0089	1.0089	1.0265	0.9862	0.9808	0.9882	0.9927	0.9972
15:20	0.9805	0.9870	0.9950	0.9867	0.9935	0.9686	0.9927	0.9927	0.9771	0.9924	0.9826	0.9986	0.9912	0.9797
20:25	0.9756	0.9829	0.9949	0.9823	0.9914	0.9712	0.9906	0.9910	0.9590	0.9917	0.9776	0.9927	0.9873	0.9822
25:30	0.9793	0.9850	0.9942	0.9794	0.9912	0.9659	0.9899	0.9886	0.9610	0.9899	0.9784	0.9932	0.9874	0.9805
30:40	0.9773	0.9828	0.9947	0.9782	0.9891	0.9685	0.9903	0.9883	0.9614	0.9887	0.9770	0.9940	0.9876	0.9828
40:60	0.9766	0.9851	0.9957	0.9796	0.9908	0.9713	0.9911	0.9894	0.9638	0.9898	0.9786	0.9945	0.9871	0.9774
60:100	0.9618	0.9794	0.9963	0.9773	0.9917	0.9597	0.9914	0.9886	0.9655	0.9883	0.9764	0.9948	0.9858	0.9796
100:200	0.9803	0.9743	0.9831	0.9740	0.9949	0.9634	0.9954	0.9814	0.9482	0.9870	0.9732	0.9962	0.9782	0.9895

Color Definition

Scale Factors	0.99-1.01	0.98-0.99,1.01-1.02	0.97-0.98,1.02-1.03	0.96-0.97	0.95-0.96	< 0.95
---------------	-----------	---------------------	---------------------	-----------	-----------	--------

(a) 2016 muon efficiencies

$p_T/\eta$	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.0	0.0:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4
10:15	0.9880	0.9827	0.9981	0.9961	1.0018	1.0099	1.0129	1.0129	1.0193	0.9893	0.9767	0.9978	0.9901	0.9852
15:20	0.9735	0.9831	0.9955	0.9849	0.9995	0.9582	0.9932	0.9932	0.9400	0.9891	0.9813	0.9970	0.9899	0.9807
20:25	0.9737	0.9839	0.9976	0.9863	0.9940	0.9561	0.9936	0.9963	0.9757	0.9915	0.9800	0.9975	0.9892	0.9771
25:30	0.9741	0.9819	0.9975	0.9852	0.9931	0.9580	0.9936	0.9952	0.9619	0.9914	0.9790	0.9979	0.9905	0.9826
30:40	0.9719	0.9794	0.9973	0.9842	0.9935	0.9570	0.9932	0.9942	0.9635	0.9921	0.9781	0.9972	0.9900	0.9815
40:60	0.9709	0.9815	0.9979	0.9849	0.9941	0.9556	0.9945	0.9953	0.9636	0.9924	0.9790	0.9966	0.9896	0.9800
60:100	0.9622	0.9779	0.9959	0.9842	0.9935	0.9525	0.9938	0.9944	0.9563	0.9918	0.9747	0.9967	0.9868	0.9887
100:200	0.9649	0.9732	0.9899	0.9851	0.9884	0.9645	1.0060	0.9953	0.9831	0.9872	0.9738	0.9981	0.9868	0.9635

Color Definition

Scale Factors	0.99-1.01	0.98-0.99	0.97-0.98	0.96-0.97	0.95-0.96	< 0.95
---------------	-----------	-----------	-----------	-----------	-----------	--------

(b) 2017 muon efficiencies

$p_T/\eta$	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.0	0.0:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4
10:15	0.9686	0.9806	0.9996	0.9801	0.9975	0.9895	1.0053	1.0053	0.9446	0.9908	0.9763	0.9984	0.9967	0.9726
15:20	0.9683	0.9817	0.9974	0.9836	0.9951	0.9513	0.9925	0.9925	0.9678	0.9901	0.9849	0.9945	0.9888	0.9810
20:25	0.9681	0.9826	0.9979	0.9874	0.9963	0.9670	0.9936	0.9986	0.9725	0.9921	0.9835	0.9967	0.9899	0.9786
25:30	0.9679	0.9821	0.9977	0.9858	0.9960	0.9779	0.9943	0.9918	0.9795	0.9941	0.9804	0.9964	0.9890	0.9742
30:40	0.9688	0.9789	0.9970	0.9848	0.9956	0.9778	0.9933	0.9946	0.9740	0.9930	0.9814	0.9957	0.9885	0.9749
40:60	0.9672	0.9802	0.9967	0.9856	0.9964	0.9752	0.9950	0.9952	0.9743	0.9943	0.9821	0.9958	0.9875	0.9761
60:100	0.9652	0.9756	0.9959	0.9835	0.9957	0.9725	0.9919	0.9946	0.9760	0.9931	0.9803	0.9945	0.9853	0.9702
100:200	0.9500	0.9750	0.9944	0.9839	0.9976	0.9688	0.9984	0.9890	0.9460	0.9925	0.9778	0.9954	0.9780	0.9881

Color Definition

Scale Factors	0.99-1.01	0.98-0.99	0.97-0.98	0.96-0.97	0.95-0.96	< 0.95
---------------	-----------	-----------	-----------	-----------	-----------	--------

(c) 2018 muon efficiencies

Figure 6.3: Tight muon efficiencies for this analysis, based on the Muon POG tight WP with additional cuts for 2016, 2017 and 2018.

in this analysis if it passes the following cuts:

- $p_T > 20 \text{ GeV}$
- $|\eta| < 2.4$
- $\text{deepCSV} > 0.1241$

## 6.2 Signals regions

It is important to note that a strict **blinding policy** has been followed for this search, in order to avoid optimizing the analysis based on what has already seen. The data available to be plotted in the following signal regions has therefore been limited to  $1 \text{ fb}^{-1}$  for each year.

Two signal regions have been defined in order to isolate on one hand the  $t/\bar{t}+\text{DM}$  process and on the other hand the  $t\bar{t}+\text{DM}$  signal.

Variable	$ \eta  \leq 2.4$	$ \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  > 3.0$
Neutral Hadron Fraction	< 0.90	< 0.90	< 0.98	-
Neutral EM Fraction	< 0.99	< 0.99	> 0.01	< 0.90
Number of Constituents	> 1	> 1	-	-
Charged Hadron Fraction	> 0	-	-	-
Charged Multiplicity	> 0	-	-	-
Charged EM Fraction	< 0.99	-	-	-
Number of Neutral Particles	-	-	> 2	> 10

Table 6.6: JET/MET POG tight WP for 2016.

Variable	$ \eta  \leq 2.4$	$ \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  > 3.0$
Neutral Hadron Fraction	< 0.90	< 0.90	-	> 0.02
Neutral EM Fraction	< 0.90	< 0.90	> 0.02 and < 0.99	< 0.90
Number of Constituents	> 1	> 1	-	-
Charged Hadron Fraction	> 0	-	-	-
Charged Multiplicity	> 0	> 0	-	-
Number of Neutral Particles	-	-	> 2	> 10

Table 6.7: JET/MET POG tight WP for 2017 and 2018.

### 6.2.1 $t/\bar{t}$ +DM region

As shown in Figure 6.4

To be added once defined

### 6.2.2 $t\bar{t}$ +DM region

As shown in Figure 6.5

Variable	$ \eta  \leq 2.4$	$ \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  > 3.0$
Neutral Hadron Fraction	< 0.90	< 0.90	< 0.99	> 0.02
Neutral EM Fraction	< 0.90	< 0.90	-	< 0.90
Number of Constituents	> 1	> 1	-	-
Charged Hadron Fraction	> 0	-	-	-
Charged Multiplicity	> 0	-	-	-
Number of Neutral Particles	-	-	-	> 2 and < 15

Table 6.8: JET/MET POG tight WP for 2017 and 2018, for jets having a  $p_T > 50$  GeV.

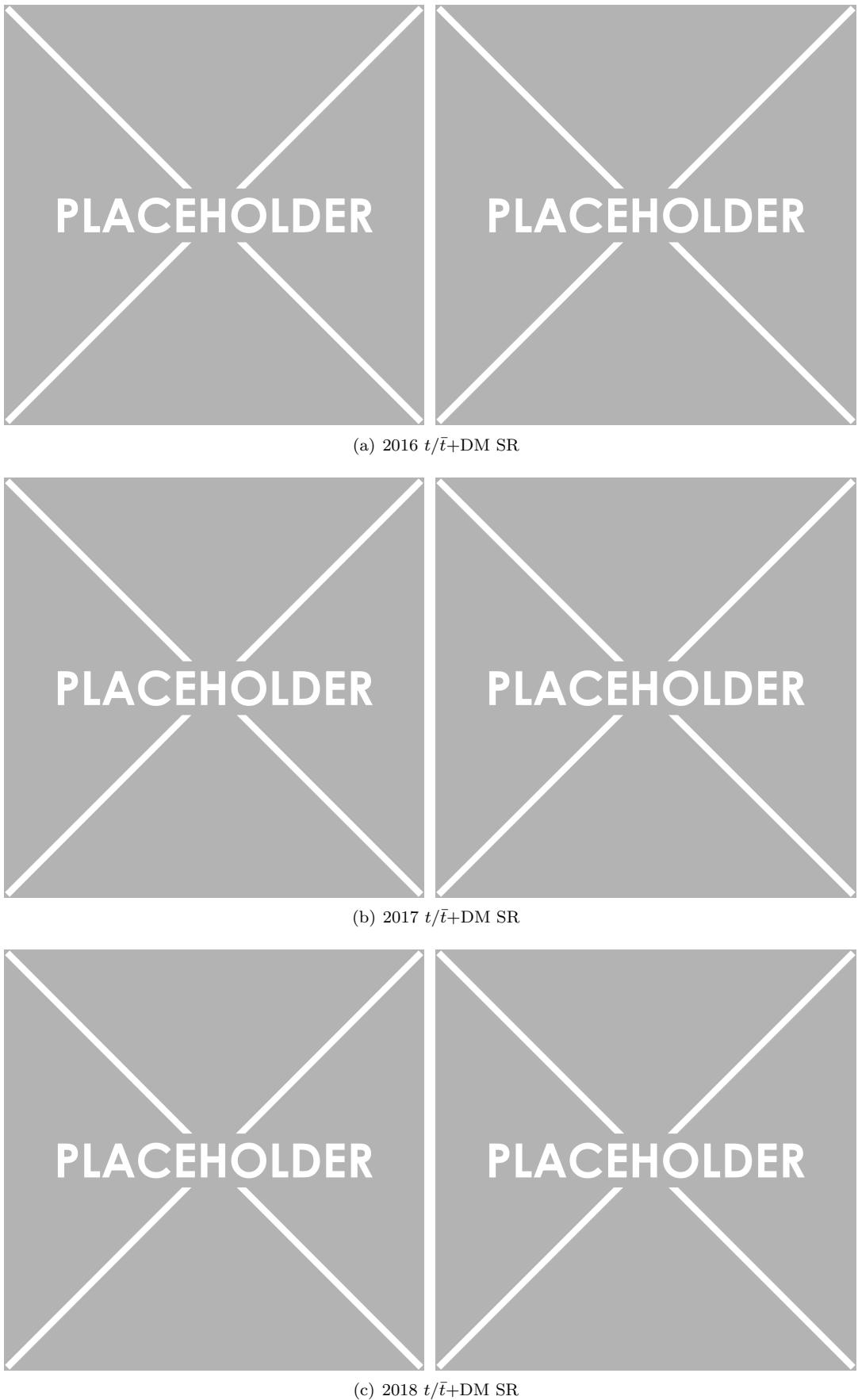


Figure 6.4: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the  $t/\bar{t}$ +DM process.

