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 Ruíz del Árbol

Santander, February 2022

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 Ruíz del Árbol

Santander, Febrero 2022

Abstract

A search for dark matter particles produced in association with one or two top quarks in the dilepton final state using the analytic top reconstruction technique and Boosted Decision Trees is presented in this work. This analysis has been done by considering the full (137.1 ± 2.0) fb^{-1} of proton-proton collisions data collected by the **CMS!** (**CMS!**) detector during the Run II of operation of the **LHC!** (**LHC!**), at a center of mass energy of $\sqrt{s} = 13$ TeV. This is the first time that dark matter is searched for combining its production with a top (or antitop) quark, and with a top antitop pair in such a final state. No evidence for the existence of dark matter has been found, but upper limits on the signal strength have been obtained by considering several different production models. An expected (observed) exclusion for scalar and pseudoscalar mediators up to 215 (180) and 250 (220) GeV was respectively achieved, at the 95% confidence level. Scalar exclusion limits were therefore improved by a factor of 3 with respect to the previous results obtained in 2016 for the $t\bar{t}$+DM model alone, while a pseudoscalar exclusion has been achieved for the first time when considering this particular final state.

En este trabajo se presenta una búsqueda de partículas de materia oscura producidas en asociación con uno o dos quarks top en el estado final dileptónico utilizando técnicas de reconstrucción analítica del sistema $t\bar{t}$ y Boosted Decision Trees. Este análisis se ha realizado considerando los (137.1 ± 2.0) fb^{-1} de datos de colisiones protón-protón recopilados por el detector **CMS!** durante la segunda fase de operación del **LHC!**, en un centro de energía de masa de 13 TeV. Esta es la primera vez que se busca materia oscura para combinar su producción con un quark top (o antitop), y con un par top antitop en tal estado final. No se ha encontrado evidencia de la existencia de materia oscura, pero límites superiores sobre la sección eficaz de producción de la señal se han obtenido considerando diferentes modelos de señales. Se logró obtener una exclusión esperada (observada) para mediadores escalares y pseudoescalares hasta 215 (180) y 250 (220) GeV, respectivamente, con un nivel de confianza del 95%. Por lo tanto, los límites de exclusión escalar se mejoraron en un factor de 3 con respecto a los resultados anteriores obtenidos en 2016 solo para el modelo ttbar+DM, mientras que se logró obtener por primera vez una exclusión pseudoescalar al considerar este estado final.

Acknowledgments

I would like to thank the group of high energy physics of the **IFCA!** (**IFCA!**) in general for the opportunity they gave me by offering me such an interesting PhD position within the group. This was a great opportunity for me to start working as a CERN collaborator, which had been a dream of mine for many years. All these years spent at **IFCA!** were for sure extremely interesting, brought me a lot in both the personal and professional points of view, and something I will definitely benefit from in the future.

I would also like to personally thank Jónatan Piedra for supervising this thesis, but also for all your help and patience during all these years, especially at the beginning. You taught me a lot and I cannot count the number of bugs and issues you helped me solve, so I am really thankful that you kept your door open at all times. Pablo Martínez, I am grateful for the great work you did as the co-director of this thesis, as I am sure this work would not have been possible without your help and dedication to this analysis. I will never be able to thank you enough for your help, especially in this last year which has undoubtedly been quite challenging for me.

I would also like to thank all the people who helped me at some point: Alicia, Rocío, Pupi, Chus, Luca and Celso. This work would not have been possible either without the help of all the people involved in this particular analysis, such as Deborah and Victor, both doing an amazing job on the hadronic and semileptonic final states, while Nicole, Alexander, Danyer and Dominic were a great help at all times in the dilepton final state. Thank you as well to all the people involved in the development of the Latino framework, which quickly became a great tool to work with. I am also deeply grateful to all the students and seniors from the University of Oviedo for all these interesting meetings, making the Wednesday mornings a highlight of each week. Thank you also to my past and present PhD colleagues who fought alongside me: Pedro, Nicolò, Celia, Clara, Andrea, Pablo, Barbara and Juan, you made the students office a better place to work in and I wish you all the best, even though I am sure you will all achieve great things in the future.

Of course, I would especially like to thank my parents Philippe and Chantal, and my brothers Antoine and Lucas for giving me the opportunity to come and live in Spain for so long. I never could have reached this far without your presence and support that I felt daily, even more than a thousand kilometers away. Thank you to Inés as well, as I would never have started this thesis if I had not met you. Finally, I would like to especially thank Aline for her patience and incredible support over the past year, which helped me a lot and for which I am extremely grateful. Even though one chapter now ends, I cannot wait to get started on the next one by your side, as I am sure many great things are waiting for us in the future.

Acronyms used

ADMX Axion Dark Matter Experiment

ALICE A Large Ion Collider Experiment

AMS Alpha Magnetic Spectrometer

ANN Analysis Neural Network

AOD Analysis Object Data

ATLAS A Toroidal LHC ApparatuS

BDT Boosted Decision Tree

BR Branching Ratio

BSM Beyond the Standard Model

BW Breit-Wigner

CAST CERN Axion Solar Telescope

CERN European Council for Nuclear Research

CL Confidence Level

CMB Cosmic Microwave Background

CMS Compact Muon Solenoid

CSC Cathode Strip Chamber

CR Control Region

CSV Combined Secondary Vertex

CTA Cherenkov Telescope Array

DAQ Data AcQuisition system

DAS Data Aggregation System

DCS Detector Control System

DQM Data Quality Monitoring

DM Dark Matter

DMWG Dark Matter Working Group

DT Drift tube

DY Drell-Yan

ECAL Electromagnetic Calorimeter

EDM Event Data Model

EFT Effective Field Theory

EWK Electroweak

FR Fake Rate

FSR Final State Radiation

GEM Gas Electron Multiplier

GSF Gaussian Sum Filter

HCAL Hadronic Calorimeter

HLT High-Level Trigger

HO Hadron Outer

IACT Imaging Atmospheric Cherenkov Telescopes

IAXO International AXion Observatory

IFCA Instituto de Física de Cantabria

IP Interaction Point

ISR Initial State Radiation

JEC Jet Energy Correction

JER Jet Energy Resolution

JES Jet Energy Scale

KF Kalman Filter

L1 Level-1 Trigger

LAT Fermi Large Telescope

LEP Large Electron Positron collider

LHC Large Hadron Collider

LNGS Laboratori Nazionali del Gran Sasso

LO Leading Order

LS Long Shutdown

LSP Lightest Supersymmetric Particle

MACHO Massive Compact Halo Object

MC Monte Carlo

MET Missing Transverse Momentum

MFV Minimal Flavour Violation

ML Machine Learning

MPI Multiple Parton Interaction

MSSM Minimal Supersymmetric Standard Model

MVA Multi-Variate Analysis

NFW Navarro-Frenk-White

NLO Next to Leading Order

PDF Parton Density Function

PF Particle Flow

POG Physics Object Group

PR Prompt Rate

PS Proton Synchrotron

PU Pile Up

PUPPI Pileup Per Particle Identification

PV Primary Vertex

QCD Quantum ChromoDynamics

QFT Quantum Field Theory

RMS Root Mean Square

ROC Receiver Operating Characteristic

RPC Resistive Plate Chamber

SC Super Cluster

SD Spin Dependent

SF Scale Factors

SI Spin Independent

SM Standard Model

SPS Super Proton Synchrotron

SR Signal Region

TEC Tracker EndCap

TIB/TBD Tracker Inner Barrel and Disks

TOB Tracker Outer Barrel

TPG Trigger Primitive Generators

UE Underlying Event

UED Universal Extra Dimensions

VBF Vector Boson Fusion

WIMP Weakly Interactive Massive Particle

WP Working Point

Table of contents

CHAPTER 1

Introduction

The **SM!** (**SM!**) of particle physics [**?**] is nowadays the most accepted mathematical model used to describe the elementary particles and 3 of the 4 fundamental forces of nature (electromagnetic, weak and strong interactions, while the gravitational interaction is out of reach of this model). Even though quite simple in concept, it has been able to describe most of the phenomena observed in nature so far with an incredible level of precision, and has made a lot of predictions that have now been proven to be true, such as the postulate of the Brout-Englert-Higgs mechanism [**?**, **?**] followed by the discovery of the Higgs boson itself [**?**, **?**] by the **CMS!** [**?**] and **ATLAS!** (**ATLAS!**) [**?**] collaborations, analyzing at the time the proton-proton collisions produced by the **LHC!** (**LHC!**) at a center of mass energy $\sqrt{s} = 7$ and 8 TeV, announced at the **CERN!** (**CERN!**) on the 4th of July 2012.

However, as accurate as it seems to be, this theory, introduced in Chapter **??**, is known to have several shortcomings which require further investigation, still on-going today. Eventual exotic particles which do not fit in the current model could in this sense 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!** (**DM!**) hypothesis was introduced in the 1930s because of gravitational anomalies observed by several astrophysicists, as a way to explain the apparent non-luminous missing mass in the Universe [**?**]. 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 [**?**], 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 [**?**]) and the anisotropies observed in the **CMB!** (**CMB!**) [**?**] are other evidences for the existence of **DM!**. 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 total energy density of the Universe (the rest is being considered as dark energy) [**?**]. Trying to understand the nature and fundamental properties of this new kind of exotic matter is therefore crucial to try and understand the laws of physics in the Universe, with many theorists and large experiments around the world currently involved in such searches.

Nowadays, the existence of **DM!** is well motivated 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!** candidates are the so-called **WIMP!**s (**WIMP!**s), predicted to interact both gravitationally and weakly with **SM!** particles. This would allow direct and indirect direction of such candidates, used as the driving process of many experiments over the last decades, trying to find hints for a possible interaction between standard baryonic particles and **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 **BSM!** (**BSM!**) physics.

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 section 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 **DM!** particles, we would not expect them to interact with our detector, making their actual detection even harder. The presence of such matter has to be inferred from the study of the interaction between **SM!** particles and the **CMS!** detector itself, since a typical **DM!**-like event consists of at least one energetic **SM!** particle produced in association with a large imbalance in the momentum due to the presence of an eventual **DM!** candidate that was able to escape our detection. This kind of collider searches considered in this work can be grouped under the name of *mono-X* searches, where the X stands for any kind of **SM!** particle (a jet, a lepton or a photon for example).

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 actually trigger the event. The top quark, the most most massive of all the fundamental particles observed by far, is indeed an excellent object to study in this context, mainly because of its high mass and because of the expected Yukawa-like coupling structure of the new physics model [**?**]. 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, almost always 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 mostly categorized depending on the decay of the W itself. This work will actually be focused on the two leptons final state, also known as the dilepton channel, mostly because leptons are by reconstruction much cleaner than jets and because this channel does not have lots of background processes raising to a similar final state, even though its branching ratio is the smallest.

The **LHC!** has now been running for a decade, 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. At 13 TeV, the **ATLAS!** collaboration published on one hand several studies, considering different final states and different luminosities, the most recent and relevant one during ICHEP 2020 [**?**]. On the other hand, the **CMS!** collaboration published a few extremely important papers for this study as well, such as a study of the $t\bar{t}$+DM production mode using 2016 data, in the dilepton final state along and then combining the three possible final states (hadronic, semi and dileptonic [**?**]). For the first time in 2019, the results obtained by both the $t/\bar{t}$+DM and $t\bar{t}$+DM analyses have also been combined and published [**?**]. Our main objective is now to repeat and improve this analysis but considering this time the full Run II dataset, globally improving the analysis strategy and including the dilepton final state for the first time in this combination.

In particular, after a general introduction about dark matter in Chapter **??**, the experimental setup used will be detailed in Chapter **??**. This will include a discussion about the **LHC!**, along with a complete description of **CMS!**, the detector used to collect the data that will be analyzed throughout this work. The data analyzed here has been collected during the years 2016, 2017 and 2018 and corresponds to an integrated luminosity of (137.1 ± 0.2) fb^{-1}, collected during the Run II of operation of the **LHC!** and at a center of mass energy $\sqrt{s} = 13$ TeV. A particular care will be given to the explanation of the **PF!** (**PF!**) algorithm, used to reconstruct the different objects of the analysis and that will be defined in Chapter **??**, while the different samples and objects used throughout this analysis will be defined in Chapter **??**. The selection applied to select only interesting events along with a description of the different ways used in order to estimate the impact that the different background processes have in this analysis will be done in Chapter **??**. Furthermore, distinguishing between the signals we are studying 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 momentum due to the presence of physical neutrinos is even obtained). To enhance the signals and to obtain some discrimination between these kind of processes, an algebraic reconstruction of the event and top-notch **ML!** (**ML!**) techniques are used in this work, in order to train Boosted Decision Trees 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 Chapter **??**.

Finally, a statistical interpretation of our data will be performed and different sources of systematic uncertainties will be considered in Chapter **??**. This allows us to set upper limits on the signal strength of **DM!** particles produced in our particular channel and for the simplified models considered in this analysis. The conclusions of this work and some additional prospects for the future will then finally be presented in Chapter **??**.

CHAPTER 2

The dark matter case

In this chapter, some general explanations about the **SM!** will first of all be given in Section **??** as a general introduction about today's most accepted mathematical model describing all the known particles and their interactions. The case for **DM!** will then be presented in Section **??**, along with a summary of the main evidences, mostly astrophysical, which lead to the introduction of this kind of **BSM!** (**BSM!**) physics. Then, the main properties expected by such exotic matter will be introduced in Section **??** and nowadays most accepted **DM!** candidates, such as the **WIMP!** (**WIMP!**) will be presented in Section **??**. 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 **??**. Finally, the searches performed in colliders such as the **LHC!** and our particular channels of interest (**DM!** produced in association with either one or two top quarks) will be detailed in Section **??**, the particular channel of interest of this thesis will be first introduced in Section **??** and the latest similar results and exclusion limits published over the course of the previous years by the **ATLAS!** and **CMS!** collaborations will be summarized in Section **??**.

2.1 The SM! (SM!)

The **SM!** of particle physics is a relativistic **QFT!** (**QFT!**) able to mathematically summarize our current understanding of all the known particles and to describe 3 out of the 4 main interactions between these particles (the weak, strong and electromagnetic forces, while an explanation of the origin of the gravitational interaction is out of the reach of this theory).

If the neutrinos are considered to be normal Dirac fermions in the sense that antineutrinos are expected to be different than neutrinos (the question whether or not they could actually be Majorana particles is still well under discussion [?]), then the **SM!** contains 26 free parameters, mainly the masses of the 12 predicted fermions shown in Figure **??**, along with the couplings describing the strengths of the 3 interactions, 2 parameters describing the Higgs potential, 8 mixing angles and, maybe, the phase of the eventual strong CP violation that will be mentioned in Section **??** (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.

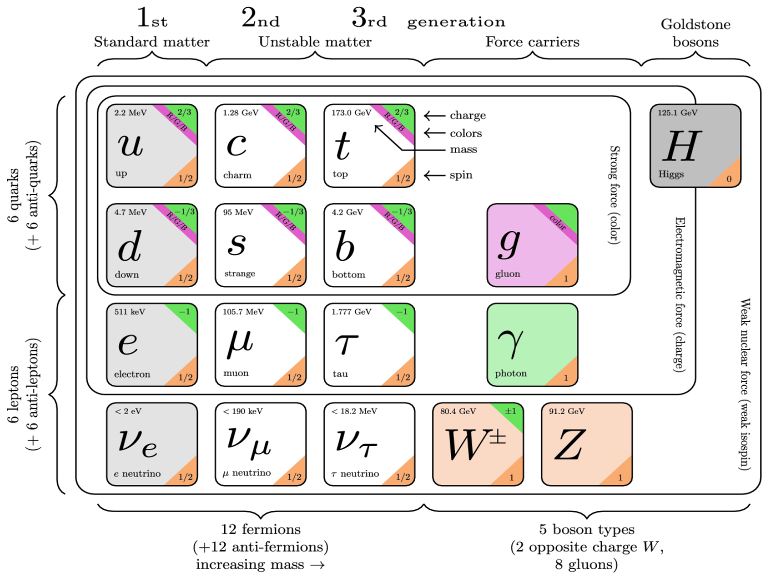


Figure 2.1: Representation of the 12 fermions of the **SM!** [?] along with the main force carriers and the Higgs boson, discovered in 2012 and completing the **SM!**.

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 in Equation **??**, where $F_{\mu\nu}$ is the field strength tensor accounting for the different interactions, ψ is the interacting field describing quarks and leptons and ϕ 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 also be described later on in this work.

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

A complete description of this model is 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 successfully describe the outcome of all the experimental data and to make predictions on new phenomena, so far always confirmed (we can quote as example that it successfully predicted the existence of the gluons, the top quarks, along with the W, Z and Higgs bosons [?]).