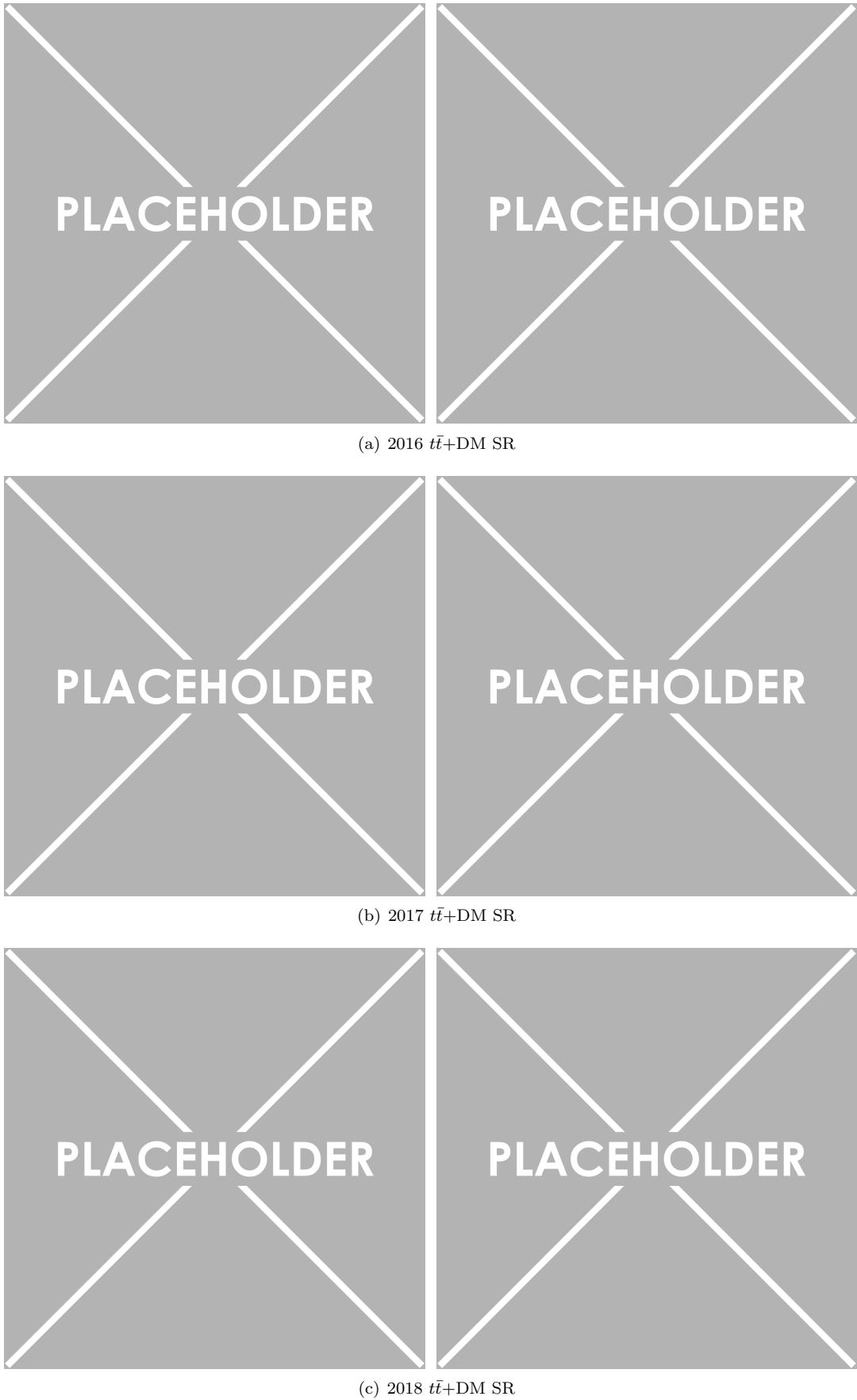


Figure 6.5: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the  $t\bar{t}+DM$  process.

To be added once defined

## 6.3 Control regions

Different control regions where the data/MC agreement can be checked have been defined in order to check the validity of the MC simulations performed for the SM processes corresponding to the main backgrounds of this analysis.

In this context, the DY control region will be defined in Section 6.3.1 while the SM  $t\bar{t}$ , WW and ttV processes will be checked in the Sections 6.3.2, ?? and 6.3.3 respectively.

### 6.3.1 DY control region

As shown in Figure 6.6

To be added once defined

### 6.3.2 Top control region

As shown in Figure 6.7

To be added once defined

### 6.3.3 ttV control region

As shown in Figure 6.8

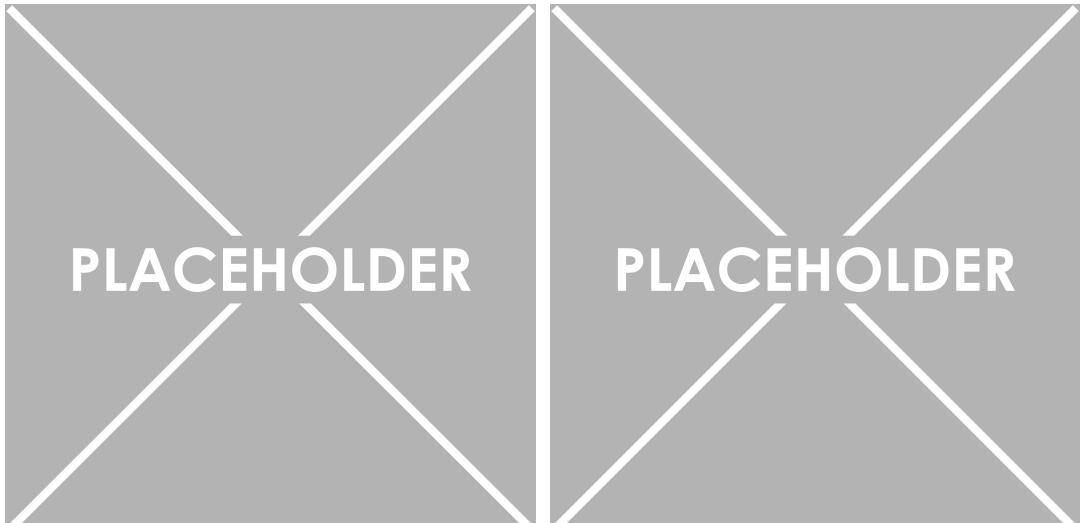
To be added once defined

### 6.3.4 Same sign control region

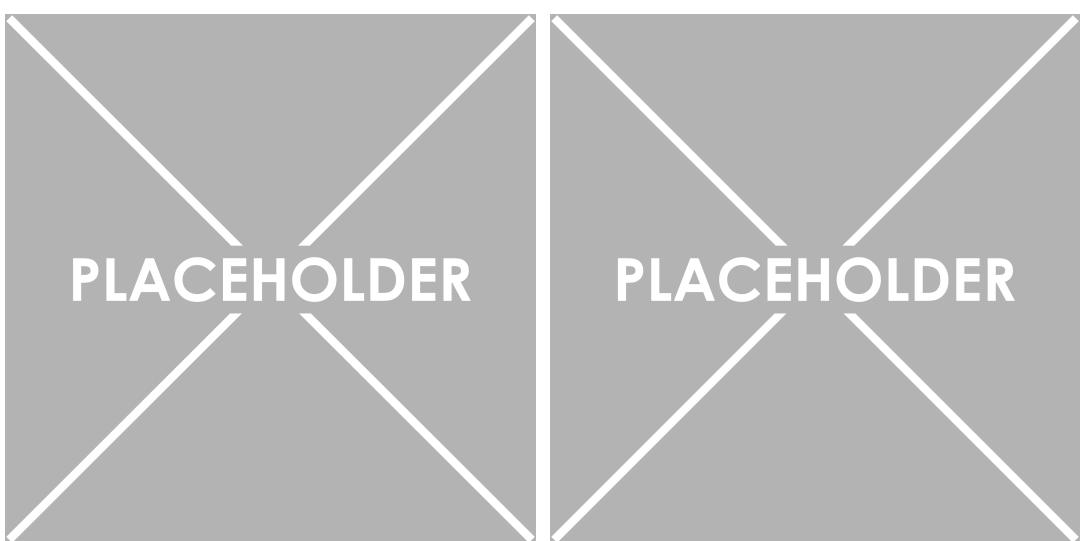
A same sign control region has also been defined in order to check the non-prompt background, calculated using a data-driven tight-to-loose method described in Section 5.6.4. This CR is defined with the following cuts:

- Exactly 2 same sign leptons
- $p_{T,1} > 20$  (25) GeV for  $e$  ( $\mu$ )
- $p_{T,2} > 13$  GeV
- $|\eta| < 2.5$  for both leptons
- $m_{ll} > 12$  GeV
- PuppiMET  $> 20$  GeV
- $p_T^{ll} > 30$  GeV
- $m_{th} > 60$  GeV

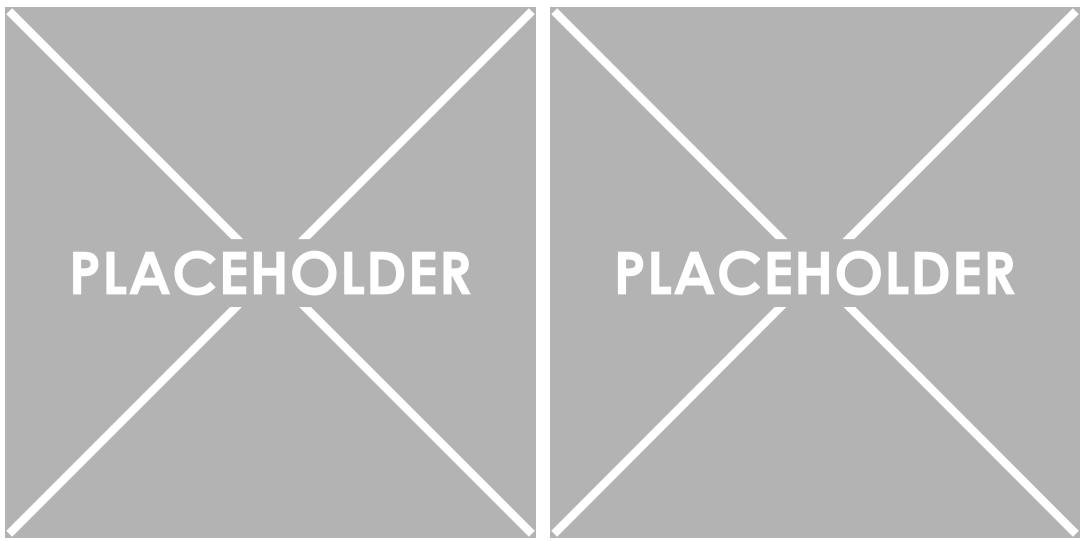
Several regions are then defined according to this cut, depending on the data taking period, on the channel, on the number of jets observed and on the  $p_T$  of the second lepton. Some plots can be found in Figure 6.9.



(a) 2016 DY CR

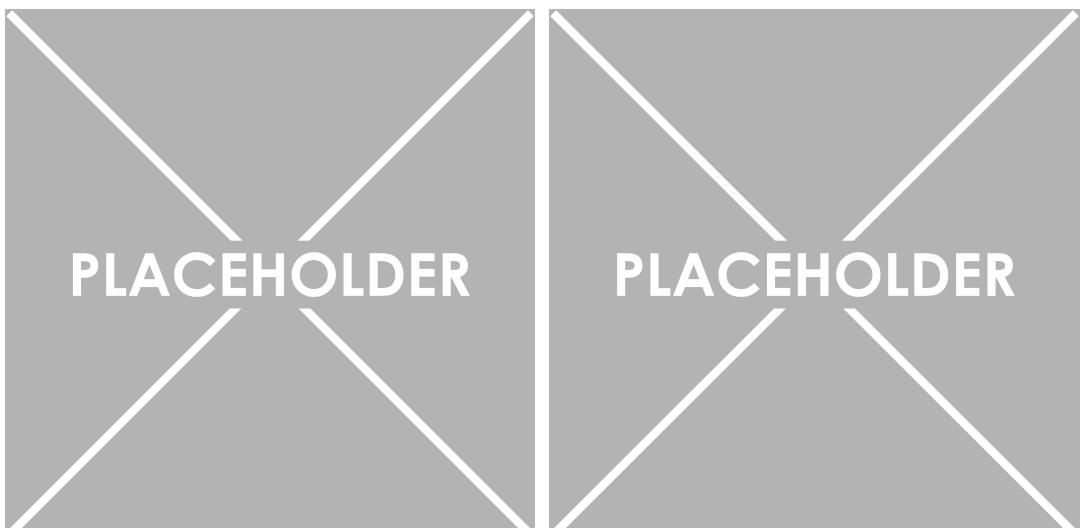


(b) 2017 DY CR

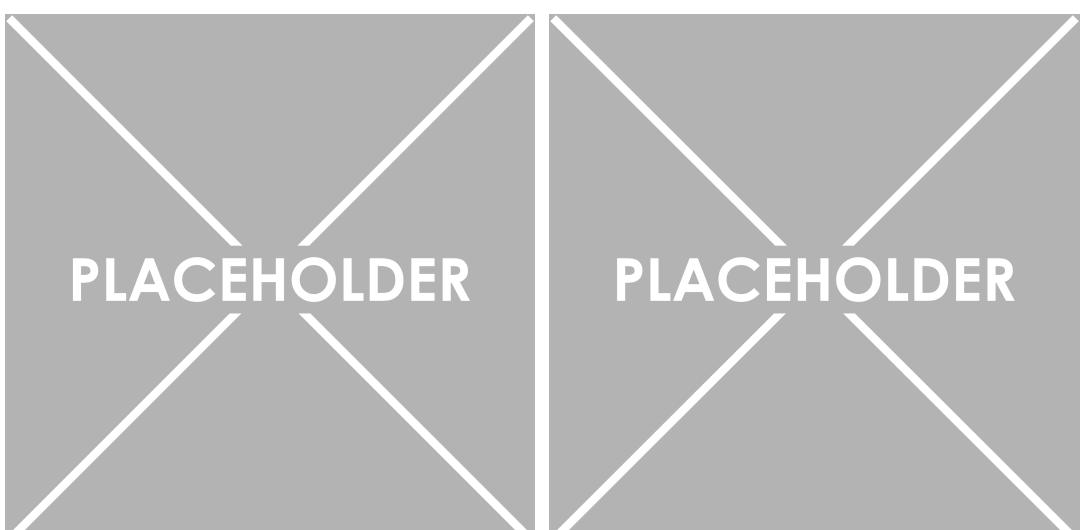


(c) 2018 DY CR

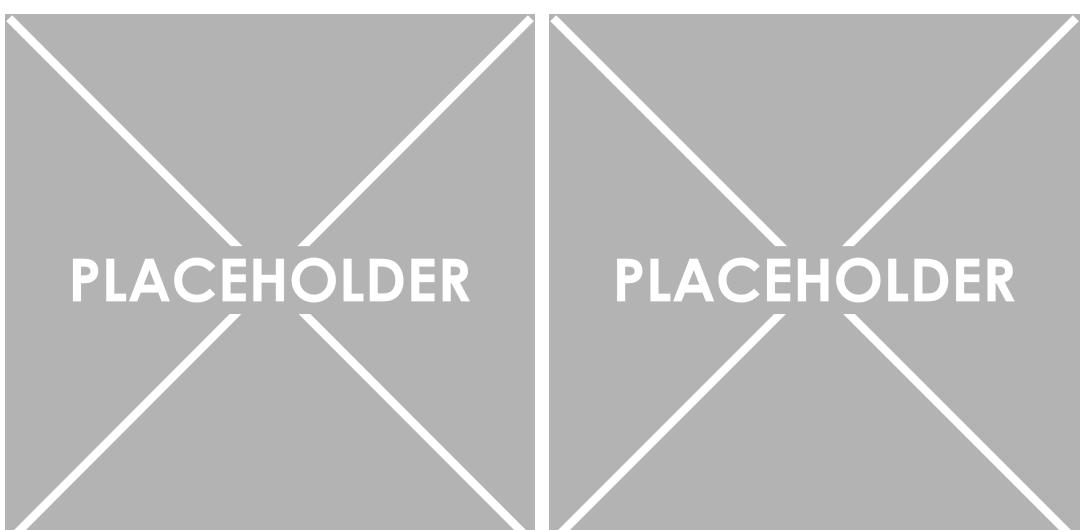
Figure 6.6: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the DY control region defined.



(a) 2016 top CR

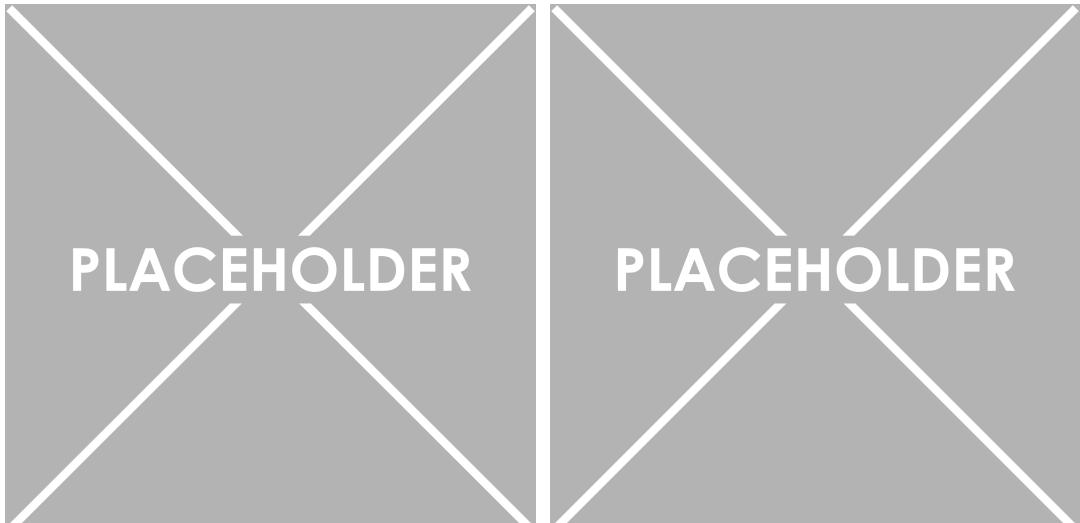


(b) 2017 top CR

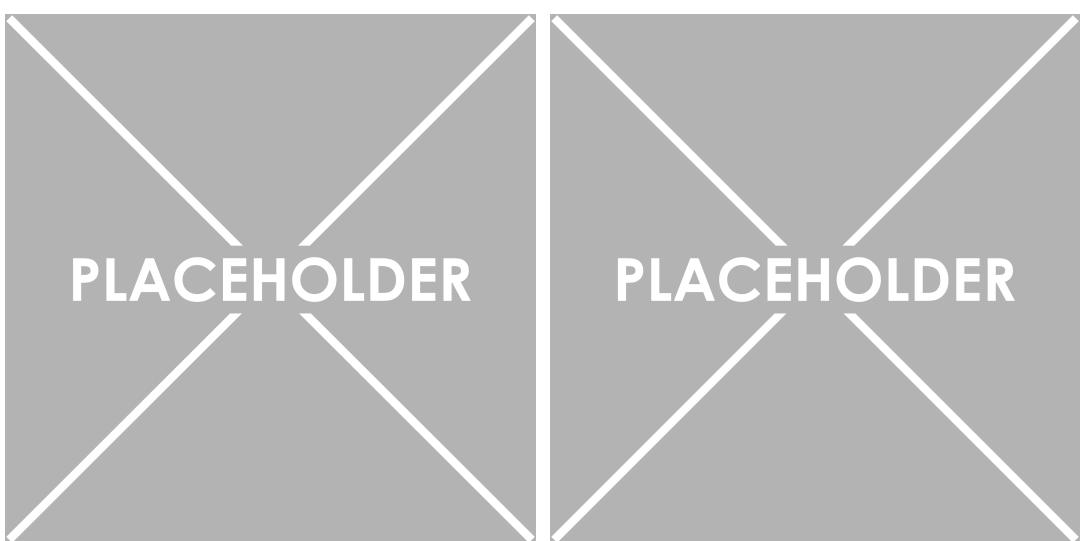


(c) 2018 top CR

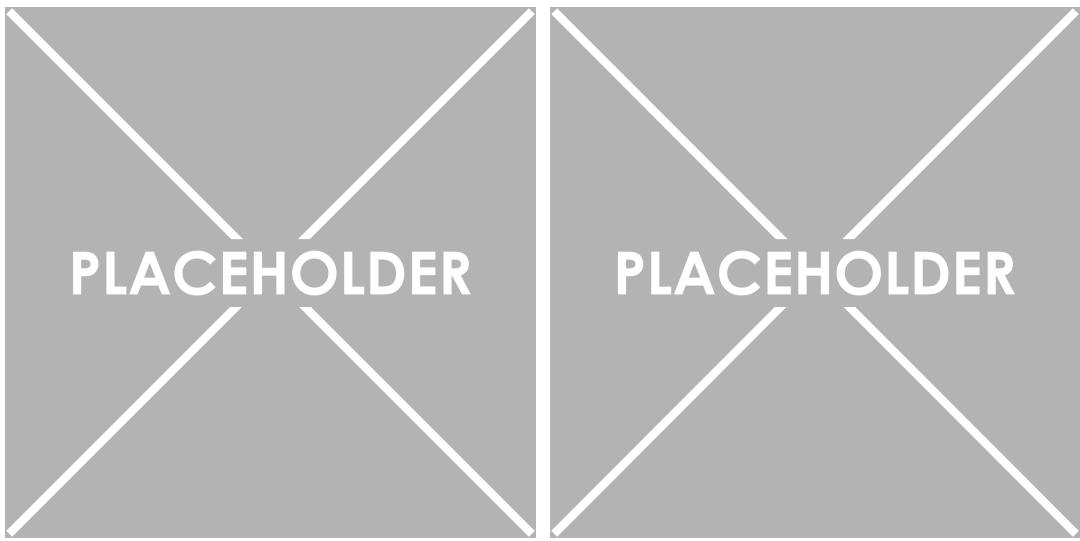
Figure 6.7: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the top control region defined.