However, we know that the **SM!** is not a complete theory of Particle Physics as it fails to explain some known phenomena in Nature. There are in this sense some open questions and many **BSM!** theories trying to explain such observations, such as an eventual inclusion of the gravitational interaction within this model. We can quote for example as **BSM!** theories the possible existence of the supersymmetry [?], telling us that each particle should have a super partner whose spin differs by 1/2 or the possible existence of **DM!** particles, the main subject 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 19th century, even though this concept has 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 did have a strong gravitation impact because of its mass. In the 20th century, astronomical and cosmological observations established the existence of such a species of matter embedded on the core of giant astronomical objects such as galaxies and clusters [**?**].

2.2.1 Zwicky and the virial theorem

In the 20th century, the first experimental evidences for the existence of dark matter were found. In 1933, Fritz Zwicky managed to determine the mass of the Coma Cluster using the virial theorem [**?**], 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 **??**, where the brackets represent the mean value of the quantity obtained over time or position, the universal constant of gravitation is $G = 6.67{\cdot}10^{-11}$ m^3 kg^{-1} 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 for the calculation.

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

Solving these simple equations gives us an approximate value of the mass of the cluster in Equation **??**, 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.

$$M \propto \frac{\langle\langle v^2\rangle\rangle R}{G} \tag{2.3}$$

Zwicky studied the velocity dispersion of the galaxies in the Coma Cluster and used this formula to estimate the total mass of the cluster, concluding that this value 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 [**?**], but additional observations came to enforce the 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 [?].

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 **??**, where $M(r)$ accounts for the total mass encountered in a radius r.

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

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 suggests 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 [?]. 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 **??**. This is another hint that can motivate the introduction of the concept of dark matter.

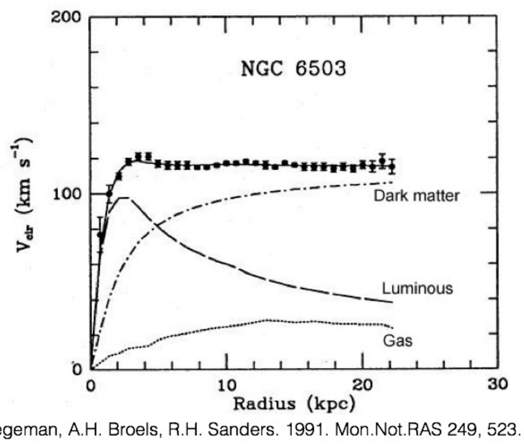


Figure 2.2: Expected and observed rotation curves of the galaxy NGC 6503 [?]. The black dots correspond to the data and the *luminous* line corresponds to the rotation curve decreasing as $r^{-1/2}$ expected from Newtonian dynamics.

2.2.3 CMB! (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 [?]. 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 ± 0.00057)K [**?**]. However, we now know that this temperature is actually not constant as some small anisotropies can be observed (at the 10^{-5} level), depending on the value of the solid angle of observation, as represented in Figure **??**.

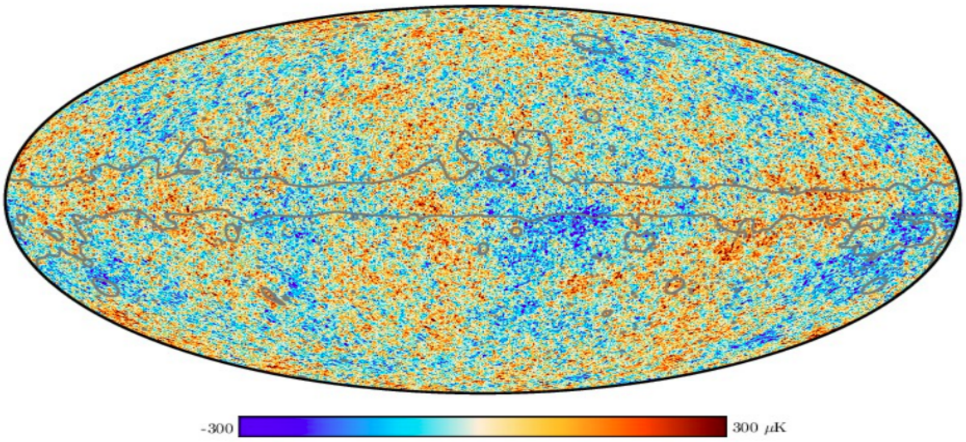


Figure 2.3: Anisotropies at the 10^{-5} level in the temperature of the **CMB!**, as observed by the Planck satellite in 2018 [**?**].

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 solid angle in the sky ($l = 100$ corresponds to $\sim 1°$) and m, the number of poles, such as $-l \leq m \leq l$ [**?**].

It is possible to show that these spherical harmonics form a complete orthonormal basis on this space and therefore that any function that 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 θ and ϕ can therefore be expanded using these generic functions, according to Equation **??**.

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

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 be extracted according to Equation **??**, and from which most of the cosmological information of the **CMB!** is derived.

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

Interestingly enough, this spectrum is directly affected by the value of the density of the dark matter in the Universe. Indeed, if we remove the effect of the $l = 0$ and $l = 1$ poles, which do not have anything to do with the origin of the Universe, the $l = 2$ poles can be related to the fluctuations happening before the recombination phase. Such fluctuations depend on the mass but also on the radiative pressure at this epoch, and should be null for dark matter but not for baryonic matter. Doing a multi-parameters fit on the observed data represented in the power spectrum (cf. Figure **??**) is then able to give us directly the energy density of baryonic Ω_b and dark Ω_χ matter, along with other important parameters of the Λ_{CDM} cosmological model. Today's most precise measurements have been obtained in 2018 using the Planck satellite, and lead to the determination of these two quantities: $\Omega_b h^2 = (0.02220 \pm 0.00020)$ and $\Omega_\chi h^2 = (0.1185 \pm 0.0015)$ [**?**]. This is one of the strongest constraint we have on dark matter so far, since any candidate will need to comply with this measurement.

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

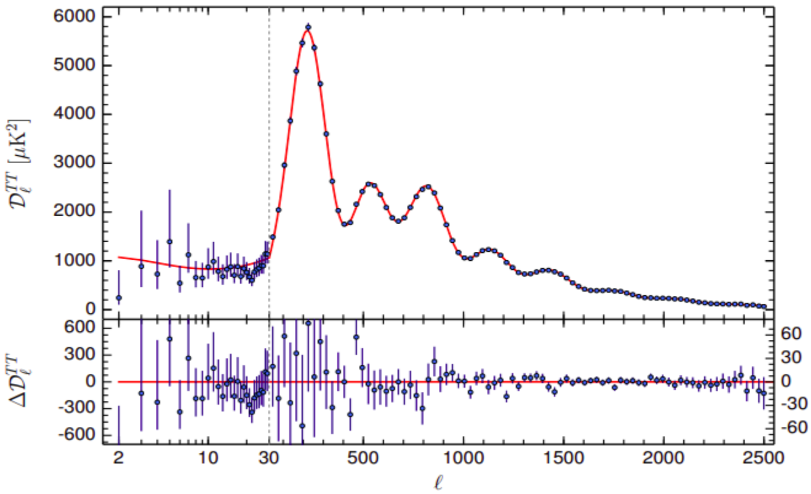


Figure 2.4: Power spectrum of the **CMB!** obtained by Planck, representing the fluctuations of the temperature of the radiation with respect to the angular angle of observation [**?**].

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 space-time, 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 dark matter.

The Bullet Cluster is particularly interesting in this context since it actually provides an evidence for the eventual existence of such particles 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 a modification of the gravitational force law alone, with a statistical significance of 8σ [?].

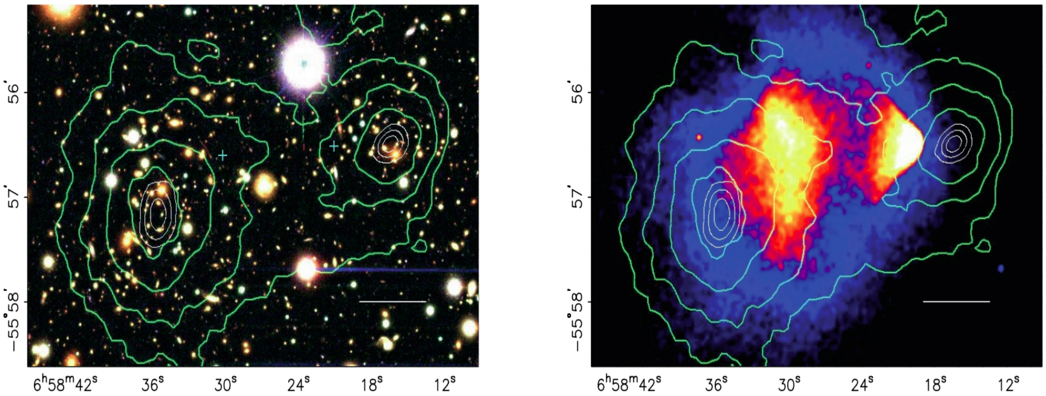


Figure 2.5: Mass distributions obtained by the Magellan telescope in the visible (on the left) and the Chandra telescope on the X-rays spectrum (on the right) telescopes of the Bullet Cluster. Being shifted with respect to each other, this is another evidence for the existence of **DM!** [?].

As seen in Figure **??**, 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 dark matter within galaxy clusters.

2.3 Dark matter properties

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

- 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 **DM!** candidate should of course be **dark**. This means that it should not interact at all with electromagnetic radiation such as light, and that it should consequently be **electrically neutral**. However, it has to interact at least gravitationally because of the observations for the evidence of such a particle explained before, mostly relying on gravitational effects. We also have to assume in this work that it interacts weakly as well in order to have a chance to discover it with particle accelerators.
- It is **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. Indeed, according to these results, baryonic matter accounts 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 Λ_{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 be explained if **DM!** is made of hot/relativistic particles, as represented in Figure **??**. 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 history, dark matter was supposed to be in thermal equilibrium with other primordial **SM!** particles, meaning that its production and annihilation rates were equal. 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 **??**.

  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 coupling g, the mass of the **DM!** particle m_χ and its relic abundance Ω_χ that has been precisely measured, as expressed in Equation **??** [**?**].

$$\Omega_\chi \propto \frac{m_\chi^2}{g^4} \tag{2.7}$$

  By using a typical value for g of the order of the Fermi coupling constant $G_0^F \simeq 4.54 \cdot 10^{14}$ 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 **CMB!** power spectrum is therefore able to put constraints on the **DM!** cross-section with the baryonic sector, and all the possible dark matter candidates 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 the relic abundance of **DM!** would simply violate the Pauli-exclusion principle [**?**].

- 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.

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 investigated, before introducing some **BSM!** theories providing additional possible **DM!** candidates.

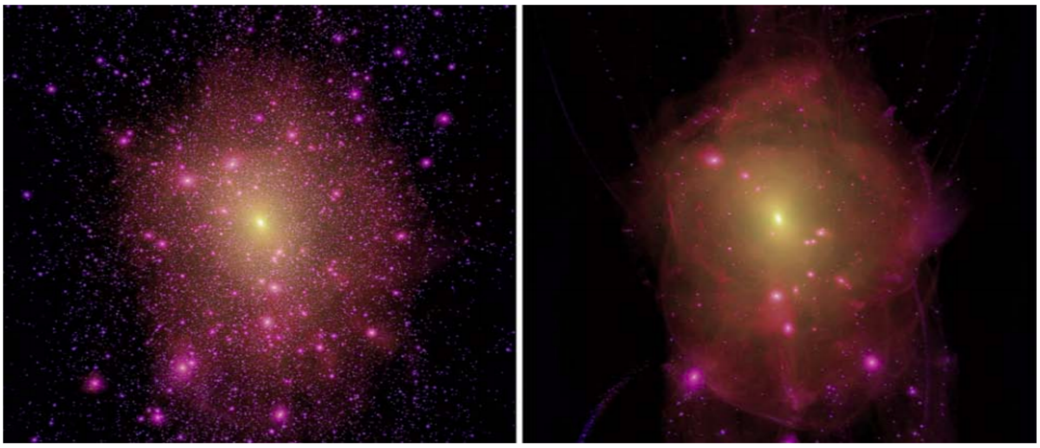


Figure 2.6: Computer simulations for cold (on the left) and warm (on the right) **DM!** scenarios and their impact on a galactic halo at 0 redshift [**?**].

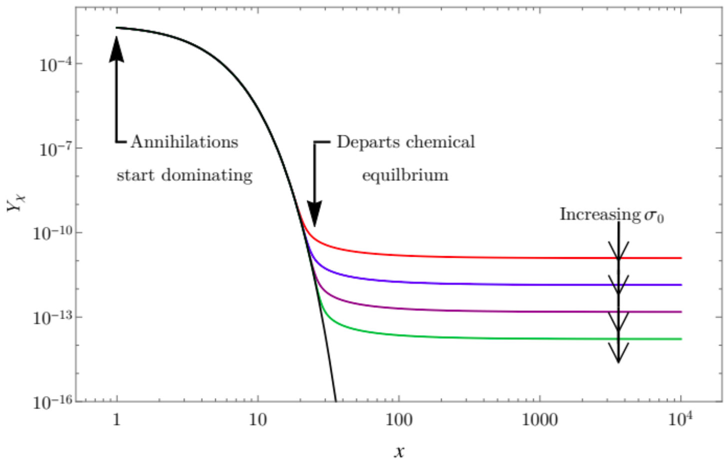


Figure 2.7: Schematic representation of the freeze-out process, representing the abundance of a 500 GeV **DM!** as Y_χ with respect to the time and the impact of increasing cross-section annihilation values on this freeze-out abundance [**?**].

MACHO!s (MACHO!s)

The first obvious **DM!** candidates are the so-called **MACHO!**s. 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 just saw in Section **??**, **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 micro-lensing effect 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 micro lensing events in order to constrain this particular dark matter 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 **MACHO!**s.

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

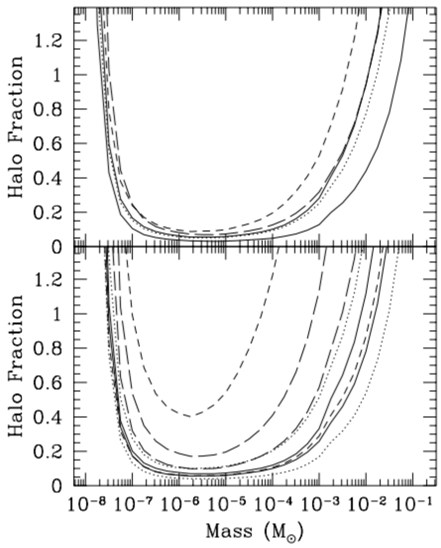


Figure 2.8: Halo fraction upper limits at the 95% **CL!** compared to the mass of the lensing object for different **MACHO!** models considered by the EROS (on the top) and the MACHO (on the bottom) collaborations [**?**].

Active neutrinos

SM! *active* neutrinos ν (as opposed to *sterile* neutrinos, which 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 masses has recently been measured to be lower than 0.17 eV at the 95% **CL!** from cosmological studies [**?**]. Even though this is not fully understood, such value 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 **??** 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.

  This figure clearly shows that an approximate value of the cross section for such neutrinos is of the order of magnitude of 10^{-38} cm^2 GeV^{-1}, which is typically tens of orders of magnitude

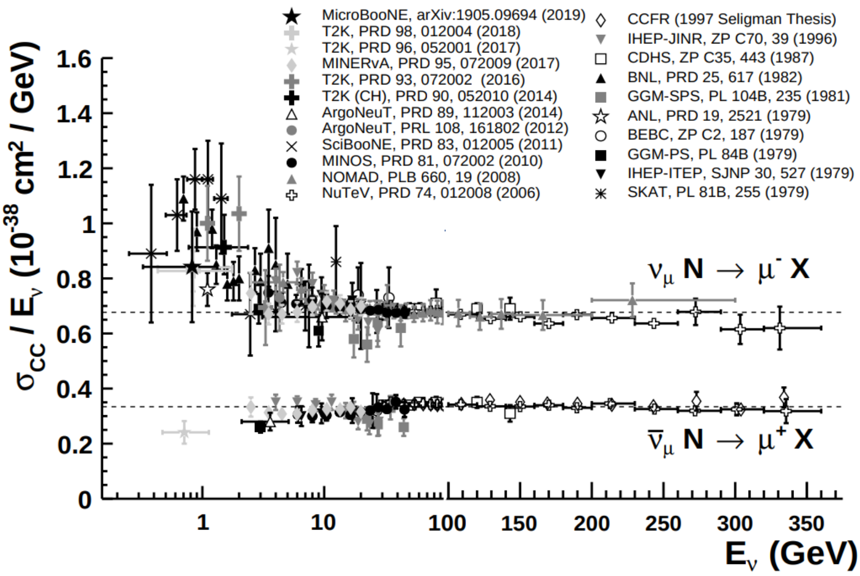


Figure 2.9: Neutrino cross section of interaction from the charged current as measured by different experiments over a large range of energies, for both neutrinos ν and antineutrinos $\bar{\nu}$ [?].

lower than the interaction cross section of a photon ($\sigma_\gamma \sim 10^{-25}$ cm^2 [?]). 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.