(a) 2016 ttV CR



(b) 2017 ttV CR



(c) 2018 ttV CR

Figure 6.8: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the ttV control region defined.

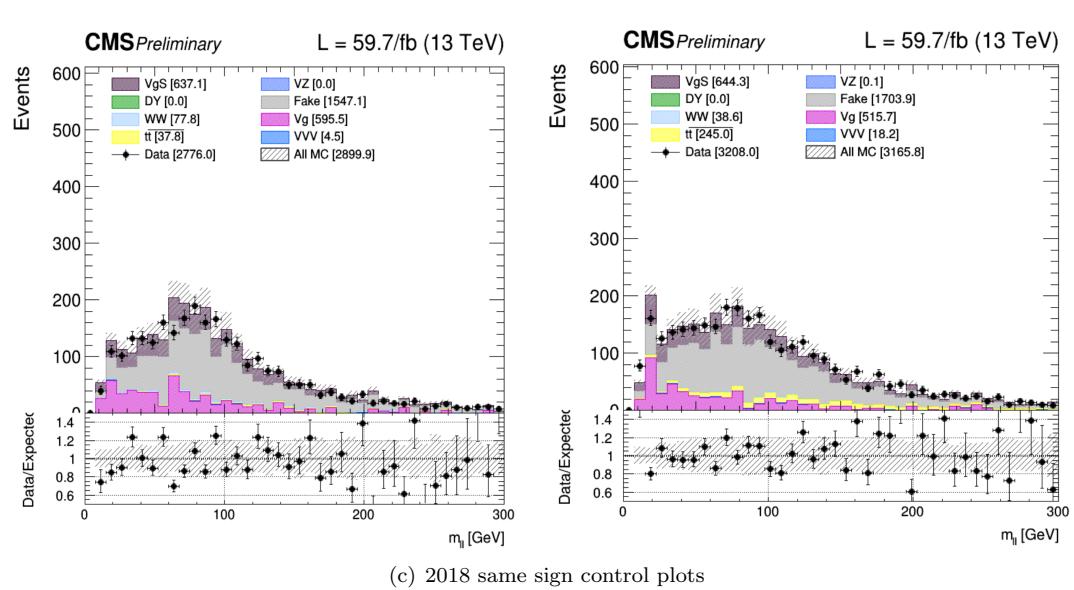
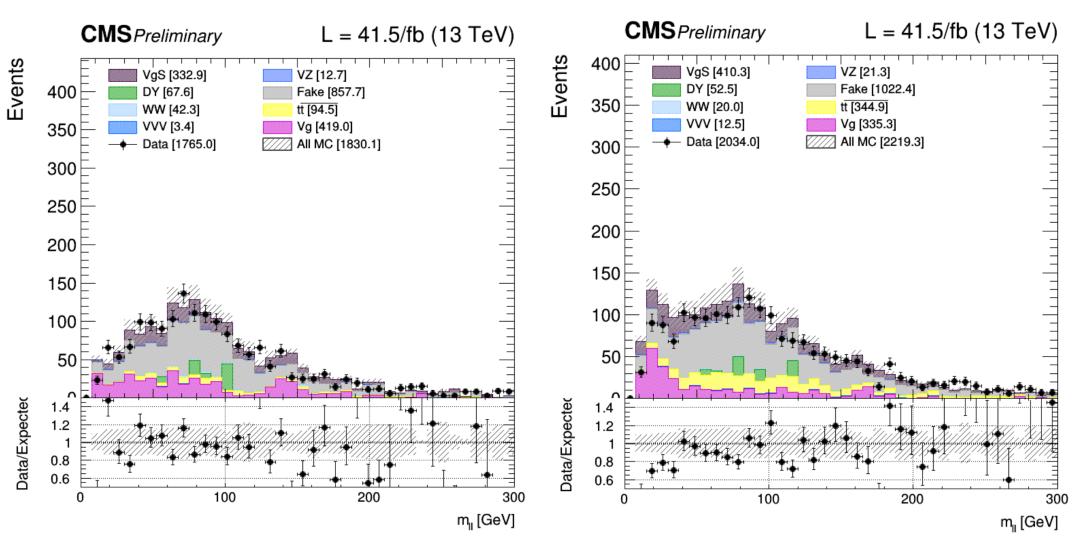
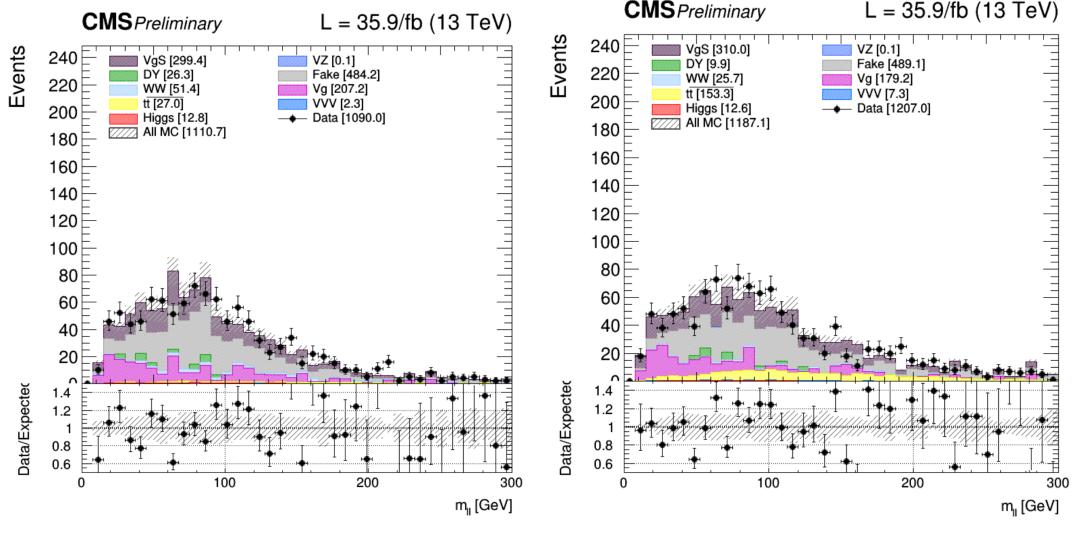


Figure 6.9: Same sign control region  $m_{ll}$  distributions for the years 2016, 2017 and 2018 and for the  $e\mu$  channel and for the 0-jet (on the left) and 1j (on the right) categories.

In order to cover for most of the discrepancies between the data and the simulation observed in the distributions of this control region, a flat 30% systematic uncertainty is typically associated to this background, as will be discussed in more details in Section 7.1.

## 6.4 Background-signal discrimination

### 6.4.1 Discriminating variables

#### Missing Transverse Energy (MET)

This variable has already been defined in Section 4.5, and corresponds to the imbalance in transverse momentum which can be left by different phenomena, such as the apparition of a SM neutrino or the existence of DM particles, able to escape the detector without being detected.

This variable is expected to induce some discrimination between the signal and the backgrounds because, even tough the  $t\bar{t}$  in the dilepton final state is expected to produce two neutrinos and therefore some MET, the  $t\bar{t}$ +DM signal model is expected to have mostly the same contribution to the MET from its own two neutrinos, and an additional contributions from the pair  $\chi\bar{\chi}$  produced. The MET spectrum is therefore expected to reach higher values for the signal than the backgrounds.

**TALK ABOUT SINGLE TOP?**

#### Stransverse mass

The  $m_{T2}$  variable, also called **stransverse mass**, is an extension of the definition of the transverse mass  $m_T$  to cases when pairs of particles with the same flavor decay into one visible and one invisible particle, such as what happens in the  $W \rightarrow l\nu$  decay, for example.

In this particular case, two particles contribute to the presence of Missing Transverse Energy (MET) and the individual contribution of each particle ( $\not{p}_{T_1}$  and  $\not{p}_{T_2}$ ) to this missing energy cannot be inferred. The stransverse mass is then defined according to Equation 6.2, where  $\not{p}_{T_i} = \overrightarrow{p_{T_i}}$  is the (visible) transverse momentum of the particle  $i$  and  $\alpha$  is the angle between the visible and invisible  $p_T$  of the decay considered [114].

$$\begin{cases} M_{T2}^2 = \min_{\not{p}_{T_1} + \not{p}_{T_2} = \not{p}_{T_{\text{tot}}}} \left( \max \left( m_T^2(\not{p}_{T_1}, \not{p}_{T_1}), m_T^2(\not{p}_{T_2}, \not{p}_{T_2}) \right) \right) \\ m_T^2(\not{p}_T, \not{p}'_T) = 4 |\not{p}_T| |\not{p}'_T| \sin^2 \left( \frac{\alpha}{2} \right) \end{cases} \quad (6.2)$$

This equation can be understood in the following way: to compute the  $m_{T2}$  variable, different combinations  $(\not{p}_{T_1}, \not{p}_{T_2})$  satisfying the condition  $\not{p}_{T_1} + \not{p}_{T_2} = \not{p}_{T_{\text{tot}}}$  need to be probed, keeping only the combination which results in the lowest value.

In this particular analysis,  $M_{T2}(ll)$  is calculated, since the role of the visible particles is played by the two final state leptons. This variable is expected to introduce some discrimination because, according to the definition just given, the  $M_{T2}(ll)$  variable for a SM  $t\bar{t}$  process is expected to have an endpoint exactly at the mass of the W boson, while an eventual  $t\bar{t}$ +DM signal does not have

this limitation in the  $M_{T2}(ll)$  spectrum because of the pair of DM particles produced, which also contributes to the total MET of the event.

However, in practice, we do observe a tail in this spectrum even for SM  $t\bar{t}$  without DM, because of the instrumental MET sometimes observed or the fact that some selected leptons are not actually prompt leptons but can be jets misidentified as leptons by the detector.

TALK ABOUT SINGLE TOP?

#### 6.4.2 Neural network



# Chapter 7

## Results and interpretations

### 7.1 Systematics and uncertainties

### 7.2 Results



# Chapter 8

## Conclusions

### 8.1 Future prospects



# **Appendices**



# Appendix A

## Samples used

### A.1 Data samples

All the data samples considered for this analysis are listed in Tables A.1, A.2 and A.3. The luminosity of each dataset has been computed using the Brilcalc tool provided by CMS [126], while the number of generated events has been obtained using the CERN official Data Aggregation System (DAS).