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 **??**. Indeed, their freeze-out abundance can be computed quite easily from Equation **??** [?], where the sum of the masses of the three neutrino flavors has been calculated to be lower than 0.17 eV [?] instead of the 11.5 eV expected to obtain the correct **DM!** relic abundance as observed today from the power spectrum of the **CMB!**.

$$\Omega_\nu h^2 = \sum_{i=1}^{3} \frac{m_{\nu_i}}{93 \text{ eV}} \tag{2.8}$$

Additionally, for several reasons explained in Section **??**, a good **DM!** candidate is expected to interact gravitationally, and to be cold or in other words, 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!** active neutrinos.

Sterile neutrinos

The most obvious **SM!** particles being ruled out as **DM!** candidates, we can now consider other **BSM!** theories responsible for the introduction of additional particles that could have the properties searched for. The first of these theories introduces 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 [**?**]. A superior bound of 50 keV on such particles can also be obtained considering limits on the observation of the monochromatic decay γ line originating from the one loop radiative decay $N \to \nu + \gamma$ of such particles.

Several experiments are already searching for this kind of particles at this level of energy. Most of these experiments focus on the analysis of γ-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 **??**) which did not match any known atomic emission line and which is actually consistent with an eventual **DM!** signal [**?**, **?**], since most of the possible instrumental contamination effects have been ruled out over the course of the last few years.

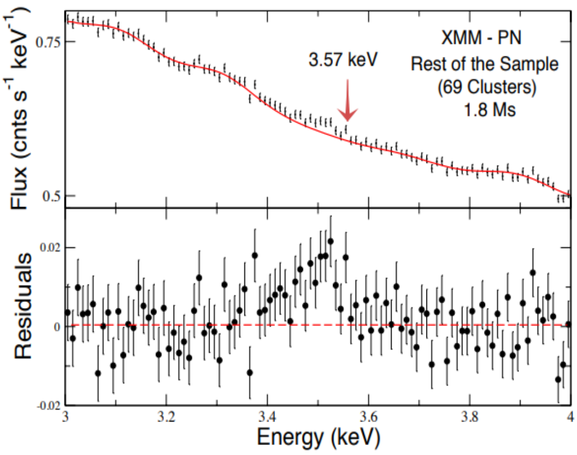


Figure 2.10: 3.57 keV emission line detected with a 4.5σ **CL!** by the XMM-Newton telescope in 2014, which could be a hint of the presence of **DM!** [**?**].

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 [**?**]. Recent observations actually ruled out at the 99% **CL!** an eventual **DM!** origin for this particular line [**?**], but further studies are still ongoing.

Axions

Axions could also explain the particle nature of **DM!**, since their existence is enough to 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 **QCD!** (**QCD!**). This issue is the following: the usual Θ term of the **QCD!** Lagrangian, the **QCD!** vacuum angle shown in Equation **??** [**?**] 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 have 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 Θ, the **QCD!** vacuum angle, is small enough so that this term becomes negligible. However, by definition, the Θ angle should be between 0 and 2π, so there is no physical reason for this parameter to be small, unless some new physics can be introduced, such as the theory developed in 1977 by Peccei and Quinn [**?**] by relaxing Θ from a parameter to a dynamic variable and absorbing it through the introduction of a new pseudoscalar particle that was called the axion.

$$\mathcal{L}_\Theta = \frac{\Theta}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}G_a^{\mu\mu}G_a^{\rho\sigma} \tag{2.9}$$

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, because their actual abundance can easily be computed from Equation **??** [**?**], from which we could conclude that an axion having a mass of $\sim$20 μeV could account for the **DM!** relic density of the Universe, as observed today.

$$\Omega_a \simeq \left(\frac{6\mu\text{eV}}{m_a}\right)^{7/6} \tag{2.10}$$

Several axion search experiments have therefore been set up, such as the **ADMX!** (**ADMX!**), a resonant microwave cavity installed at the University of Washington, the **CAST!** (**CAST!**), a CERN experiment observing the Sun which came on-line in 2002 and which managed in 2014 to turn up definitely the existence of solar axions [**?**], or the **IAXO!** (**IAXO!**), whose aim will 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 **??**.

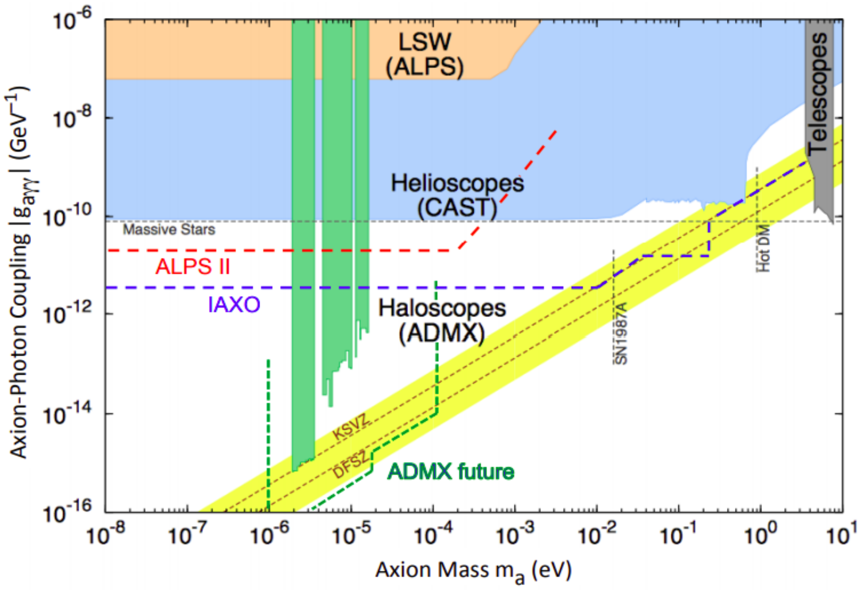


Figure 2.11: Axions exclusion summary plot and projected coverage of axion searches experiments, such as **ADMX!**, **CAST!** and **IAXO!** [**?**].

WIMP!s (WIMP!s)

The actual **DM!** candidates that will be mostly considered throughout this work are the so-called **WIMP!**s, which are expected to interact, even though very 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 dark matter 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 **??**, we can find an expression relating the relic abundance of **DM!** Ω_χ with its annihilation cross section $\langle \sigma_A v \rangle$ through Equation **??** [**?**].

$$\Omega_\chi h^2 \sim \frac{3 \cdot 10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle \sigma_A v \rangle} \tag{2.11}$$

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 $\sim 3 \cdot 10^{-27}$ cm^3 s^{-1}, which corresponds to the typical value given by **WIMP!**s for a range of dark matter masses matching the expected one.

This kind of particle basically arises in various **BSM!** theories, such as the **LSP!** (**LSP!**) 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 dark matter candidate in most of the **MSSM!** (**MSSM!**) theories, the so-called neutralino χ [**?**].

The **WIMP!**s are also interesting in the sense that introducing them in the terms of this supersymmetric theory would not only give us a strong dark matter candidate, but would also solve the hierarchy problem, the apparent large discrepancy among the masses of the different **SM!** particles ranging several orders of magnitude.

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 to try and detect **DM!** particles in order to study them, as represented in Figure **??**, 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 of 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.

2.5.1 Direct searches

From cosmological observations, we know that we live in a halo of dark matter. In this case, **WIMP!**s 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 the atoms. Indeed, the transfer of momentum between these two particles in this case might be detectable with the correct experimental device, typically

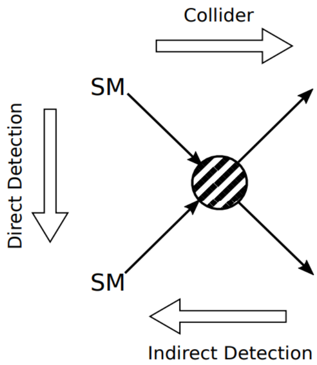


Figure 2.12: Schematic view of the three main **DM!** detection strategies: direct, indirect and collider production searches [**?**].

placed deep underground to have the lowest possible levels of backgrounds, 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 **??**, rate which ends up being described by a simple steeply falling exponential function as shown in Figure **??** [**?**], where E_{nr} is the nuclear recoil energy measured, m_χ is the **WIMP!** mass, σ its cross section, $\rho_0 = 0.3$ GeV cm^{-3} is the local dark matter density and $f(v)$ the normalized **WIMP!** velocity distribution.

$$\frac{dR}{dE_{nr}} = \frac{\rho_0 M}{m_N m_\chi} \int_{v_{\min}}^{v_{\text{esc}}} v f(v) \frac{d\sigma}{dE_{nr}} dv \propto \exp\left(-\frac{E_{nr}}{E_0} \frac{4 m_\chi m_N}{(m_\chi + m_N)^2} \right) \tag{2.12}$$

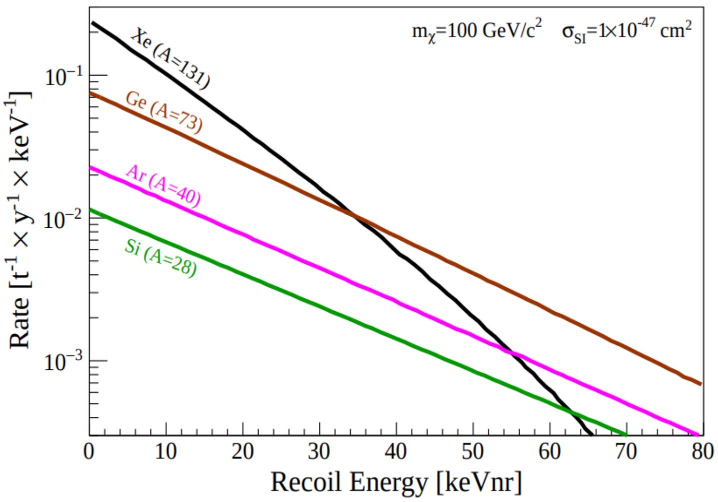


Figure 2.13: Nuclear recoil spectra induced in different materials for a given **DM! WIMP!** of 100 GeV, assuming a WIMP-nucleon **SI!** (**SI!**) cross section [**?**].

From this relation, the Equation **??** can be easily derived, representing this time the number of 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.

$$N = T \int_{E_{\min}}^{E_{\max}} \epsilon(E_{nr}) \frac{dR}{dE_{nr}} \, dE_{nr} \tag{2.13}$$

The maximal velocity v_{esc} used as superior bound of the integral in Equation **??** has actually been measured to be in the range $[498 - 608]$ km/s at the 90% **CL!** [**?**], 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 the annual modulation of the observed count rate, due to the movement of the Earth around the Sun, as shown in Figure **??** [**?**], since this velocity is not negligible compared to the escape velocity v_{esc}.

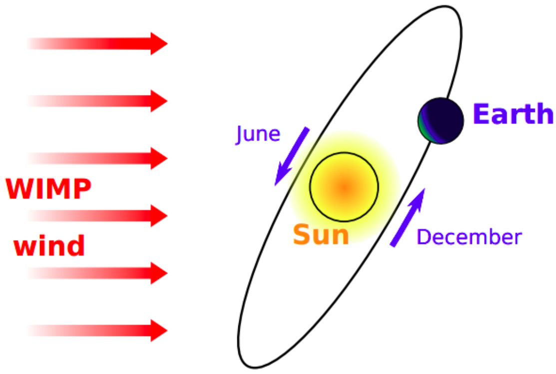


Figure 2.14: Schematic representation of the annual modulation of the WIMP wind introduced by the motion of rotation of the Earth around the Sun [**?**].

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. It is extremely important to take into account this effect since, as we saw on the previous equations such as Equation **??**, 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 **??**, where the periodical part usually introduces a $\sim 5\%$ deviation [**?**].

$$N(t) = B + N_0 + N_m \cos(\omega(t - t_0)) \tag{2.14}$$

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 **SI!** (**SI!**) (proceeding through the scalar term) and **SD!** (**SD!**) (proceeding through the axial term of the Lagrangian) searches, since the interaction cross section σ of Equation **??** is expected to be different for **DM!** particles having a spin 0 or not, as shown in Equation **??**, where F is a factor accounting for the dependence of the scattering on the energy. This means that results obtained by either hypothesis cannot be compared with each other.

$$\frac{d\sigma}{dE_{nr}} \propto \sigma_{SI} F_{SI}^2(E_{nr}) + \sigma_{SD} F_{SD}^2(E_{nr}) \tag{2.15}$$

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 is usually reduced either by placing the detector underground (to reduce the contamination due to cosmic rays), by increasing the statistics (by repeating the experiment and/or observing for a larger amount of time) 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 **??**.

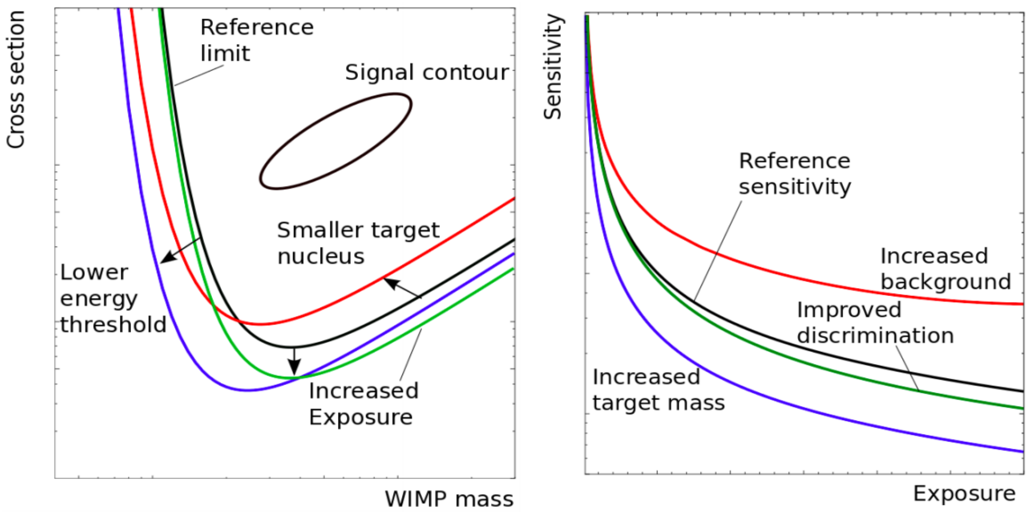


Figure 2.15: Impact of different experimental parameters on the final limits depending on the cross section and **WIMP!** mass (on the left) or sensitivity and exposure (on the right), with respect to the expected limits (black curve) [?].

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 de-excitation of the scattered nucleus, and some even search for the heat produced by these collisions under the form of phonons (collective excitation phenomena in condenses matter) in a crystal. These different search strategies have been summarized schematically in Figure **??**.

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 limits on the scattering cross section depending on the models parameters, as seen in Figure **??** for multiple experiments at once.

However, the DAMA experiment at the **LNGS!** (**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 **??** [?]. 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.

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 γ-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 **UED!** (**UED!**) do provide solid dark matter candidates (the lightest supersymmetric particle

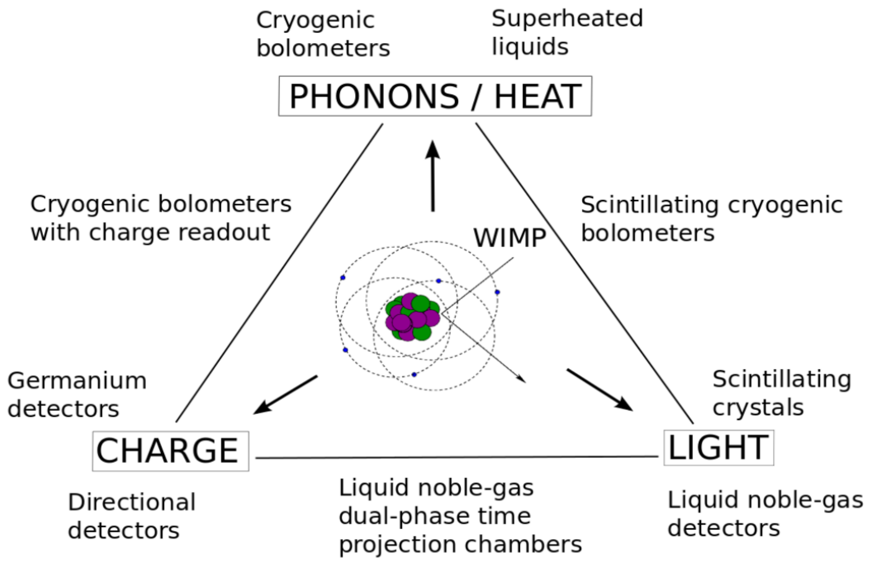


Figure 2.16: Schematic representation of the three main strategies to detect directly the interaction between **DM!** particles and an ordinary nucleus [**?**].

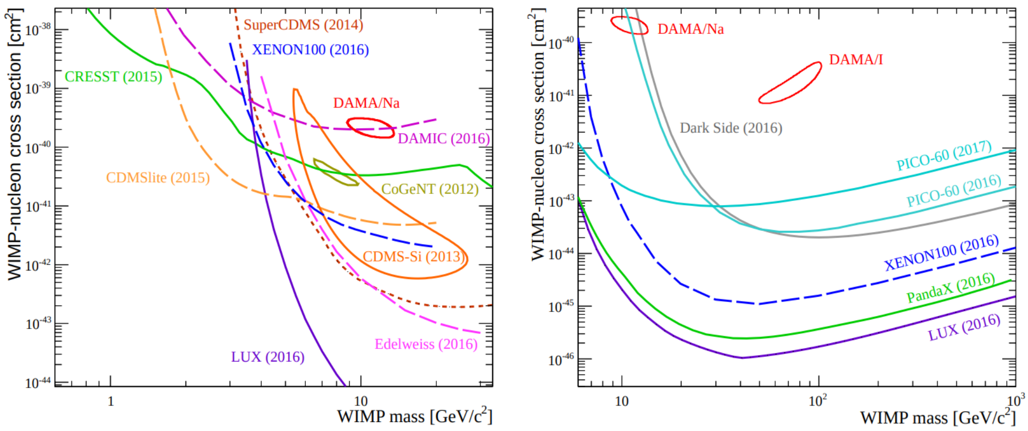


Figure 2.17: Exclusion limits obtained by various direct detection experiments considering a **SI!** interaction cross section for low **WIMP!** (on the left) or high **WIMP!** masses (on the right) [**?**].

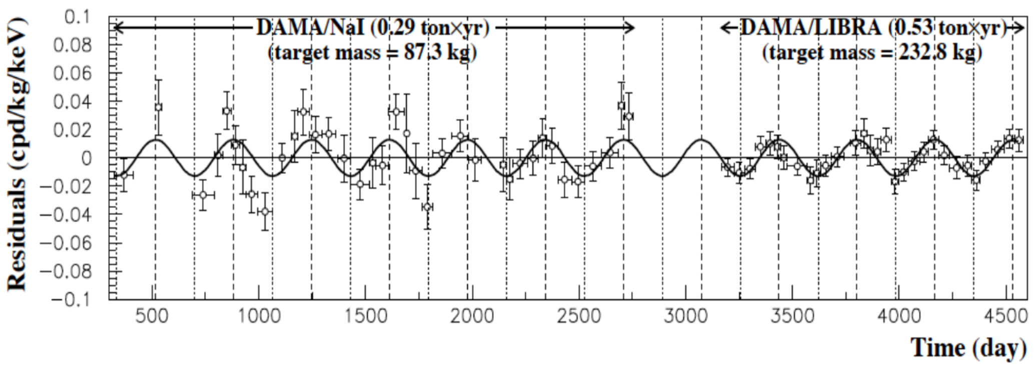


Figure 2.18: Observed and expected annual modulation in single hits events in the 2-6 keV energy range by the DAMA experiment [**?**].

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 Section **??**.

Indirect searches are actually extremely useful since they are sensitive to the **DM!** annihilation cross section, mass and the density profile of **DM!** halos $\rho(\overrightarrow{r}(s,\Omega))$, usually represented by a **NFW!** (**NFW!**) profile, as shown in Equation **??**, 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 [**?**].

$$\rho(r) \propto \frac{r_s}{r\left(1+\left(\frac{r}{r_s}\right)^2\right)} \tag{2.16}$$

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

$$\frac{d\Phi}{d\Omega dE} = \frac{\sigma v}{8\pi m_\chi^2} \cdot \frac{dN}{dE} \iint_{l,\Omega} \rho^2(\overrightarrow{r}(s,\Omega))\ dl\ d\Omega \equiv P \cdot J(\Delta\Omega) \tag{2.17}$$

This equation is extremely important for two reasons. First of all, it shows that, if a signal is found in indirect 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 observation are available for us to analyze the eventual annihilation of **DM!**, several strategies can be used in order to detect **DM!** indirectly, by searching different kinds of **SM!** particles. Anyway, all these strategies have one goal in common: try to reduce the backgrounds, since the signal searched for is usually quite low while the uncertainties associated to the backgrounds in astrophysics are usually quite high.

Through γ-rays detection

The golden channel for such searches is through the production of γ-rays by **DM!** annihilation $\chi\chi \rightarrow \gamma + X$ or decay $\chi \rightarrow \gamma + X$, mainly because the energy scale of the **WIMP!**s implies that most of the annihilation and decay radiation will be emitted in this range of energies and because γ-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 in this category is the **LAT!** (**LAT!**), a pair production detector launched in June 2008 and mostly sensitive to γ-rays between 20 MeV and 300 GeV [?]. This experiment has managed to exclude a large portion of the phase space, as seen in Figure **??**. The GAMMA-400 experiment, whose launch has not happened yet, will pick up the work of **LAT!**, by studying a similar range but with much improved angular and energy resolutions [?].

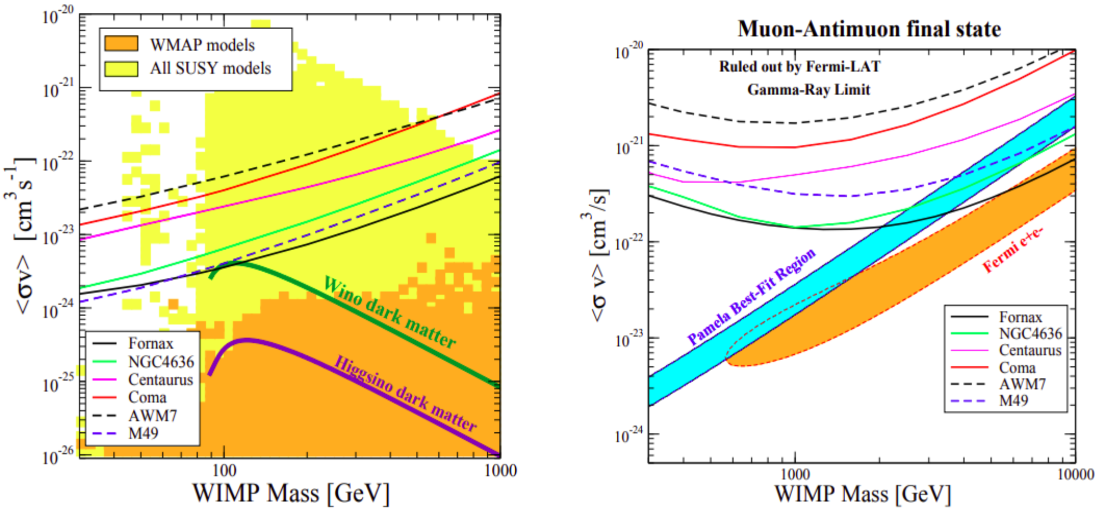


Figure 2.19: Upper limits on the **DM!** annihilation cross section considering $b\bar{b}$ (on the left) and $\mu^+\mu^-$ (on the right) final states as a function of the **WIMP!** mass, for different clusters studied [?].

It is much harder for **DM!** to produce high energy γ-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 **IACT!** (**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 **CTA!** (**CTA!**) is a brand new telescope of this kind whose construction is supposed to end in 2025, that should improve greatly the sensitivity of high masses indirect **DM!** searches.

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 much harder than detecting γ-rays and usually involves huge tanks of water in which neutrinos can be detected with the Cherenkov effect, which consists of the emission of an electromagnetic radiation when a charged particle moves through a dielectric medium with a speed greater than the speed of light in this particular medium.

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 requires 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 **??**, typically several orders of magnitude larger than the signal. Multiple strategies therefore need to be put in place in order to reduce the background levels 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.

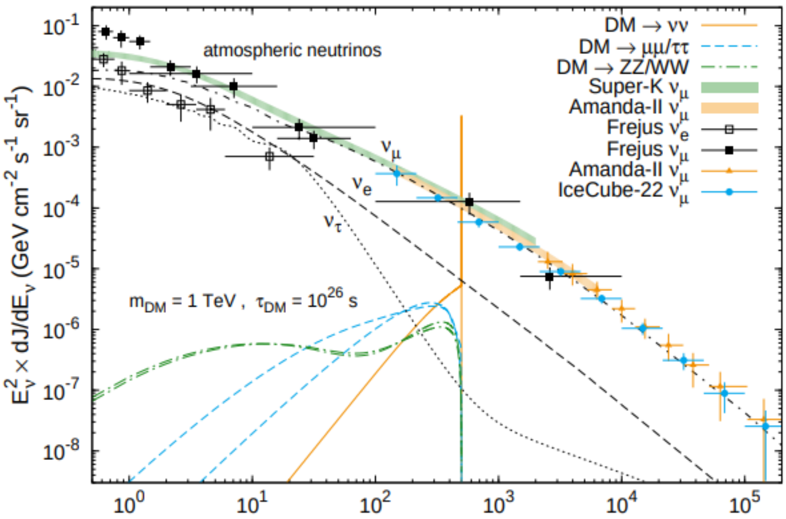


Figure 2.20: Neutrino spectra for a scalar **DM!** candidate of 1 TeV for different indirect detection experiments and the corresponding background level expected [**?**].

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 or even impossible.

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 **AMS!** (**AMS!**), installed in the International Space Station has also studied such radiation from a range of a few 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 **??**.

2.5.3 Collider production

In this particular kind of searches, we are 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.

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 **EFT!** (**EFT!**) approach is usually considered to be the easiest, even though it can still

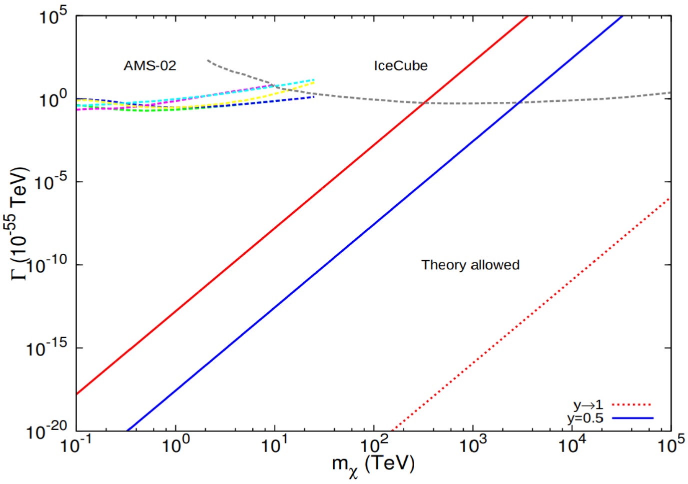


Figure 2.21: Limits of the decay width of the interaction with respect to the **DM!** mass obtained by both IceCube and **AMS!** [**?**].

give us plenty of information about this kind of interaction. It relies the assumption 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 in this case. According to this approach, represented in Figure **??**, the interaction can be described using simplified operators (the most simple ones being either scalar $\bar{\chi}\chi\bar{q}q$, pseudoscalar $\bar{\chi}\gamma^5\chi\bar{q}\gamma^5 q$, vector $\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu q$ or axial-vector $\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu\gamma^5 q$) since, according to the assumption made, its mediator can usually be integrated out [**?**].

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 Λ, which can be related to the different couplings of the interaction as in Equation **??**, where g_q and g_χ are the coupling between the mediator and the **SM!** or **DM!** particle, and m_{med} is the mass of the mediator.

$$\frac{1}{\Lambda^2} = \frac{g_q g_\chi}{m_{\text{med}}} \tag{2.18}$$

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 **??**, while in collider experiments such as the **LHC!**, this is of the order of Θ(GeV-TeV).

$$E = \frac{m_\chi v_\chi^2}{2} \simeq 100\ \text{GeV} \cdot 10^{-6} \simeq 50\ \text{keV} \tag{2.19}$$

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 alternative 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 ϕ or pseudoscalar a (both having a spin 0), described by the Lagrangians in Equation **??**, considering the **DM!** candidate to be a Dirac fermion coupling to the **SM!**

through the mediator considered and under the assumption of **MFV!** (**MFV!**) (a proposal made to characterize the effects of flavor transitions in new theories of particle physics). 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 high mass, which will allow us to simplify the following equations [?].

$$\begin{cases} \mathcal{L}_{\text{fermion},\phi} \propto -g_\chi \phi \bar{\chi}\chi - \dfrac{\phi}{\sqrt{2}} \sum_i \left(g_u y_i^u \bar{u}_i u_i + g_d y_i^d \bar{d}_i d_i + g_l y_i^l \bar{l}_i l_i \right) \\ \mathcal{L}_{\text{fermion},a} \propto -i g_\chi a \bar{\chi}\gamma^5\chi - \dfrac{ia}{\sqrt{2}} \sum_i \left(g_u y_i^u \bar{u}_i \gamma^5 u_i + g_d y_i^d \bar{d}_i \gamma^5 d_i + g_l y_i^l \bar{l}_i \gamma^5 l_i \right) \end{cases} \tag{2.20}$$

An important parameter in this case is the decay width of this mediator Γ_{med}, given by Equation **??** for a scalar mediator ϕ or Equation **??** 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.

$$\begin{cases} \Gamma_\phi = \sum_f N_C \dfrac{y_f^2 g_\nu^2 m_\phi}{16\pi} \left(1 - \dfrac{4m_f^2}{m_\phi^2}\right)^{3/2} + \dfrac{g_\chi^2 m_\phi}{8\pi} \left(1 - \dfrac{4m_f^2}{m_\phi^2}\right)^{3/2} + \dfrac{\alpha_S^2 g_\nu^2 m_\phi^3}{32\pi^3 \nu^2} \left| f_\phi \left(\dfrac{4m_t^2}{m_\phi^2}\right)\right|^2 \\ f_\phi(\tau) = \tau \left(1 + (1-\tau) \arctan^2 \left(\dfrac{1}{\sqrt{\tau - 1}}\right)\right) \end{cases} \tag{2.21}$$

$$\begin{cases} \Gamma_a = \sum_f N_C \dfrac{y_f^2 g_\nu^2 m_a}{16\pi} \left(1 - \dfrac{4m_f^2}{m_a^2}\right)^{1/2} + \dfrac{g_\chi^2 m_a}{8\pi} \left(1 - \dfrac{4m_f^2}{m_a^2}\right)^{1/2} + \dfrac{\alpha_S^2 g_\nu^2 m_a^3}{32\pi^3 \nu^2} \left| f_a \left(\dfrac{4m_t^2}{m_a^2}\right)\right|^2 \\ f_a(\tau) = \tau \arctan^2 \left(\dfrac{1}{\sqrt{\tau - 1}}\right) \end{cases} \tag{2.22}$$

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

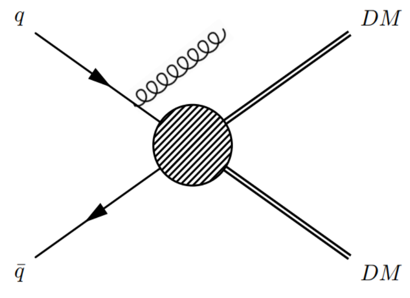


Figure 2.22: Schematic representation of a typical **EFT!** modelization of an **LHC!** event with an **ISR!** (**ISR!**) object used to trigger the event [?].

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 **??**. 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.

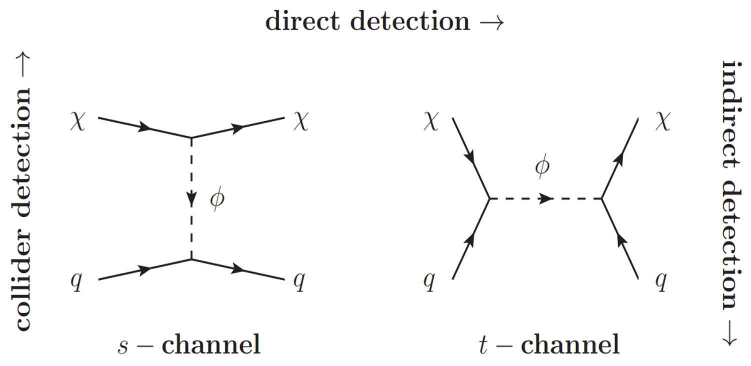


Figure 2.23: Schematic representation of a typical collider **DM!** production through s-channel and t-channel processes [**?**].

In this work, simplified models following the ATLAS-CMS Dark Matter Forum recommendations [**?**], with either scalar or pseudoscalar mediators will always be considered because of their relative simplicity and the lack of strong assumptions behind them.