### A.2 Signal samples

To be completed once the files are actually available

### A.3 Backgrounds samples

All the background MC samples considered for this analysis are listed in Tables A.4, A.5 and A.6 for 2016, 2017 and 2018 respectively.

Dataset	Events (size)	$\mathcal{L}$ [fb $^{-1}$ ]
<b>Run 2016B</b>		
/DoubleEG/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	143073268 (99.4Gb)	
/DoubleMuon/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	82535526 (53.2Gb)	
/MuonEG/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	32727796 (26.8Gb)	5.8
/SingleElectron/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	246440440 (167.8Gb)	
/SingleMuon/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	158145722 (96.4Gb)	
<b>Run 2016C</b>		
/DoubleEG/Run2016C-Nano1June2019-v1/NANOAOD	47677856 (35.3Gb)	
/DoubleMuon/Run2016C-Nano1June2019-v1/NANOAOD	27934629 (19.7Gb)	
/MuonEG/Run2016C-Nano1June2019-v1/NANOAOD	15405678 (12.8Gb)	2.6
/SingleElectron/Run2016C-Nano1June2019-v1/NANOAOD	97259854 (69.3Gb)	
/SingleMuon/Run2016C-Nano1June2019-v1/NANOAOD	67441308 (42.4Gb)	
<b>Run 2016D</b>		
/DoubleEG/Run2016D-Nano1June2019-v1/NANOAOD	53324960 (39.6Gb)	
/DoubleMuon/Run2016D-Nano1June2019-v1/NANOAOD	33861745 (24.1Gb)	
/MuonEG/Run2016D-Nano1June2019-v1/NANOAOD	23482352 (19.4Gb)	4.2
/SingleElectron/Run2016D-Nano1June2019-v1/NANOAOD	148167727 (104.4Gb)	
/SingleMuon/Run2016D-Nano1June2019-v1/NANOAOD	98017996 (61.3Gb)	
<b>Run 2016E</b>		
/DoubleEG/Run2016E-Nano1June2019-v1/NANOAOD	49877710 (37.9Gb)	
/DoubleMuon/Run2016E-Nano1June2019-v1/NANOAOD	28246946 (20.8Gb)	
/MuonEG/Run2016E-Nano1June2019-v2/NANOAOD	22519303 (19.0Gb)	4.0
/SingleElectron/Run2016E-Nano1June2019-v1/NANOAOD	117321545 (86.5Gb)	
/SingleMuon/Run2016E-Nano1June2019-v1/NANOAOD	90984718 (58.7Gb)	
<b>Run 2016F</b>		
/DoubleEG/Run2016F-Nano1June2019-v1/NANOAOD	34577629 (26.9Gb)	
/DoubleMuon/Run2016F-Nano1June2019-v1/NANOAOD	20329921 (15.3Gb)	
/MuonEG/Run2016F-Nano1June2019-v1/NANOAOD	16002165 (13.6Gb)	3.1
/SingleElectron/Run2016F-Nano1June2019-v1/NANOAOD	70593532 (51.4Gb)	
/SingleMuon/Run2016F-Nano1June2019-v1/NANOAOD	65489554 (42.4Gb)	
<b>Run 2016G</b>		
/DoubleEG/Run2016G-Nano1June2019-v1/NANOAOD	78797031 (61.6Gb)	
/DoubleMuon/Run2016G-Nano1June2019-v1/NANOAOD	45235604 (34.2Gb)	
/MuonEG/Run2016G-Nano1June2019-v1/NANOAOD	33854612 (29.0Gb)	7.6
/SingleElectron/Run2016G-Nano1June2019-v1/NANOAOD	153363109 (109.2Gb)	
/SingleMuon/Run2016G-Nano1June2019-v1/NANOAOD	149912248 (94.6Gb)	
<b>Run 2016H</b>		
/DoubleEG/Run2016H-Nano1June2019-v1/NANOAOD	85388734 (67.7Gb)	
/DoubleMuon/Run2016H-Nano1June2019-v1/NANOAOD	48912812 (37.3Gb)	
/MuonEG/Run2016H-Nano1June2019-v1/NANOAOD	29236516 (26.0Gb)	8.6
/SingleElectron/Run2016H-Nano1June2019-v1/NANOAOD	128854598 (93.8Gb)	
/SingleMuon/Run2016H-Nano1June2019-v1/NANOAOD	174035164 (110.2Gb)	

Table A.1: Datasets collected in 2016 and considered for this analysis.

Dataset	Events (size)	$\mathcal{L}$ [fb $^{-1}$ ]
<b>Run 2017B</b>		
/DoubleEG/Run2017B-Nano1June2019-v1/NANOAOD	58088760 (46.6Gb)	
/DoubleMuon/Run2017B-Nano1June2019-v1/NANOAOD	14501767 (10.8Gb)	
/SingleElectron/Run2017B-Nano1June2019-v1/NANOAOD	60537490 (42.2Gb)	4.8
/SingleMuon/Run2017B-Nano1June2019-v1/NANOAOD	136300266 (86.2Gb)	
/MuonEG/Run2017B-Nano1June2019-v1/NANOAOD	4453465 (4.1Gb)	
<b>Run 2017C</b>		
/DoubleEG/Run2017C-Nano1June2019-v1/NANOAOD	65181125 (53.8Gb)	
/DoubleMuon/Run2017C-Nano1June2019-v1/NANOAOD	49636525 (39.5Gb)	
/SingleElectron/Run2017C-Nano1June2019-v1/NANOAOD	136637888 (102.5Gb)	9.7
/SingleMuon/Run2017C-Nano1June2019-v1/NANOAOD	165652756 (109.5Gb)	
/MuonEG/Run2017C-Nano1June2019-v1/NANOAOD	15595214 (15.0Gb)	
<b>Run 2017D</b>		
/DoubleEG/Run2017D-Nano1June2019-v1/NANOAOD	25911432 (21.6Gb)	
/DoubleMuon/Run2017D-Nano1June2019-v1/NANOAOD	23075733 (18.6Gb)	
/SingleElectron/Run2017D-Nano1June2019-v1/NANOAOD	51526710 (38.5Gb)	4.2
/SingleMuon/Run2017D-Nano1June2019-v1/NANOAOD	70361660 (47.2Gb)	
/MuonEG/Run2017D-Nano1June2019-v1/NANOAOD	9164365 (8.9Gb)	
<b>Run 2017E</b>		
/DoubleEG/Run2017E-Nano1June2019-v1/NANOAOD	56233597 (49.8Gb)	
/DoubleMuon/Run2017E-Nano1June2019-v1/NANOAOD	51589091 (44.4Gb)	
/SingleElectron/Run2017E-Nano1June2019-v1/NANOAOD	102121689 (81.3Gb)	9.3
/SingleMuon/Run2017E-Nano1June2019-v1/NANOAOD	154630534 (111.0Gb)	
/MuonEG/Run2017E-Nano1June2019-v1/NANOAOD	19043421 (19.2Gb)	
<b>Run 2017F</b>		
/DoubleEG/Run2017F-Nano1June2019-v1/NANOAOD	74307066 (67.1Gb)	
/DoubleMuon/Run2017F-Nano1June2019-v1/NANOAOD	79756560 (68.0Gb)	
/SingleElectron/Run2017F-Nano1June2019-v1/NANOAOD	128467223 (105.2Gb)	13.5
/SingleMuon/Run2017F-Nano1June2019-v1/NANOAOD	242135500 (178.3Gb)	
/MuonEG/Run2017F-Nano1June2019-v1/NANOAOD	25776363 (26.3Gb)	

Table A.2: Datasets collected in 2017 and considered for this analysis.

Dataset	Events (size)	$\mathcal{L} [\text{fb}^{-1}]$
<b>Run 2018A</b>		
/DoubleMuon/Run2018A-Nano25Oct2019-v1/NANO AOD	75499908 (62.6Gb)	
/EGamma/Run2018A-Nano25Oct2019-v1/NANO AOD	327843843 (261.8Gb)	13.5
/SingleMuon/Run2018A-Nano25Oct2019-v1/NANO AOD	241608232 (167.7Gb)	
/MuonEG/Run2018A-Nano25Oct2019-v1/NANO AOD	32958503 (32.3Gb)	
<b>Run 2018B</b>		
/DoubleMuon/Run2018B-Nano25Oct2019-v1/NANO AOD	35057758 (28.3Gb)	
/EGamma/Run2018B-Nano25Oct2019-v1/NANO AOD	153822427 (123.1Gb)	6.8
/SingleMuon/Run2018B-Nano25Oct2019-v1/NANO AOD	119918017 (82.3Gb)	
/MuonEG/Run2018B-Nano25Oct2019-v1/NANO AOD	16211567 (15.8Gb)	
<b>Run 2018C</b>		
/DoubleMuon/Run2018C-Nano25Oct2019-v1/NANO AOD	34565869 (27.6Gb)	
/EGamma/Run2018C-Nano25Oct2019-v1/NANO AOD	147827904 (119.2Gb)	6.6
/SingleMuon/Run2018C-Nano25Oct2019-v1/NANO AOD	110032072 (75.7Gb)	
/MuonEG/Run2018C-Nano25Oct2019-v1/NANO AOD	15652198 (15.3Gb)	
<b>Run 2018D</b>		
/DoubleMuon/Run2018D-Nano25Oct2019_ver2-v1/NANO AOD	168605834 (128.6Gb)	
/EGamma/Run2018D-Nano25Oct2019-v1/NANO AOD	751348648 (583.6Gb)	32.0
/SingleMuon/Run2018D-Nano25Oct2019-v1/NANO AOD	513867253 (344.5Gb)	
/MuonEG/Run2018D-Nano25Oct2019_ver2-v1/NANO AOD	71961587 (68.6Gb)	

Table A.3: Datasets collected in 2018 and considered for this analysis.

Process	Sample	Cross section [pb]
DY	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18610.0
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6025.20
TTTo2L2Nu	TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.360
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	26.38
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	44.33
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.60
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.60
WW	WWTo2L2Nu_13TeV-powheg	12.178
	WWJJToLNuLNu_EWK_noTop_13TeV-madgraph-pythia8	0.34520
	GluGluWWTo2L2Nu_MCFM_13TeV	0.5905
V $\gamma$ /V $\gamma^*$	WGToLNuG_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	405.271
	ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	131.300
	WZTo3LNu_mllmin01_13TeV-powheg-pythia8	58.59
VZ	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_powheg_pythia8	3.22
	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.212
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
VVV	ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.01398
	WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.05565
	WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.16510
	WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.18331
Non-Prompt	Data-driven (tight-to-loose method)	

Table A.4: Main 2016 MC simulations for the different background processes considered for this analysis and their respective cross sections.

Process	Sample	Cross section [pb]
DY	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	6189.39
TTTo2L2Nu	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_mtop1715_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	3.360
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	26.38
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	44.33
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60
WW	WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8	12.178
	WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8	0.34520
	GluGluToWWTo*_13TeV_MCFM701_pythia8	0.06387
V $\gamma$ /V $\gamma^*$	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	405.271
	ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	58.83
	WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg_pythia8	58.59
VZ	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.212
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
VVV	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01398
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
	WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.16510
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.18331
Non-Prompt	Data-driven (tight-to-loose method)	

Table A.5: Main 2017 MC simulations for the different background processes considered for this analysis and their respective cross sections.