2.6 Dark matter at the LHC! (LHC!)

The **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 **??**. 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 momentum 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!** (**MET!**), a key variable to all these searches that will be described in Section **??**. Depending on the nature of the X particle, several searches can be performed: we can for example mention the mono-γ [**?**, **?**] or mono-jet [**?**, **?**] searches performed by the **ATLAS!** and **CMS!** collaborations.

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

- The searches for **DM!** production in **association with heavy quark(s)** belong to the second category, such as the one performed in this work. Other analyses such as the one probing the $b\bar{b}$ + DM to its different final states depending on the decay of the bottom quark (at 8 or 13 TeV) also belong to this group [**?**].

  These searches typically 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** provide the best exclusion limits for most of the spin-1 mediated models considered (up to a few TeV for typical coupling choices) [**?**, **?**]. In this case, the sensitivity is obtained by searching for either narrow or large resonances on the exponentially falling QCD background, while the other categories of searches were mostly dedicated in searching for bumps in the **MET!** spectrum.
- **Multi-cascade searches**, involving the production of large cascade chains in which a stable dark matter particle is produced at the end. This kind of signature is very popular in supersymmetry models, which not only solve the problem of **DM!** but also the hierarchy problem while giving us perfect **DM!** candidates such as the neutralino χ, defined as the lightest stable supersymmetric particle, obtained in many of the **MSSM!** theories, having an hypothetical mass below the TeV scale [**?**].
- **Searches with Higgs driven states**: several models such as 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 requires a minimal modification of the **SM!** Lagrangian. It is then necessary to study the different Higgs production modes, such as the gluon fusion and **VBF!** (**VBF!**) mechanisms, to search for an eventual invisible decay of the Higgs into a couple of **DM!** particles, at least if we assume that its mass is lower than ~ 62.5 GeV, $m_H/2$ [**?**].
- Finally, and this is quite new, **long-lived searches** can also be performed. These are interesting 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 [**?**]. 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 from the different searches performed by the **CMS!** collaboration using the full (35.9 ± 0.9) fb^{-1} dataset collected at 13 TeV can be summarized in Figure **??** for spin-0 mediators and in Figure **??** when considering spin-1 mediators [**?**].

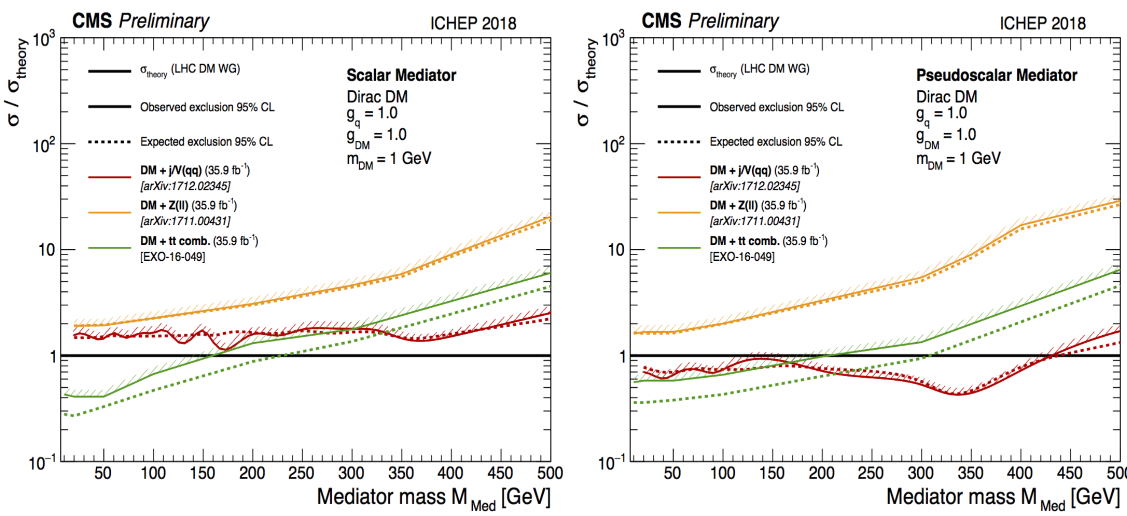


Figure 2.24: Observed and expected 95% exclusion limits obtained by different searches of the **CMS!** collaboration as a function the spin-0 scalar (on the left) or pseudoscalar (on the right) mediator.

These results have also been compared to the results obtained by several direct detection experiments in Figure **??**, considering both the **SD!** (**SD!**) and **SI!** (**SI!**) cases,.

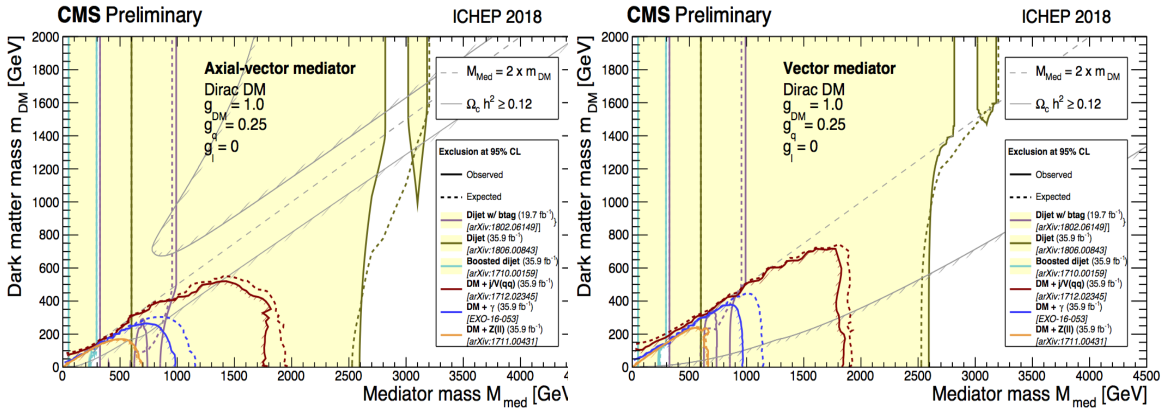


Figure 2.25: Observed and expected 95% exclusion limits obtained by different searches of the **CMS!** collaboration as a function of the spin-1 mediator, considering axial-vector (on the left) and axial (on the right) interactions [**?**].

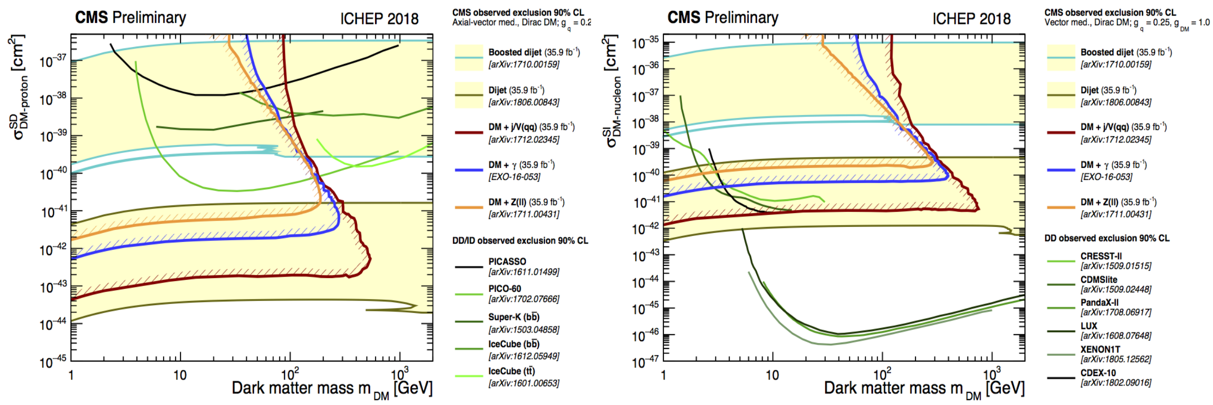


Figure 2.26: **CMS!** 90% exclusion limits compared to the most famous direct detection experiments for the **SD!** (on the left) and **SI!** (on the right) scenarios, obtained using similar couplings [**?**].

This particular analysis and our signal of interest, the search for **DM!** production together with a single or a pair of top quarks falls into the second category. It is already important to note that this analysis has been performed considering the **DM!** candidate to be a Dirac fermion with all the couplings equal and set to 1, as recommended by the **DMWG!** (**DMWG!**) [**?**].

2.7 The focus of this thesis

This thesis focuses on a particular search for **DM!**, produced in association with either one or two top quarks, considering only the dilepton final state of such processes. In particular, two different categories of signal, later referred to as the $t/\bar{t}$+DM and $t\bar{t}$+DM processes, will be considered and combined in order to compute upper limits on the **DM!** production signal strength.

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 the $t/\bar{t}$+DM analysis. This process is expected to be mediated by a spin-0 mediator, either scalar ϕ or pseudoscalar a, 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 **??**.

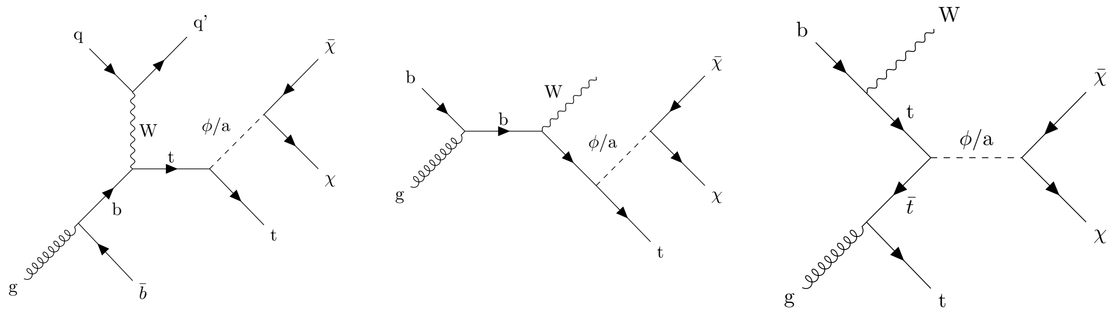


Figure 2.27: Feynman diagrams involving the production of **DM!** with a single top quark according to its t-channel W boson (on the left), or tW (on the center and on the right) production modes.

As discussed previously, the top quark will be dynamically favored due to its high mass and therefore high Yukawa coupling, but this has another consequence as well, since the lifetime of this quark is extremely low, of the order of 10^{-15} s [**?**]. This means that this particle usually decays before being able to form hadrons, 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 quarks and/or leptons. Even though this will be detailed later on, 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.

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 **??**.

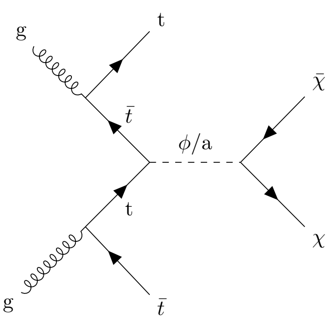


Figure 2.28: Feynman diagram of a typical $t\bar{t}$+DM event.

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. In this case as well, both scalar ϕ and pseudoscalar a mediators will be considered, with a mass range from 10 to 500 GeV.

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, giving us an additional contribution of **MET!** ($10.8 \pm 0.09\%$ for each charged lepton) [**?**].

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 W bosons decay into two neutrinos and two leptons of opposite charge).

Given the **BR!** (**BR!**) of the W decay, it is easy to see that the dileptonic channel, the focus of this work, is not favored statistically. However, this channel features less backgrounds than other channels, resulting in a better signal isolation, and leptons can usually be reconstructed in a better way than jets (cf. Chapter **??**), leading to lower uncertainties and improved limits.

2.8 Previous relevant results

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

First of all, the **ATLAS! collaboration** published interesting results at this center of mass energy, being the first results for a combined search of stop and for a similar dark matter search in the dilepton final state, and using the full Run II dataset, as shown in Figure **??** [**?**]. According to these results, the collaboration has been able to achieve an exclusion of scalar (pseudoscalar) mediators with masses up to 250 (300) GeV. It is important to note though that the **ATLAS!** collaboration used **NLO!** (**NLO!**) cross-sections, while the **CMS!** collaboration uses by convention **LO!** (**LO!**) cross-sections for the signals of interest, typically around 30% lower around the exclusion region, making the **CMS!**limits of exclusion artificially worse.

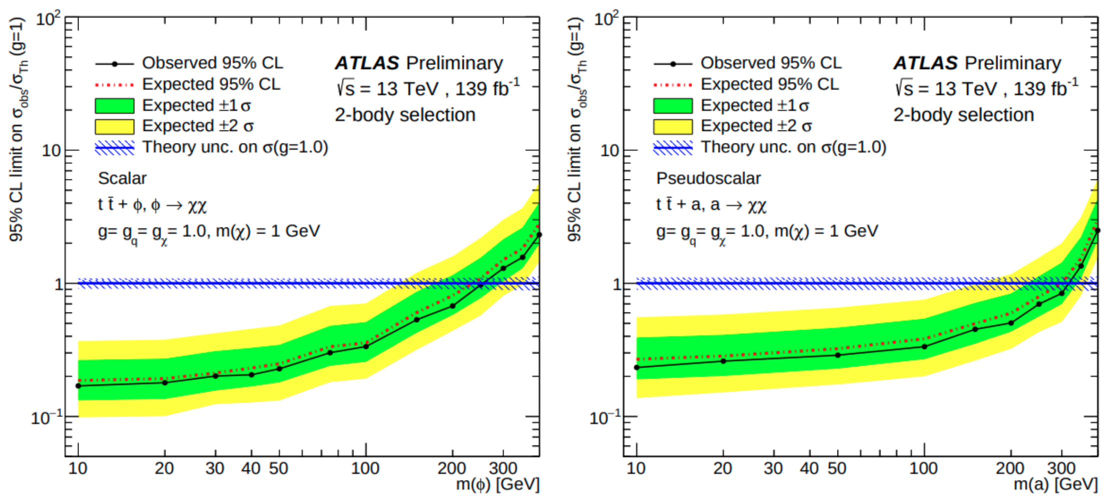


Figure 2.29: Exclusion limits at the 95% **CL!** obtained by **ATLAS!** considering scalar (on the left) and pseudoscalar (on the right) mediators, for a **DM!** mass of 1 GeV, considering the dilepton final state of such model and the full Run II dataset [**?**].

Additionally, summary plots for ATLAS dark matter searches, showing the exclusion limits obtained for the three possible final states have been published, as shown in Figure ??. The collaboration did not perform any combination between the different channels though.

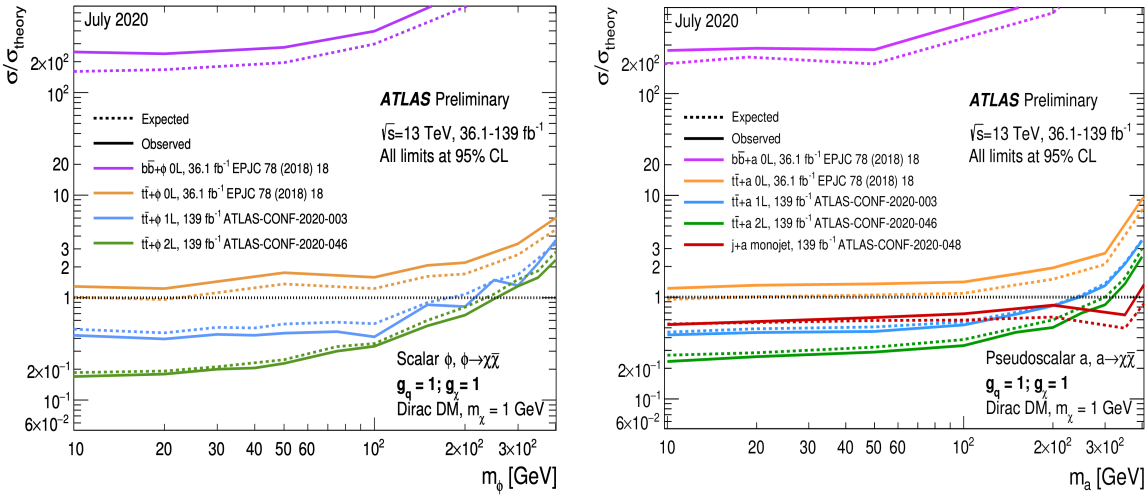


Figure 2.30: Exclusion limits at the 95% **CL!** obtained by **ATLAS!** considering scalar (on the left) and pseudoscalar (on the right) mediators, for a **DM!** mass of 1 GeV [?].

The **CMS! collaboration** published on the other hand in 2018 some results in the dilepton final state for the $t\bar{t}$+DM signal only, using the 35.9 fb^{-1} of data collected during the year 2016, as shown in Figure ??. Scalar models were excluded up to a mass of 80 GeV in this case, while no exclusion was obtained for the pseudoscalar mediators.