Process	Sample	Cross section [pb]
DY	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610.0
	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6189.39
TTTo2L2Nu	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	3.360
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	26.38
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	44.33
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60
WW	WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8	12.178
	WWJJToLNuLNu_EWK_TuneCP5_13TeV-madgraph-pythia8	0.4286
	GluGluToWWTo*_TuneCP5_13TeV_MCFM701_pythia8	0.06387
V $\gamma$ /V $\gamma^*$	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	405.271
	ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	131.300
	WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg_pythia8	58.59
VZ	ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.212
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
VVV	ZZZ_TuneCP5_13TeV-amcatnlo_pythia8	0.01398
	WZZ_TuneCP5_13TeV-amcatnlo_pythia8	0.05565
	WWZ_TuneCP5_13TeV-amcatnlo_pythia8	0.16510
	WWW_4F_TuneCP5_13TeV-amcatnlo_pythia8	0.18331
Non-Prompt	Data-driven (tight-to-loose method)	

Table A.6: Main 2018 MC simulations for the different background processes considered for this analysis and their respective cross sections.



## Appendix B

### **Neural network optimization**



# List of figures

2.1	Representation of the 12 fermions of the SM [21]. . . . .	6
2.2	Expected and observed rotation curves of the galaxy NGC 6503 [6]. The black dots correspond to the data and the <i>luminous</i> line corresponds to the expected rotation curve decreasing as $r^{-1/2}$ , as expected from Newtonian dynamics. . . . .	8
2.3	Anisotropies at the $10^{-5}$ level in the temperature of the CMB, as observed by the Planck satellite in 2018 [29]. . . . .	9
2.4	Power spectrum of the CMB obtained by Planck, representing the fluctuations of the temperature of the radiation with respect to the angular angle of observation [31].	10
2.5	Mass distributions obtained by the Magellan in the visible (on the left) and Chandra on the X-rays spectrum (on the right) telescopes of the Bullet Cluster. Being shifted compared to each other, this is yet another clear evidence for the existence of DM [33]. . . . .	11
2.6	Computer simulations for cold (on the left) and warm (on the right) DM scenarios and their impact on a galactic halo at 0 redshift [35]. . . . .	13
2.7	Schematic representation of the freeze-out process, representing the abundance of a 500 GeV DM as $Y_\chi$ with respect to the time and the impact of increasing cross-section annihilation values on this freeze-out abundance [36]. . . . .	13
2.8	Halo fraction upper limit at the 95% CL compared to the mass of the lensing object for different MACHO models considered by the EROS (on the top) and the MACHO (on the bottom) collaborations [40]. . . . .	14
2.9	Neutrino cross section of interaction from the charged current as measured by different experiments over a large range of energies, for both neutrinos $\nu$ and antineutrinos $\bar{\nu}$ [41]. . . . .	15
2.10	3.57 keV emission line detected with a $4.5\sigma$ CL by the XMM-Newton telescope in 2014, which could be a hint of the presence of DM [45]. . . . .	16
2.11	Axions exclusion plot summary and projected coverage of axion searches experiments, such as ADMX, CAST and IAXO [51]. . . . .	18
2.12	Schematic view of the three main DM detection strategies: direct, indirect and collider production searches [53]. . . . .	19
2.13	Nuclear recoil spectra induced in different materials for a given DM WIMP of 100 GeV, assuming a WIMP-nucleon SI cross section [54]. . . . .	20

2.14 Schematic representation of the annual modulation of the WIMP wind introduced by the motion of rotation of the Earth around the Sun [56]. . . . .	20
2.15 Impact of different experimental parameters on the final limits depending on the cross section and WIMP mass (on the left) or sensitivity and exposure (on the right), with respect to the expected limits (black curve) [57]. . . . .	22
2.16 Schematic representation of the three main strategies to detect directly the interaction between DM particles and an ordinary nucleus [57]. . . . .	23
2.17 Exclusion limits obtained by various direct detection experiments considering a SI interaction cross section for low WIMP (on the left) or high WIMP masses (on the right) [57]. . . . .	23
2.18 Observed and expected annual modulation in single hits events in the 2-6 keV energy range by the DAMA experiment [58]. . . . .	23
2.19 Upper limits on the DM annihilation cross section considering $b\bar{b}$ (on the left) and $\mu^+\mu^-$ (on the right) final states as a function of the WIMP mass, for different clusters studied [61]. . . . .	25
2.20 Neutrino spectra for a scalar DM candidate of 1 TeV for different indirect detection experiments and the corresponding background level expected [63]. . . . .	26
2.21 Limits of the decay width of the interaction with respect to the DM mass obtained by both IceCube and AMS [64]. . . . .	26
2.22 Schematic representation of a typical EFT modelization of an LHC event with an ISR object used to trigger the event [53]. . . . .	28
2.23 Schematic representation of a typical collider DM production through s-channel and t-channel processes [66]. . . . .	29
2.24 Schematic representation of a typical mono-X event, with an ISR jet in this case going back-to-back with some MET [53]. . . . .	29
2.25 Observed and expected 95% exclusion limits obtained by different searches of the CMS collaboration as a function the spin-0 scalar (on the left) or pseudoscalar (on the right) mediator. . . . .	31
2.26 Observed and expected 95% exclusion limits obtained by different searches of the CMS collaboration as a function the spin-1 mediator, considering an axial-vector (on the left) and axial (on the right) interaction. . . . .	31
2.27 CMS 90% exclusion limits compared to the most famous direct detection experiments for the SD (on the left) and SI (on the right) scenarios, obtained using similar couplings. . . . .	32
2.28 Feynman diagrams involving the production of DM with a single top quark with its associated t-channel W boson (on the left), or tW (on the center and on the right) production. . . . .	32
2.29 Schematic representation of a typical $t\bar{t}$ +DM event. . . . .	33
2.30 Limits on the mass on the DM and mediator masses obtained by ATLAS using 13.3 $\text{fb}^{-1}$ of 13 TeV data, considering scalar (on the left) and pseudoscalar (on the right) mediators [14]. . . . .	34

2.31	Exclusion limits at the 95% CL obtained by ATLAS considering scalar (on the left) and pseudoscalar (on the right) mediators, for a DM mass of 1 GeV [17] . . . . .	34
2.32	95% CL exclusion plots on the signal strength computed as a function of the mediator and DM masses obtained by CMS considering a scalar (on the left) and a pseudoscalar (on the right) mediator for the interaction [19]. . . . .	34
2.33	Expected and observed 95% CL limits on the DM production cross sections shown considering scalar (on the left) and pseudoscalar (on the right) mediators for the interaction [20]. . . . .	35
3.1	LHC injection chain and experiments performed at CERN [83]. . . . .	38
3.2	Integrated luminosity collected by CMS over its different years of operation so far.	41
3.3	Mean PU distribution and luminosity recorded by CMS over the different years of operation of the LHC. . . . .	42
3.4	Schematic representation of the CMS detector, along with all its subdetectors and main characteristics. . . . .	43
3.5	Schematic representation of the CMS coordinate system used by convention. . . . .	44
3.6	Schematic representation of the CMS tracker, for different pseudorapidity values and along the z-axis [87]. . . . .	45
3.7	Tracker reconstruction efficiency of muons (on the left) and pions (on the right) for different pseudorapidities and particle momenta (1, 10 and 100 GeV) [86]. . . . .	46
3.8	Schematic representation of the sub-systems of the CMS ECAL [86]. . . . .	47
3.9	Schematic representation of a typical electromagnetic shower and the radiation length $X_0$ concept [86]. . . . .	47
3.10	Schematic representation of the HCAL sub-system in CMS [86]. . . . .	49
3.11	Distribution of the measured energy scaled to the incident energy for pions with incident energies of 200 GeV at $\eta = 0$ (on the left) and 225 GeV at $\eta = 0.5$ (on the right), with and without the inclusion of the HO in the HCAL system [86]. . . . .	49
3.12	Picture of the solenoid system of CMS being setup in the assembly hall. . . . .	50
3.13	Geometrical repartition along the z-axis of the different muons chambers in CMS [88].	51
3.14	Lateral geometrical division of the different DT chambers in one of the 5 wheels of CMS [86]. . . . .	52
3.15	Muon reconstruction efficiency with different $p_T$ and $\eta$ values, considering only the muon system (on the left), and the combined information from the muon system and the tracker (on the right) [86]. . . . .	54
3.16	Location of the new GEM muon subsystem currently being installed in the very-forward region of CMS [86]. . . . .	54
3.17	Architecture of Level-1 Trigger (L1) of CMS [86]. . . . .	55
3.18	Architecture of CMS Data Acquisition System (DAQ) system [86]. . . . .	56

4.1	Transverse section of CMS showing the different tracks expected by different kind of particles in the detector. . . . .	59
4.2	Different kind of vertices typically observed in a $pp$ collision in the LHC. . . . .	59
4.3	Muon $p_T$ resolution obtained in simulation in the barrel (on the left) and endcap (on the right) for different kind of reconstructed muons [92]. . . . .	61
4.4	Lepton isolation cone typically used to reduce the background levels from SM processes. . . . .	62
4.5	Schematic representation of the full electron reconstruction workflow in CMS [94].	64
4.6	Schematic representation of the typical development of a jet within the CMS detector.	64
4.7	Comparison of the jet response (on the left) and jet energy resolution (on the right) for dijets simulated events in the barrel for jets reconstructed using only the calorimeters (in blue) and jet candidates from the PF algorithm (in red) [97]. . . . .	65
4.8	Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex. . . . .	66
4.9	B-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [98]. . . . .	67
4.10	Schematic representation of the Missing Transverse Energy (MET). . . . .	67
4.11	MET and jet $\phi$ distributions with and without MET filters applied [99]. . . . .	68
5.1	Structure of a $pp$ collision and different steps of the MC simulation used by the event generators, such as the parton shower (in green), the UE (in pink), the hadronization (in blue) and the decay of unstable particles (in red) [101]. . . . .	72
5.2	Top $p_T$ (on the left) and rapidity (on the right) distributions obtained using different MC generators [110]. . . . .	73
5.3	Proton energy distribution at 3 (on the left) and 6 (on the right) GeV compared for the test beam data (in black) and two different GEANT4 versions [112]. . . . .	75
5.4	Production cross section of the most common SM processes considering different center of mass energies, such as the 13 TeV of the LHC. . . . .	77
5.5	Main feynman diagrams for the production of the SM $t\bar{t}$ process. . . . .	78
5.6	Feynman diagrams for s-channel production mode of a single top quarks. . . . .	78
5.7	Feynman diagrams for t-channel (on the left) and tW (on the right) production modes of a single top quarks. . . . .	79
5.8	Feynman diagrams for the leptonic decay of the top (on the left) and antitop (on the right) quarks. . . . .	79
5.9	Feynman diagram for the DY process involving a virtual $\gamma^*$ or Z boson. . . . .	79
5.10	Possible Feynman diagrams for the Initial State Radiation (ISR) $t\bar{t}$ with a W/Z boson (on the left) and for the production of an Final State Radiation (FSR) ttZ (on the center and right). . . . .	80