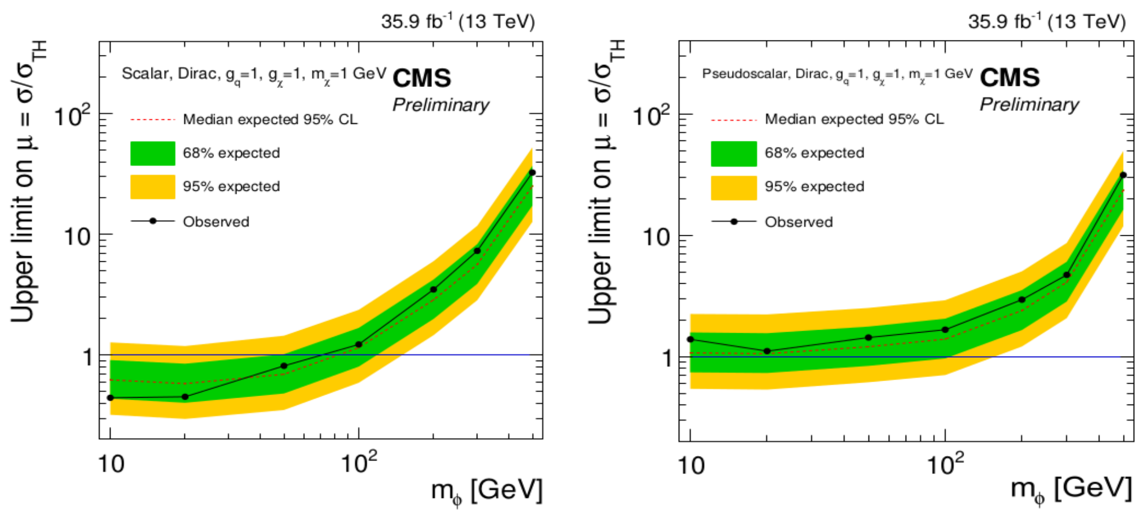


Figure 2.31: Exclusion limits at the 95% **CL!** obtained using the 2016 data by **CMS!** considering the dilepton final state of the $t\bar{t}$+DM scalar (on the left) and pseudoscalar (on the right) mediators, for a **DM!** mass of 1 GeV.

The three different final states possible (hadronic, semileptonic and dileptonic) were then combined, considering both scalar and pseudoscalar mediators [?], as shown in Figure ??. Models with a scalar (pseudoscalar) mediator up to 240 (320) GeV were respectively excluded in this case.

Finally, in 2019, a **CMS!** combination of the $t/\bar{t}$+DM and $t\bar{t}$+DM analyses has also been published [?], combining this time only the hadronic and semileptonic channels of both analyses. The limits obtained in this case are represented in Figure ??, 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.

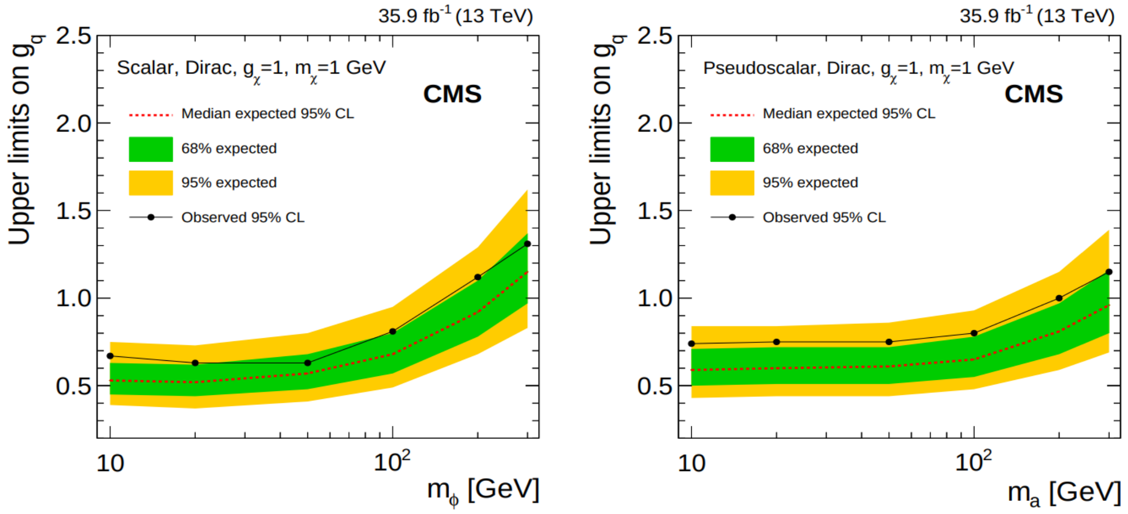


Figure 2.32: Expected and observed 95% **CL!** limits on the **DM!** production cross sections shown considering scalar (on the left) and pseudoscalar (on the right) mediators for the interaction [**?**].

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!** and for the couplings considered. This combined analysis actually leads to the most stringent exclusion limits of the **LHC!** so far on the production of **DM!** through these categories of spin-0 mediators.

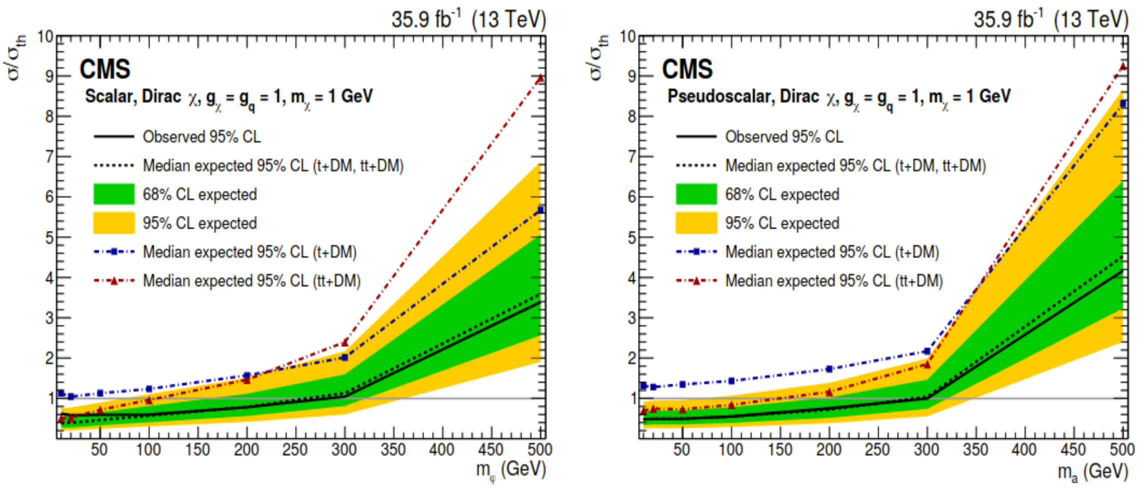


Figure 2.33: Expected and observed 95% **CL!** limits on the **DM!** production cross sections shown considering scalar (on the left) and pseudoscalar (on the right) mediators for the interaction [**?**].

CHAPTER 3

The experimental setup

The data analyzed throughout this work is the result of several years of proton-proton collisions produced at the **LHC!** with a center of mass energy of 13 TeV, and the particles emerging from these collisions have been recorded with the **CMS!** detector. In order to provide the experimental context of this thesis, the Section **??** of this section will be dedicated to the description of the accelerator, while the detector itself, whose construction involved thousands of people and made out of several different layers, will be described in Section **??**.

3.1 The LHC! (LHC!)

The **LHC!** (**LHC!**) is a superconducting particle accelerator able to accelerate protons and lead ions up to velocities close to the speed of light (0.99999999c). Planned since the end of the 20th century, this accelerator, a 27 kilometers ring put 100 meters underground (under France and Switzerland) to avoid part of the contamination due to the cosmic rays and located at **CERN!**, has now been running for more than 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 [**?**, **?**] but never observed in any experiment. The discovery of the Higgs boson, announced at **CERN!** on the 4th of July 2012 [**?**, **?**], was then a magnificent achievement of the accelerator, after only a few years 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, couplings or spin, many groups of scientists are involved in different kinds of **BSM!** physics since the **LHC!**, whose center of mass energy has kept increasing over the years, allows 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 searches for dark matter as the one performed in this work.

3.1.1 The LHC! in a nutshell

The **LHC!** is an underground particle accelerator built in the 27 kilometers tunnel where the **LEP!** (**LEP!**) previously sat [**?**]. This machine is accelerating two beams of 10^{10} protons or $7 \cdot 10^7$ 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 while being pushed from behind by a positive 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 is reached, the protons are then smashed together in four different places on the **LHC!**, where the four detectors (**ATLAS!**, **CMS!**, **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 mainly in order to introduce some redundancy and to check the results). **ALICE!** (**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 quark-gluon plasma [**?**]. Finally, LHCb has been designed to study in particular the CP violation phenomena, which could be the sign of some new physics [**?**].

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 the input energies required ($\sim$ 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 **PS!** (**PS!**), the **SPS!** (**SPS!**) and finally the **LHC!** itself (all this chain of acceleration can be found in Figure **??**).

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 30-35 collisions of protons happen at once, on average: this phenomena, usually referred to as **PU!** (**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 typically not possible, as setting up a beam takes time and they cannot be kept 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 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 **LS!** (**LS!**), in order to also have the time to perform additional upgrade operations of the machine.

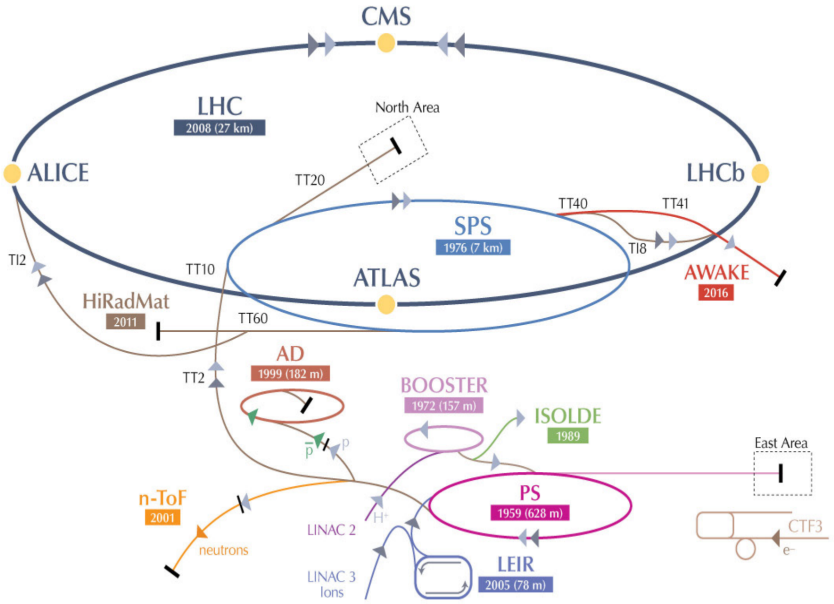


Figure 3.1: LHC injection chain and experiments performed at **CERN!** [?].

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 soon, in the Spring of 2022. The summary of the main (expected) parameters of operation of the **LHC!** across the different Runs of operations can be found in Table **??**.

3.1.2 Key parameters

Center of mass energy

The center of mass energy is defined as a Lorentz invariant quantity under any kind of boost resulting of the collisions between two protons (defined as $E_1, \overrightarrow{p_1}, m_1$ and $E_2, \overrightarrow{p_2}, m_2$) with a θ angle, as developed in Equation **??**. It is a key variable of the **LHC!** since the phase space of particles that can be explored directly depends on this value.

$$\sqrt{s} = \sqrt{(m_1)^2 + (m_2)^2 + 2\left(E_1 E_2 - 2|\overrightarrow{p_1}|\ |\overrightarrow{p_2}|\ \cos(\theta)\right)} \tag{3.1}$$

The LHC started operating in 2008 running at an energy of 7 TeV, quickly moved to 8 TeV and kept this level of energy until 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 13.5 to 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 another key variable for the operation of the **LHC!** since it gives an indication on the number of collisions per second provided 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 creation when colliding

Parameter	Run I	Run II	Run III	Design
Energy [TeV]	$7 \to 8$	13	13	14
Bunch spacing [ns]	50	25	25	25
Intensity [10^{11} protons per beam]	1.6	1.2	Up to 1.8	1.15
Bunches	1400	2500	2800	2800
Emittance [μm]	2.2	2.2	2.5	3.5
β^* [cm]	80	$30 \to 25$	$30 \to 25$	55
Crossing angle [μrad]	-	$300 \to 260$	$300 \to 260$	285
Peak luminosity [10^{34} cm^{-2} 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 [**?**].

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

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

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

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

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

- γ 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_{rev} is the frequency of revolution (11.2 kHz);
- k_B is the number of proton bunches per beam (2808 for a 25ns bunch spacing);
- N_p is the number of protons per bunch ($1.15 \cdot 10^{11}$ protons);
- ϵ_n is the transverse normalized emittance (3.75 μm);
- β^* is the betatron function at the interaction point (0.55 m);
- G 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 θ_C (285 μrad), and σ, σ_z, the transverse and longitudinal r.m.s. sizes (respectively 16.7 μm 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 radio-frequency 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 [**?**]).

The total integrated luminosity taken by the **LHC!** during its different years of operation has been summarized in Figure **??**. The final datasets available for the Run II and analyzed in this work have an integrated luminosity of (35.9 ± 0.9) fb^{-1} (2016), (41.5 ± 1.0) fb^{-1} (2017) and (59.7 ± 1.5) fb^{-1} (2018), resulting in a total dataset of (137.1 ± 2.0) fb^{-1} recorded during the Run II of operation. This roughly means that we expect to have produced around 137 events of any process whose cross section of production would be equal to 1 fb.

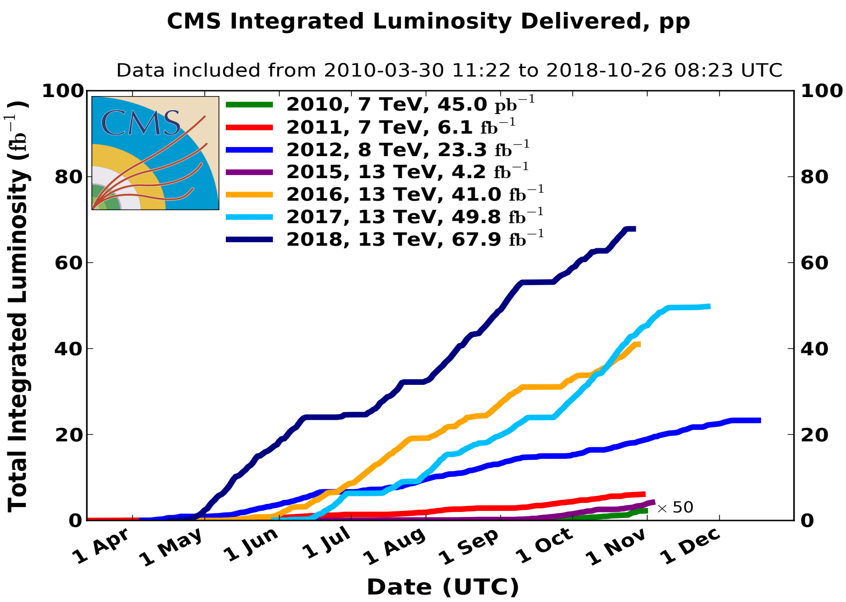


Figure 3.2: Integrated luminosity collected by **CMS!** over its different years of operation so far.

Number of simultaneous interactions: PU! (PU!)

The last key parameter of the **LHC!** discussed here is the number of simultaneous interactions, also called **PU!**. Usually, because of the high density of protons within the beams, a single bunch crossing in an experiment produces around 30-35 proton collisions, as seen in Figure **??**, defining the **PV!** (**PV!**) as the most interesting one, while the other vertices are usually referred to as the **PU!**. The tracker of **CMS!** therefore needs to be able to reconstruct all these events in order to define the **PV!** of the interaction, as explained in the next section.

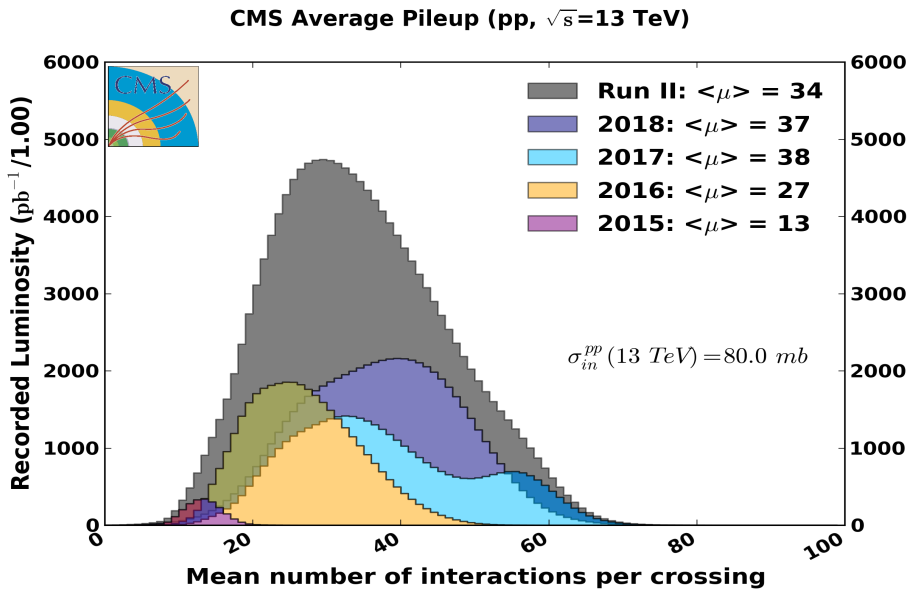


Figure 3.3: Mean **PU!** distribution and luminosity recorded by **CMS!** over the different years of operation of the **LHC!**.

3.2 The CMS! detector

The **CMS!** (**CMS!**) is one of the two general purpose detectors of the **LHC!** and is installed at the access point 5 of the **LHC!**. Its main purposes are to discover the Higgs boson, make precision measurements and to discover **BSM!** physics, such as the possible existence of dark matter.

CMS! 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. It therefore counts with a central part with cylindrical shape, the barrel, and two endcaps, one on each side of the detector. With its 14.000 tons, distributed over a diameter of 15 meters and a length of 28.7 meters, the detector can be considered quite compact and 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.

CMS! subdetectors

As shown in Figure **??**, **CMS!** is made out of different layers corresponding to different sub-detectors, each able to provide different kinds of information about the particles created by each collision [**?**]. These sub-detectors have all been designed in order to make the reconstruction of the different events efficient, fast and precise while surviving the rough conditions occurring at the **LHC!**.

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

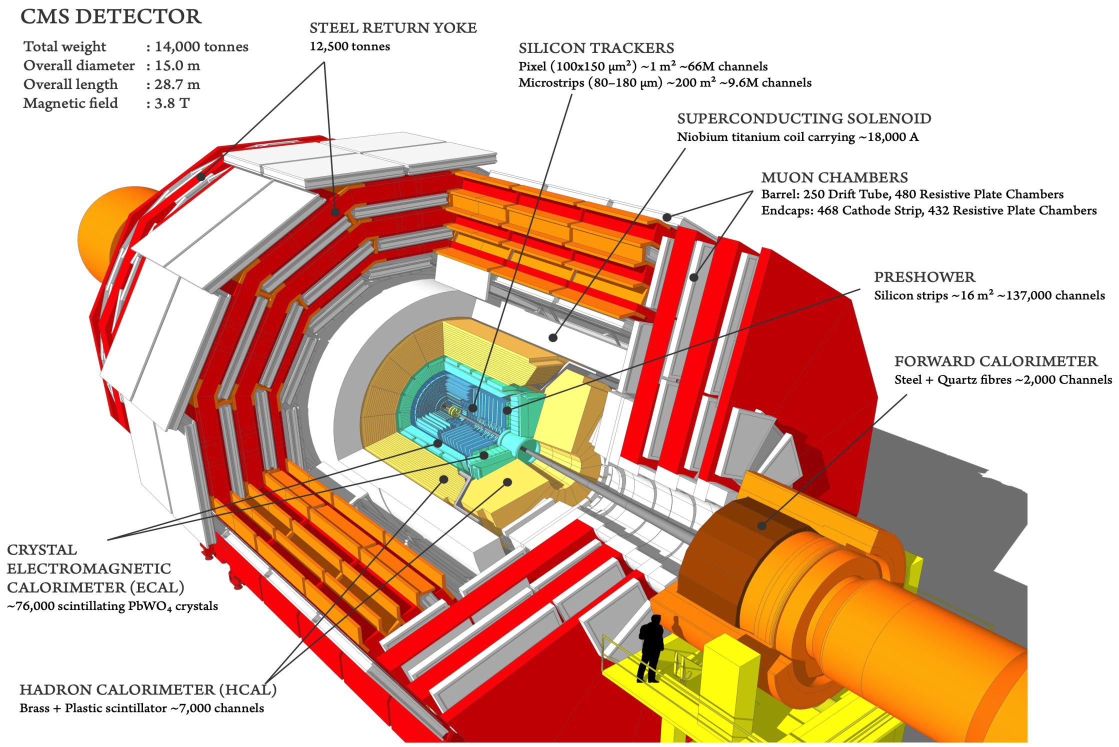


Figure 3.4: Schematic representation of the **CMS!** detector, along with all its sub-detectors and main characteristics.

The **CMS!** name partly comes from its central piece, a huge superconducting solenoid 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 **??**). 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 a way to measure their charge and energy from the measurement of the curvature of the binding induced.

Finally, on the outside of the detector the complete muon system, particularly efficient in **CMS!**, can be found. This sub-detector is currently made out of three main sub-systems (the **DT!**s (**DT!**s), **CSC!**s (**CSC!**s) and **RPC!**s (**RPC!**s)) and is responsible for the identification and measurement of the momentum of the muons produced by the collisions.

Coordinates system

As a convention, is has been decided to use within the **CMS!** collaboration a right-handed Cartesian coordinate system with the origin defined as the interaction point, 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 **??**.

The θ and ϕ angles are then defined as the angles between the z and y axes and the x and y axes respectively and the pseudorapidity η, defined in Equation **??**, a Lorentz invariant quantity under boosts quite often used in the different analyses since the multiplicity of high energy particles is roughly expected to be constant in η.

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

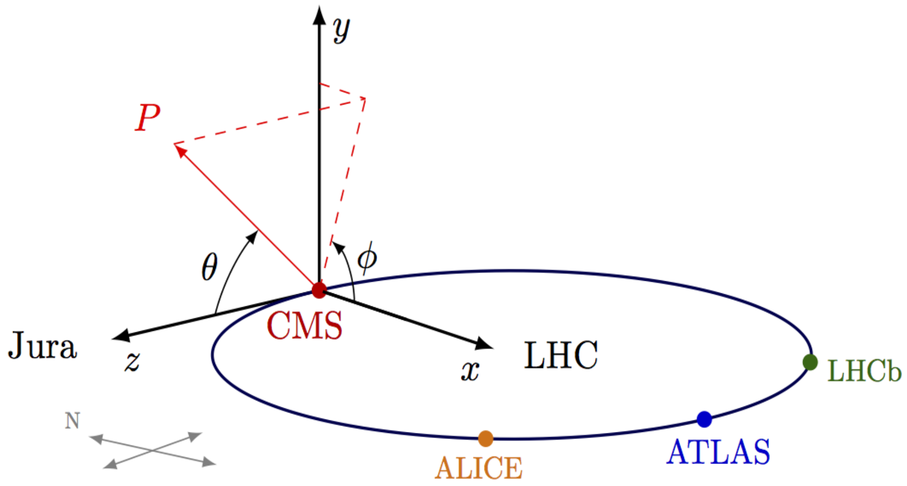


Figure 3.5: Schematic representation of the **CMS!** coordinate system used by convention.

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 harsh 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 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 **??**. A charged particle crossing the tracker will then leave a hit each time it crosses one of the silicon sensors, from which the track of the particle can be reconstructed using advanced reconstruction algorithms that will be detailed in Chapter **??**.

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 radii $r = 4.4, 7.3$ and 10.2 cm of the **PV!**), one on each side. In total,

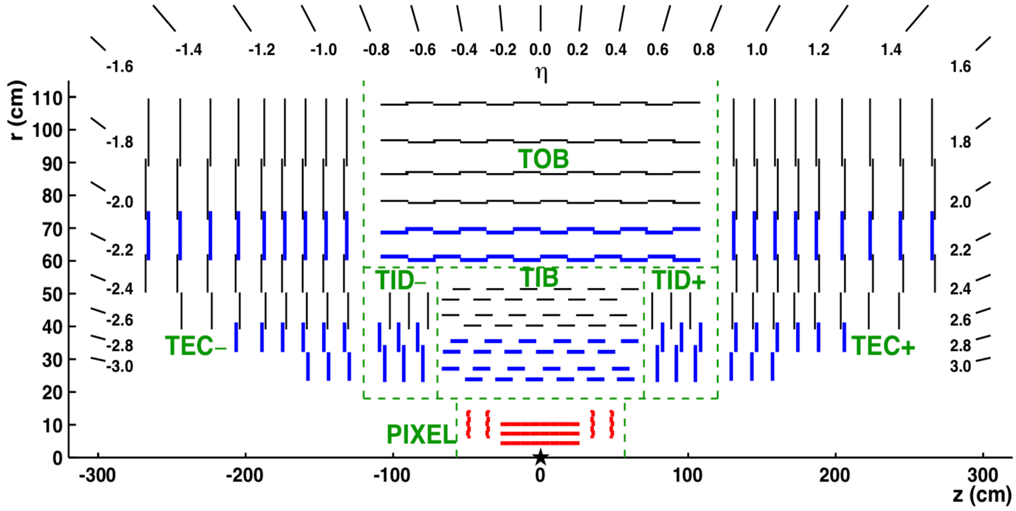


Figure 3.6: Schematic representation of the **CMS!** tracker, for different pseudorapidity values and along the z-axis [**?**].

more than 60 millions pixels make up the 1440 (1856 after an upgrade in 2017) modules of this detector, covering an area corresponding to ~ 1 m^2.

- The **silicon strip detector** on the other hand is composed of three different sub-systems, as seen in Figure **??** and covers a total area of $\sim$ 200 m^2. First of all, the **TIB/TBD!** (**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 **TOB!** (**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 **TEC!**s (**TEC!**s) are made out of 9 disks and complete the measurement of particles emitted along the z-axis and having a high pseudorapidity η.

The **CMS!** tracker is extremely performing: one can see first of all in Figure **??** that for high momentum tracks of 100 GeV, the resolution is 1-2% up to $|\eta| < 1.6$.

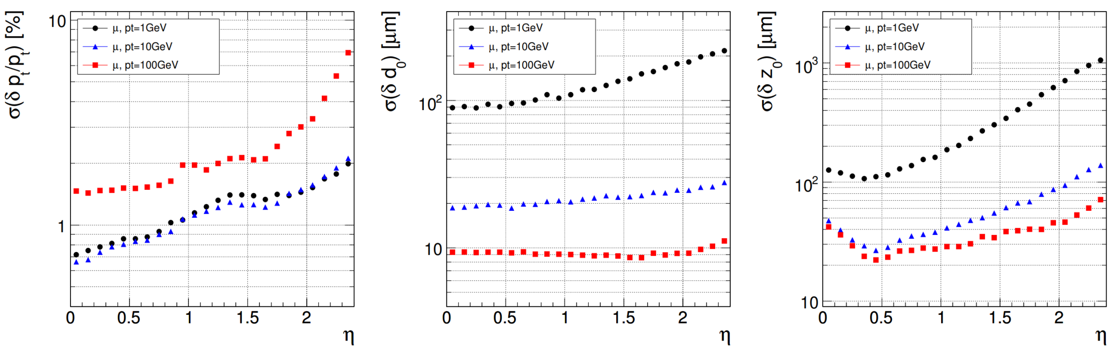


Figure 3.7: Expected resolution of muons transverse momentum (left), transverse impact parameter (middle) and longitudinal impact parameter (right), as a function of pseudorapidity and muon momentum (1, 10 and 100 Gev) [**?**].

Additionally, Figure **??** shows that muons are reconstructed with an efficiency higher than 99% for most of the pseudorapidity spectrum, even though this efficiency drops at high η 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 results in a lower reconstruction efficiency for such particles.

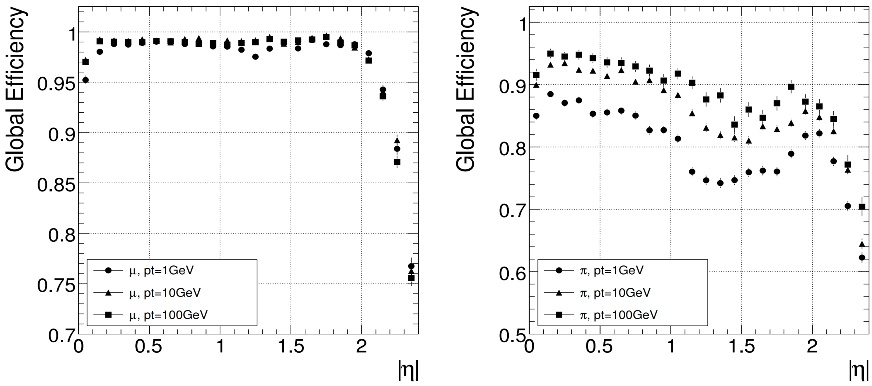


Figure 3.8: Tracker reconstruction efficiency of muons (on the left) and pions (on the right) in simulation for different pseudorapidities and particle momenta (1, 10 and 100 GeV) [**?**].

3.2.2 ECAL! (ECAL!)

The **ECAL!** of CMS is a detector mostly homogeneous which sits inside the solenoid and encloses the tracker system, giving 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 $|\eta| < 1.479$ the barrel part of the **ECAL!** (EB) can be found, made out of 61 200 lead tungstate ($PbWO_4$) 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 **??** and, finally, the preshower completes the **ECAL!**.

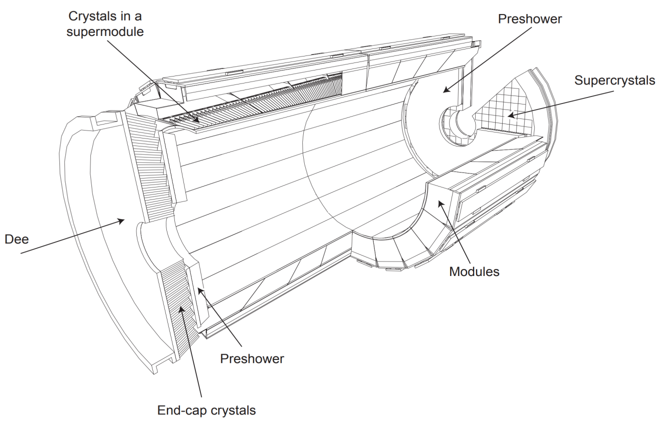


Figure 3.9: Schematic representation of the sub-systems of the **CMS! ECAL!** [**?**].

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 photo-detectors and photo-multipliers. This effect is known as electromagnetic shower, is represented in Figure **??** and is usually characterized using the so-called radiation length X_0, the mean distance over which a high-energy particle loses all but 1/e of its energy, then determining the total length of interaction of a particle in the **ECAL!**.

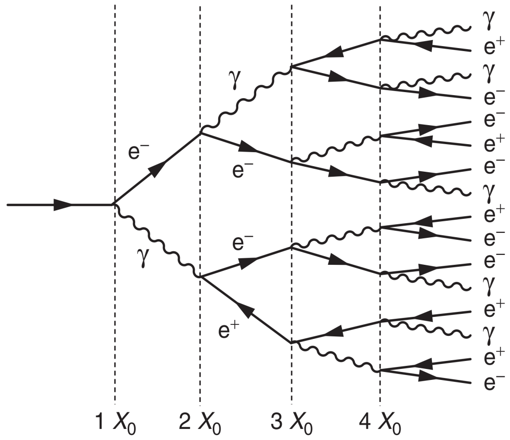


Figure 3.10: Schematic representation of a typical electromagnetic shower and the radiation length X_0 concept [**?**].

The high density, short radiation length and low scintillation decay time (smaller than the bunch spacing of 25ns) of the $PbWO_4$ 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 2.2 x 2.2 x 23 cm 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 photo-electrons per MeV of energy), they have to be connected to both avalanche photo-diodes and vacuum photo-triodes in order to multiply the signal measured. Finally, they have been mounted using a specific installation, slightly tilted with respect to both ϕ and η 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 **??** [**?**], accounting for several different effects, such as the stochastic nature of the observed scintillation, the electronics and **PU!** noise and the calibration and detector non-uniformity uncertainty.

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

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 $\pi^0 \to \gamma\gamma$ 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 HCAL! (HCAL!)

We know that charged hadrons lose energy in a continuous way 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 **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 λ parameter, the nuclear interaction length, defined as the mean distance between two interactions of relativistic hadrons. The nuclear interaction length λ 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 π^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 [?]. In this case, the energy resolution can therefore be expressed using Equation **??**.

$$\frac{\sigma_E}{E} > \frac{50\%}{\sqrt{E/\mathrm{GeV}}} \tag{3.6}$$

In **CMS!**, the **HCAL!**, represented in Figure **??**, 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), two endcaps (HE) and two symmetrical forward regions (HF) extending the pseudorapids 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 **HO!** (**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 its effective nuclear radiation length λ.