5.11	Possible Feynman diagrams for the production of a W/Z boson with a jet . . . . .	80
5.12	Schematic representation of the two jets used for the systematics and for the jet faking a lepton in the tight-to-loose datadriven method. . . . .	81
5.13	Electron (on the left) and muon (on the right) FR obtained in a QCD enriched region for different jet $E_T$ thresholds for 2016, 2017 and 2018 with respect to the $p_T$ of the lepton. . . . .	83
5.14	Electron (on the left) and muon (on the right) PR obtained in a Z+jets enriched region by a general tag and probe method for 2016, 2017 and 2018 with respect to the $p_T$ of the lepton. . . . .	84
5.15	Possible Feynman diagrams for smaller backgrounds of this analysis: WW (on the left), W $\gamma$ and WZ (on the center) and ZZ (on the right). . . . .	85
5.16	Data/MC agreement with (on the right) and without (on the left) prefiring corrections in a 2017 top control region. . . . .	86
6.1	DougleEG trigger efficiencies with respect to the $p_T$ (on the left) and $\eta$ (on the right), computed using a tag and probe method, for the 2017 data taking period. .	89
6.2	Tight electron efficiencies for this analysis, based on the Egamma POG $mva\_90p\_Iso2016$ WP, for the data taking period of 2016. . . . .	92
6.3	Tight muon efficiencies for this analysis, based on the Muon POG tight WP with additional cuts for 2016, 2017 and 2018. . . . .	93
6.4	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the $t/\bar{t}+DM$ process. . . . .	95
6.5	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the $t\bar{t}+DM$ process. . . . .	96
6.6	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the DY control region defined. . . . .	98
6.7	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the top control region defined. . . . .	99
6.8	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the ttV control region defined. . . . .	100
6.9	Same sign control region $m_{ll}$ distributions for the years 2016, 2017 and 2018 and for the $e\mu$ channel and for the 0-jet (on the left) and 1j (on the right) categories. . .	101



# List of tables

3.1	Expected and observed main parameters of $pp$ operation of the LHC across the different eras of operation [82]. . . . .	39
3.2	Comparison of the three main sub-systems currently used by CMS in order to identify muons [89]. . . . .	53
6.1	2016 trigger paths considered for this analysis. . . . .	88
6.2	2017 trigger paths considered for this analysis. . . . .	88
6.3	2018 trigger paths considered for this analysis. . . . .	89
6.4	Quality cuts applied to define a 2016 tight electron in this analysis. . . . .	91
6.5	Quality cuts applied to define a 2017/2018 tight electron in this analysis. . . . .	92
6.6	JET/MET POG tight WP for 2016. . . . .	94
6.7	JET/MET POG tight WP for 2017 and 2018. . . . .	94
6.8	JET/MET POG tight WP for 2017 and 2018, for jets having a $p_T > 50$ GeV. . . . .	94
A.1	Datasets collected in 2016 and considered for this analysis. . . . .	112
A.2	Datasets collected in 2017 and considered for this analysis. . . . .	113
A.3	Datasets collected in 2018 and considered for this analysis. . . . .	114
A.4	Main 2016 MC simulations for the different background processes considered for this analysis and their respective cross sections. . . . .	115
A.5	Main 2017 MC simulations for the different background processes considered for this analysis and their respective cross sections. . . . .	116
A.6	Main 2018 MC simulations for the different background processes considered for this analysis and their respective cross sections. . . . .	117



# Bibliography

- [1] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", Phys. Rev. Lett. 13, pp. 321-323, 1964
- [2] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett. 13, pp. 508-509, 1964
- [3] S. Chatrchyan et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys. Lett. B716, pp. 30-61, 2012 [arXiv: 1207.7235]
- [4] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys. Lett. B716, pp. 1-29, 2012 [arXiv: 1207.7214]
- [5] V.C. Rubin, W.K. Ford and N. Thonnard, "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)", Astrophysical Journal 238, pp. 471-487, 1980
- [6] K.G. Begeman, A.H. Broeils and R.H. Sanders, "Extended rotation curves of spiral galaxies - Dark haloes and modified dynamics", Monthly Notices of the Royal Astronomical Society, vol. 249, issue 3, ISSN 0035-8711, 1991
- [7] A. Robertson, R. Massey and V. Eke, "What does the Bullet Cluster tell us about self-interacting dark matter?", Monthly Notices of the Royal Astronomical Society, vol. 465, issue 1, 2017 [arXiv: 1605.04307]
- [8] J.B. Mu?oz, C. Dvorkin and A. Loeb, "21-cm Fluctuations from Charged Dark Matter", Phys. Rev. Lett. 121, 121301 (2018) [arXiv: 1804.01092]
- [9] A. Natarajan, "A closer look at CMB constraints on WIMP dark matter", Phys. Rev. D85, 2012 [arXiv:1201.3939 ]
- [10] G. D'Ambrosio G.F. Giudice, G. Isidori and A. Strumia, "Minimal Flavour Violation: an effective field theory approach", Nucl.Phys. 645, pp 155-187, 2002 [arXiv:0207.036 ]
- [11] CMS Collaboration, "Search for the production of dark matter in association with top-quark pairs in the single-lepton final state in proton-proton collisions at  $\sqrt{s} = 8$  TeV", JHEP, vol. 6 121, 2015
- [12] CMS Collaboration, "Search for the Production of Dark Matter in Association with Top Quark Pairs in the Di-lepton Final State in pp collisions at  $\sqrt{s} = 8$  TeV", CMS-PAS-B2G-13-004, 2014
- [13] "Search for dark matter in events with heavy quarks and missing transverse momentum in pp collisions with the ATLAS detector", Eur. Phys. J. C (2015) 75:92
- [14] ATLAS Collaboration, Search for the Supersymmetric Partner of the Top Quark in the Jets+Emiss Final State at  $\sqrt{s} = 13$  TeV", ATLAS-CONF-2016-077

- [15] ATLAS Collaboration, "Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in  $\sqrt{s} = 13$  TeV pp collisions with the ATLAS detector", ATLAS-CONF-2016-050, 2016
- [16] ATLAS Collaboration, "Search for direct top squark pair production and dark matter production in final states with two leptons in  $\sqrt{s} = 13$  TeV pp collisions using  $13.3 \text{ fb}^{-1}$  of ATLAS data", ATLAS-CONF-2016-076, 2016
- [17] ATLAS Collaboration, "Search for dark matter produced in association with bottom or top quarks in  $\sqrt{s} = 13$  TeV pp collisions with the ATLAS detector", Eur. Phys. J. C 78 (2018) 18 [arXiv: 1710.11412]
- [18] CMS Collaboration, Search for dark matter produced in association with heavy-flavor quark pairs in proton-proton collisions at  $\sqrt{s} = 13$  TeV", Eur. Phys. J. C (2017) 77: 845
- [19] CMS Collaboration, "Search for dark matter particles produced in association with a top quark pair at  $\sqrt{s} = 13$  TeV", Phys. Rev. Lett. 122, 011803 (2019) [arXiv: 1807.06522]
- [20] CMS Collaboration, "Search for dark matter produced in association with a single top quark or a top quark pair in proton-proton collisions at  $\sqrt{s} = 13$  TeV", JHEP, vol. 03 141, 2019 [arXiv: 1901.01553]
- [21] S. Manzoni, "The Standard Model and the Higgs Boson", Physics with Photons Using the ATLAS Run 2 Data, Springer Theses, 2019
- [22] A.B. Balantekin, A. Gouvea and B.Kayser, "Addressing the Majorana vs. Dirac Question with Neutrino Decays", FERMILAB-PUB-18-418-T, NUHEP-TH/18-09 [arXiv: 1808.10518]
- [23] J. Woithe, G.J. Wiener and F. Van der Vecken, "Let's have a coffee with the Standard Model of particle physics!", Physics education 52, number 3, 2017
- [24] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln", Helvetica Physica Acta , vol. 6, pp. 110-127, 1933
- [25] S. Van den Bergh, Phys Rev D "The early history of dark matter", Dominion Astrophysical Observatory, 1999
- [26] V.C. Rubin, W.K. Ford, "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions", Astrophysical Journal 159, p. 379, 1970
- [27] A. A. Penzias, R.W. Wilson, "A Measurement of Excess Antenna Temperature at 4080 Mc/s", Astrophysical Journal 142, pp. 419-421
- [28] D.J. Fixsen, "The temperature of the cosmic microwave background", Astrophysical Journal, 2009
- [29] Planck Collaboration, "Planck 2018 results. I. Overview and the cosmological legacy of Planck", 2018 [arXiv: 1807.06205]
- [30] R. Tojeiro, "Understanding the Cosmic Microwave Background Temperature Power Spectrum", 2006
- [31] Planck Collaboration, "Planck 2018 results. VI. Cosmological parameters", 2018 [arXiv: 1807.06209]
- [32] "Astrophysical Constants and Parameters", 2019
- [33] D. Clowe et all., "A Direct Empirical Proof of the Existence of Dark Matter", Astrophysical Journal Letters 648, 2006

- [34] K.R. Dienes, J. Fennick, J. Kumar, B. Thomas "Dynamical Dark Matter from Thermal Freeze-Out", Phys. Rev. D 97, 063522 (2018) [arXiv: 1712.09919]
- [35] C.S. Frenk, S.D.M. White, "Dark matter and cosmic structure", Annalen der Physik, p. 22 , 2012 [arXiv: 1210.0544]
- [36] R. Kirk, "Dark matter genesis"
- [37] M. Drewes et all., "A White Paper on keV Sterile Neutrino Dark Matter", 2016 [arXiv: 1602.04816]
- [38] C. Alcock et all., "The MACHO Project: Microlensing Results from 5.7 Years of LMC Observations", Astrophys.J. 542 (2000) 281-307
- [39] P. Tisserand et all., "Limits on the Macho content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds", A & A 469, pp. 387-404 (2007)
- [40] EROS and MACHO collaborations, "EROS and MACHO Combined Limits on Planetary Mass Dark Matter in the Galactic Halo", 1998
- [41] Particle Data Group, "Neutrino Cross Section Measurements", PDG 2019
- [42] K. McFarland, "Neutrino Interactions", 2008 [arXiv: 0804.3899]
- [43] E. Morgan, "Aspects of WIMP Dark Matter Searches at Colliders and Other Probes", Springer theses, 2016
- [44] F. Couchot et all., "Cosmological constraints on the neutrino mass including systematic uncertainties", A & A 606, A104 (2017)
- [45] E. Bulbul et all., "Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters", 2014 [arXiv: 1402.2301]
- [46] A. Boyarsky et all., "An unidentified line in X-ray spectra of the Andromeda galaxy and Perseus galaxy cluster", Phys. Rev. Lett. 113, 251301 (2014) [arXiv: 1402.4119]
- [47] A. Boyarsky et all., "Checking the dark matter origin of 3.53 keV line with the Milky Way center", Phys. Rev. Lett. 115, 161301 (2015) [arXiv: 1408.2503]
- [48] T. Jeltema1 and S.Profumo, "Deep XMM Observations of Draco rule out at the 99% Confidence Level a Dark Matter Decay Origin for the 3.5 keV Line", 2015 [arXiv: 1512.01239]
- [49] D. Wu, "A Brief Introduction to the Strong CP Problem", Superconducting Super Collider Laboratory, 1991
- [50] R.D. Peccei, H.R. Quinn, "CP Conservation in the Presence of Pseudoparticles", Phys. Rev. Lett. 38, 1440, 1977
- [51] P.W. Graham et all., "Experimental Searches for the Axion and Axion-like Particles", Annual Review if Nuclear and Particle Science 65, 2015 [arXiv: 1602.00039]
- [52] CAST collaboration, "New CAST limit on the axion-photon interaction", Nature Physics 13, pp. 584-590 (2017)
- [53] B. Penning, "The Pursuit of Dark Matter at Colliders - An Overview", 2017 [arXiv: 1712.01391]

- [54] M. Schumann, "Direct Detection of WIMP Dark Matter: Concepts and Status", J. Phys. G46 (2019) no.10, 103003 [arXiv: 1903.03026]
- [55] S.C. Martin et all., "The RAVE survey: constraining the local Galactic escape speed", Mon.Not.Roy.Astron.Soc.379:755-772, 2007
- [56] K. Freese, M. Lisanti, C. Savage, "Annual Modulation of Dark Matter: A Review", [arXiv: 1209.3339v3]
- [57] T.M. Undagoitia and L. Rauch, "Dark matter direct-detection experiments", J. Phys. G43 (2016) no.1, 013001 [arXiv: 1509.08767]
- [58] R. Bernabei et all., "First results from DAMA/LIBRA and the combined results with DAMA/NaI", Eur.Phys.J.C56:333-355, 2008 [arXiv: 0804.2741]
- [59] J.M. Gaskins, "A review of indirect searches for particle dark matter", Contemporary Physics, 2016 [arXiv: 1604.00014]
- [60] F.S. Queiroz, "Dark Matter Overview: Collider, Direct and Indirect Detection Searches", Max-Planck Institute of Physics
- [61] LAT collaboration, "Constraints on Dark Matter Annihilation in Clusters of Galaxies with the Fermi Large Area Telescope", JCAP 05(2010)025 [arXiv: 1002.2239]
- [62] A.A. Moiseev et all., "Dark Matter Search Perspectives with GAMMA-400", 2013 [arXiv: 1307.2345]
- [63] L. Covi et all., "Neutrino Signals from Dark Matter Decay", JCAP 1004:017, 2010 [arXiv: 0912.3521]
- [64] B. Lu and H. Zong, "Limits on the Dark Matter from AMS-02 antiproton and positron fraction data", Phys. Rev. D 93, 103517 (2016) [arXiv: 1510.04032]
- [65] J. Abdallah et all., "Simplified Models for Dark Matter Searches at the LHC", Phys. Dark Univ. 9-10 (2015) 8-23 [arXiv: 1506.03116]
- [66] H. An, L. Wang, H. Zhang, "Dark matter with t-channel mediator: a simple step beyond contact interaction", Phys. Rev. D 89, 115014 (2014) [arXiv: 1308.0592]
- [67] ATLAS Collaboration, "Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector", JHEP 01 (2018) 126 [arXiv: 1711.03301]
- [68] CMS Collaboration, "Search for new physics in the monophoton final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV", J. High Energy Phys. 10 (2017) 073 [arXiv: 1706.03794]
- [69] CMS Collaboration, "Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at  $\sqrt{s} = 13$  TeV", JHEP 07 (2017) 014 [arXiv: 1703.01651]
- [70] CMS Collaboration, "Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at  $\sqrt{s} = 13$  TeV", Phys. Rev. D 97, 092005 (2018) [arXiv: 1712.02345]
- [71] ATLAS Collaboration, "Search for dark matter in association with a Higgs boson decaying to two photons at  $\sqrt{s} = 13$  TeV with the ATLAS detector", Phys. Rev. D 96 (2017) 112004 [arXiv: 1706.03948]
- [72] CMS Collaboration, "Search for associated production of dark matter with a Higgs boson decaying to  $b\bar{b}$  or  $\gamma\gamma$  at  $\sqrt{s} = 13$  TeV", JHEP 10 (2017) 180 [arXiv: 1703.05236]

- [73] Atlas Collaboration, "Search for new phenomena in dijet events using  $37 \text{ fb}^{-1}$  of pp collision data collected at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector", Phys. Rev. D 96, 052004 (2017) [arXiv: 1703.09127]
- [74] CMS Collaboration, "Search for narrow and broad dijet resonances in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  and constraints on dark matter mediators and other new particles", JHEP 08 (2018) 130 [arXiv: 1806.00843]
- [75] C. Munoz, "Models of Supersymmetry for Dark Matter", FTUAM 17/2, IFT-UAM/CSIC-17-005, 2017 [arXiv: 1701.05259]
- [76] CMS Collaboration, "Searches for invisible decays of the Higgs boson in pp collisions at  $s_{\text{sqrt}s} = 7, 8, \text{ and } 13 \text{ TeV}$ ", JHEP 02 (2017) 135 [arXiv: 1610.09218]
- [77] J. Alimena et all., "Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider", 2019 [arXiv: 1903.04497]
- [78] A. Albert et all., "Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for heavy mediators of dark matter production in visible and invisible decay channels", 2017 [arXiv: 1703.05703]
- [79] M. Tanabashi et al., Particle Data Group, Phys. Rev. D98, 030001 (2018)
- [80] R. Schicker, "The ALICE detector at LHC", 2005
- [81] LHCb Collaboration, "LHCb Detector Performance", Int. J. Mod. Phys. A 30, 1530022 (2015) [arXiv: 1412.6352]
- [82] J.T. Boyd, "LHC Run-2 and Future Prospects", 2020
- [83] E. Gschwendtner, "AWAKE, A Particle-driven Plasma Wakefield Acceleration Experiment", CERN Yellow Report CERN 2016-001, pp.271-288 [arXiv: 1705.10573]
- [84] M. Thomson, "Modern Particle Physics", Cambridge University Press, 2013
- [85] G. Apollinari et all., "High Luminosity Large Hadron Collider HL-LHC", CERN Yellow Report CERN-2015-005, pp.1-19 [arXiv: 1705.08830]
- [86] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004
- [87] CMS Collaboration, "Precision measurement of the structure of the CMS inner tracking system using nuclear interactions", JINST 13 (2018) P10034 [arXiv: 1807.03289]
- [88] M.S. Kim, "CMS reconstruction improvement for the muon tracking by the RPC chambers", 2013 JINST 8 T03001 [arXiv: 1209.2646]
- [89] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ ", JINST 13 (2018) P06015 [arXiv: 1804.04528]
- [90] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS-PAS-PFT-09-001, 2009
- [91] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", JINST 9 (2014) P10009 [arXiv: 1405.6569]
- [92] V. Knunz, "Measurement of Quarkonium Polarization to Probe QCD at the LHC", Springer theses, 2015

- [93] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV", JINST 10 (2015) P06005 [arXiv: 1502.02701]
- [94] J. Rembser, "CMS Electron and Photon Performance at 13 TeV", J. Phys. Conf. Ser. 1162 012008, 2019
- [95] P.L.S. Connor, "Review of jet reconstruction algorithms", Ryan Atkin J. Phys. Conf. Ser. 645 012008, 2015
- [96] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", JINST 12 (2017) P02014 [arXiv: 1607.03663]
- [97] F. Beaudette, "The CMS Particle Flow Algorithm", 2014 [arXiv: 1401.8155]
- [98] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", JINST 13 (2018) P05011 [arXiv: 1712.07158]
- [99] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector", JINST 14 (2019) P07004 [arXiv: 1903.06078]
- [100] L. Sonnenschein, "Analytical solution of ttbar dilepton equations", Phys.Rev.D73:054015, 2016
- [101] M.H. Seymour and M. Marx, "Monte Carlo Event Generators", MCnet-13-05, 2013 [arXiv:1304.6677]
- [102] B. Cabouat, J.R. Gaunt and K. Ostrolenk, "A Monte-Carlo Simulation of Double Parton Scattering", JHEP11(2019)061 [arXiv: 1906.04669]
- [103] R. Placakyte, "Parton Distribution Functions", 2011 [arXiv:1111.5452]
- [104] J. Alwall et all., "MadGraph 5 : Going Beyond", 2011 [arXiv:1106.0522]
- [105] C. Oleari, "The POWHEG-BOX", Nucl.Phys.Proc.Suppl.205-206:36-41 [arXiv:1007.3893]
- [106] S. Frixione et all., "The MC@NLO 4.0 Event Generator", CERN-TH/2010-216 [arXiv: 1010.0819]
- [107] B. Webber, "Parton shower Monte Carlo event generators", Scholarpedia
- [108] M. Bahr et all., "Herwig++ Physics and Manual", Eur.Phys.J.C58:639-707, 2008 [arXiv: 0803.0883]
- [109] T. Sjostrand, "A Brief Introduction to PYTHIA 8.1" Comput.Phys.Commun.178:852-867, 2008 [arXiv: 0710.3820]
- [110] A. Karneyeu et all., "MCPLLOTS: a particle physics resource based on volunteer computing", European Physical Journal C 74 (2014) [arXiv: 1306.3436]
- [111] V. Lefebure and S. Banerjee, "CMS Simulation Software Using Geant4", CMS-NOTE-1999-072, 1999
- [112] S. Banerjee, "Validation of Geant4 Physics Models Using Collision Data from the LHC", J. Phys.: Conf. Ser. 898 042005
- [113] A. Rizzi, G. Petrucciani and M. Peruzzi, "A further reduction in CMS event data for analysis: the NANOAOD format", J. Phys.: Conf. Ser. 214 06021

- [114] C.G. Lester and D.J. Summers, "Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders", Phys.Lett.B463:99-103, 1999
- [115] K. Bloom, "CMS software and computing for LHC Run 2", ICHEP 2016 [arXiv: 1611.03215]
- [116] W. Tanenbaum, "A ROOT/IO Based Software Framework for CMS", ECONFC0303241:TUKT010, 2003
- [117] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS-PAS-LUM-17-001, 2017
- [118] CMS Collaboration, "CMS Luminosity Measurements for the 2017 Data Taking Period", CMS-PAS-LUM-17-001, 2018
- [119] CMS Collaboration, "CMS Luminosity Measurements for the 2018 Data Taking Period", CMS-PAS-LUM-17-001, 2019
- [120] CMS Twiki, "Reweighting recipe to emulate Level 1 ECAL prefireing"
- [121] CMS Collaboration, "Measurement of the differential cross sections for top quark pair production as a function of kinematic event variables in pp collisions at  $\sqrt{s} = 7$  and 8 TeV", Phys. Rev. D 94 (2016) 052006 [arXiv: 1607.00837]
- [122] CMS Twiki, "Electron and Photon Physics Object Offline Guide"
- [123] CMS Twiki, "Baseline muon selections for Run-II", 2019
- [124] CMS Twiki, "Jet Identification"
- [125] CMS Twiki, "Heavy flavor identification at CMS with deep neural networks"
- [126] CMS Collaboration, "BRIL Work Suite"