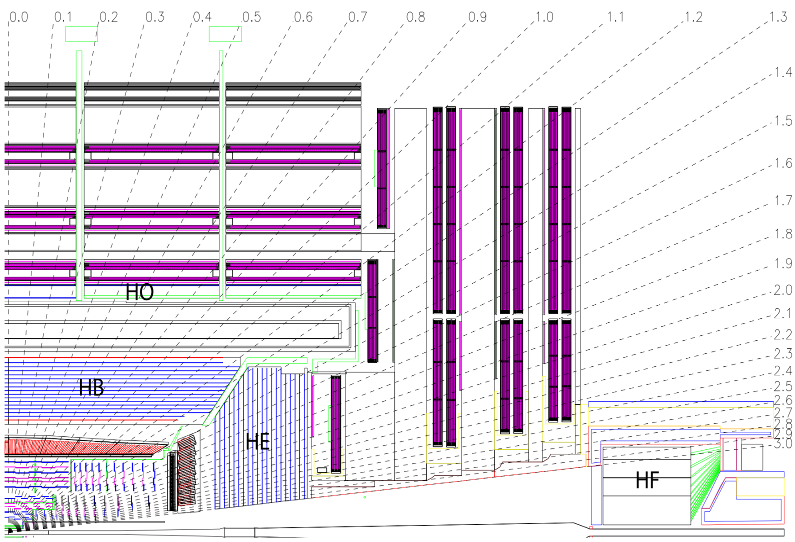


Figure 3.11: Schematic representation of the **HCAL!** sub-system in **CMS!** [?].

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 are smaller than 2 mm. The total absorber thickness at a 90° incidence angle is equal to only 5.8λ, 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 fiber concept to reduce the energy of detected photons and 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 pseudorapidity 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 **??**.

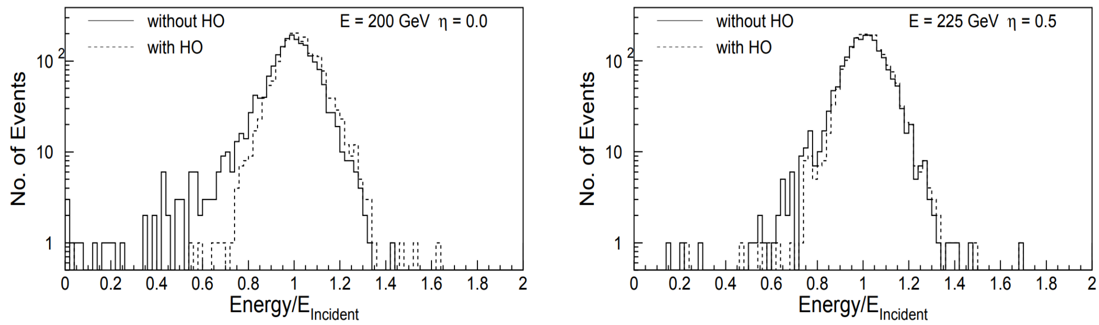


Figure 3.12: Distribution of the measured energy scaled to the incident energy for pions with incident energies of 200 GeV at $\eta = 0$ (on the left) and 225 GeV at $\eta = 0.5$ (on the right), with and without the inclusion of the **HO!** in the **HCAL!** system [?].

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 3.8 T magnetic field, storing when active a huge amount of energy (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 other 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 **??**. This solenoid has been designed to reach a momentum resolution $\Delta p/p \sim 10\%$ at $p = 1$ TeV.

$$\overrightarrow{F} = \frac{m\overrightarrow{v}^2}{R} = q\overrightarrow{E} + q\overrightarrow{v} \times \overrightarrow{B} = q\overrightarrow{v} \times \overrightarrow{B} \tag{3.7}$$

The hoop strain, normal stress parallel to the axis of cylindrical symmetry applied throughout this solenoid is quite large ($\epsilon = 130$ MPa) compared to the strain applied on other previous detectors and it 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 aluminum, is mechanically reinforced with an aluminum alloy.

The coil of the magnet is then completed with a huge steel yoke return system, as seen in Figure **??**, 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 also serves as the absorber plates of the **HO!** and muon detection system that will be described in the next section.

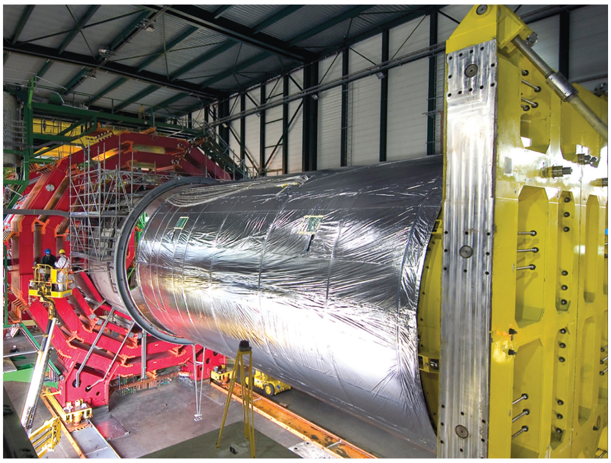


Figure 3.13: Picture of the solenoid system of **CMS!** being setup in the assembly hall.

Finally, a two pumping stations system has been put in place in order to setup a vacuum as strong as possible inside the 40 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 system 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.

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 distributed over a cylindrical area, 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 corresponding to more than 25 000 m^2.

Each category of muon chambers is typically used in a different pseudorapidity area, as shown in Figure **??**, in order to form a muon system as hermetic as possible.

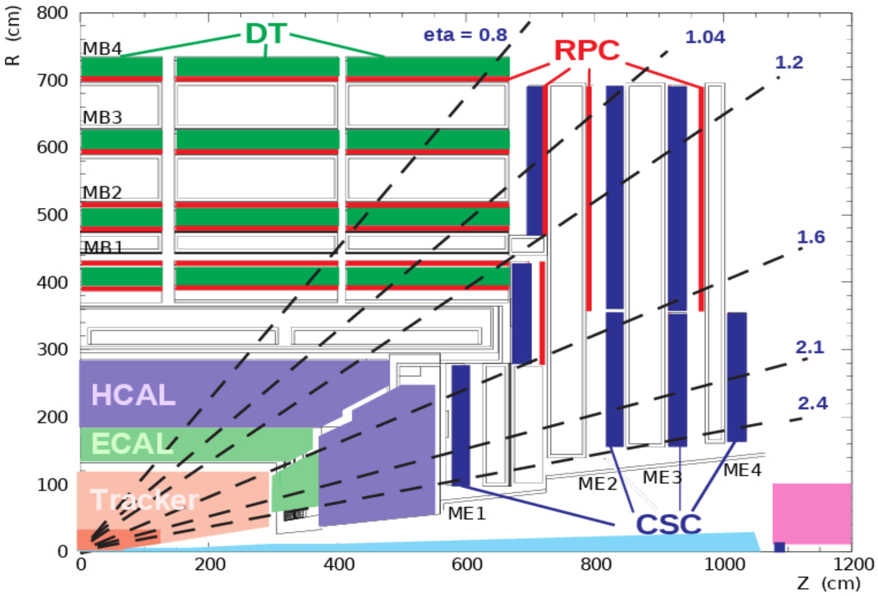


Figure 3.14: Geometrical distribution along the z-axis of the different muons chambers in **CMS!** [**?**].

DT!s (DT!s)

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 **DT!s** 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 θ 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 **DT!**s is represented in Figure **??**.

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 in such a way that the maximal drift of any charge is lower than 21mm, corresponding to a drift time of 380 ns in the gaseous chambers made out of 85% of Ar and 15% of CO_2, a value small enough to produce negligible occupancy in the different wires and to avoid the need of multi-hits electronics. Redundancy of the **DT!**s 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.

CSC!s (CSC!s)

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 **CSC!s**, multi-wire 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 **DT!**s and **CSC!**s while in the $1.2 < |\eta| < 2.4$ area, muons typically cross between 3 and 4 **CSC!**s 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 5000 m^2 and the total gas volume included in such chambers is equal to about 50 m^3.

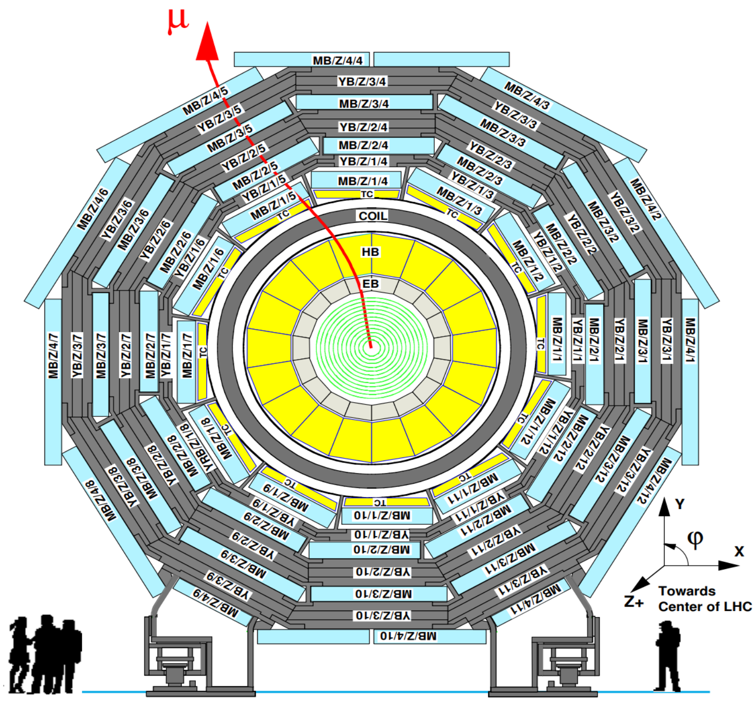


Figure 3.15: Geometrical division of the **DT!**s chambers in one of the 5 wheels of **CMS!** [?].

RPC!s (RPC!s)

Finally, some **RPC!s** 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 **DT!**s of 380 ns is way larger than the bunch spacing in the **LHC!**, while a **RPC!** is capable of tagging an ionizing event in less than 25ns, making it an ideal candidate to trigger the event.

The **RPC!**s are double-gap chambers operated in avalanche mode to ensure good operation, even 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 **DT!**s or **CSC!**s. The **RPC!**s are also useful in the sense that they can help to resolve ambiguities when attempting to construct tracks from multiple hits observed in a chamber.

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, independently of the rest of the detector and in addition of being able to detect, identify and measure several properties of muons crossing it. In summary, the different features of the three muon sub-systems used by the **CMS!** detector are summarized in Table **??**.

The muon reconstruction efficiency obtained by the muon system strongly depends on the pseudo-rapidity value of the muon considered, along with its transverse momentum, as shown in Figure **??**. 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 from the muon system and the **global muons**, defined using both the information coming from the muon system and the tracker.

Taking advantage of the **LS!**2, a new muon system has been installed in the experimental cavern: the so-called **GEM!s (GEM!s)**, placed in the endcaps, where the radiation and event rates are the highest. This new sub-detector will provide additional redundancy and measurement points to

Muon sub-system	**DT!**	**CSC!**	**RPC!**
$\lvert\eta\rvert$ 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)
Spatial resolution	80-120 μm	40-150 μm	0.8-1.2 cm
Average efficiency (13 TeV)	97.1%	97.4%	94.2% (barrel) 96.4% (endcaps)

Table 3.2: Comparison of the three main sub-systems currently used by **CMS!** in order to identify and measure muons [**?**].

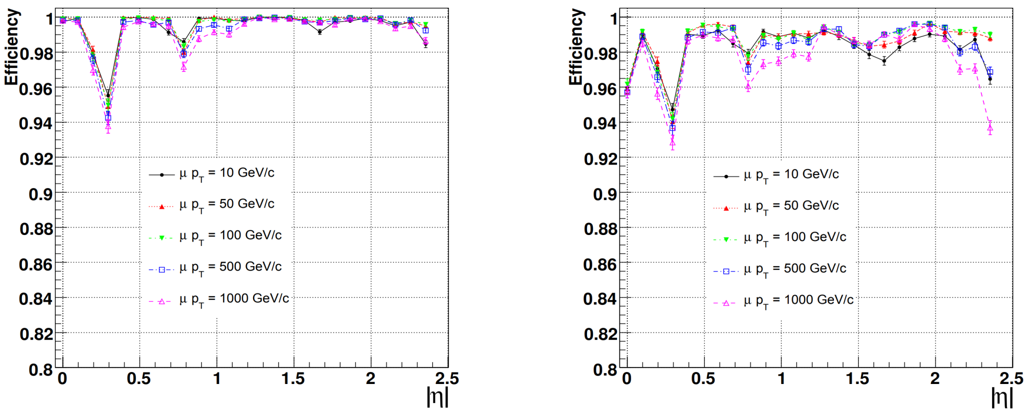


Figure 3.16: Muon reconstruction efficiency with different p_T and η values, considering only the muon system (on the left), and the combined information from the muon system and the tracker (on the right) [**?**].

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_2 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 **??**), while the rest will be set up during the next **LS!** expected in 2024, before the phase II of operation of the **LHC!**.

3.2.6 Trigger system

The **CMS!** experiment faces a data acquisition limitation 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 1 kHz only, already more or less equivalent to 1 Gb 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.

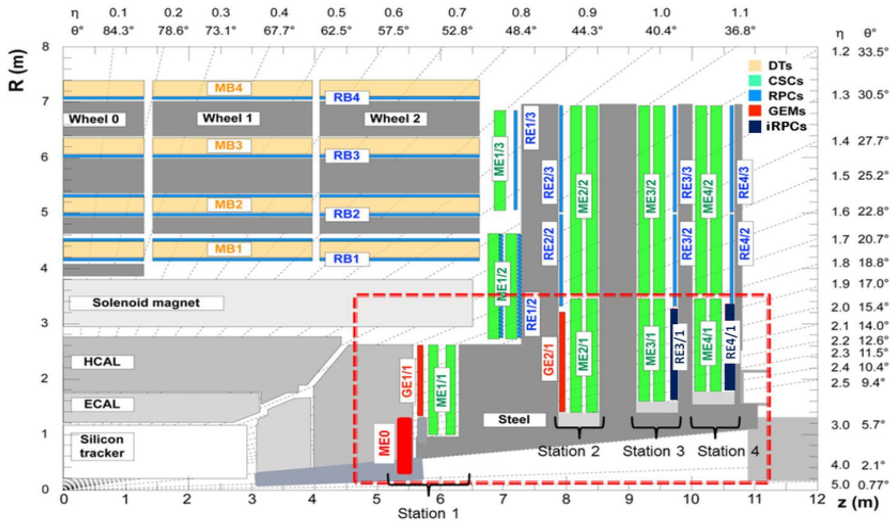


Figure 3.17: Location of the new **GEM!** muon subsystem currently being installed in the very-forward region of **CMS!** [?].

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 40 MHz. This system is based on two different levels: first of all, the **L1!** (**L1!**), a hardware set of electronics selecting around 100 kHz of data, followed by the **HLT!** (**HLT!**), a software layer refining even more the selection.

L1! (L1!)

The **L1!** is the first level of trigger [?], based directly on hardware. In order to maximize its efficiency, it is mostly 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.

The L1 trigger, whose architecture is represented in Figure **??**, has local, regional and global components. First of all, the local triggers, also referred to the **TPG!** (**TPG!**), are using the energy deposits measured in the calorimeter and the track segments or hit patterns in the muon chambers. Then, regional triggers combine this information and use pattern logic to determine ranked and sorted trigger objects such as electron or muon candidates in limited spatial regions. The rank is typically determined as a function of energy or momentum and quality, which reflects the level of confidence attributed to the L1 parameter measurements, based on detailed knowledge of the detectors and trigger electronics and on the amount of information available. Finally, the global calorimeter and muon triggers combine this local information to determine the highest-rank calorimeter and muon objects across the entire experiment and transfer them to the the top entity of the L1 hierarchy, the global trigger. This level is in charge of taking the final decision on whether to keep or reject the event, and eventually passes it to the **HLT!**.

This trigger 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 system 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.

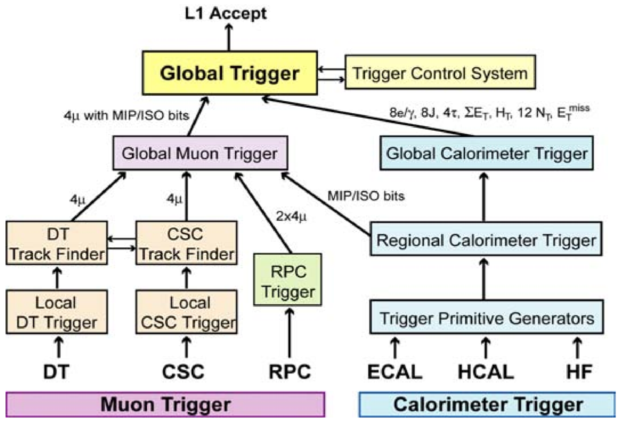


Figure 3.18: Architecture of the **L1!** trigger of **CMS!** [**?**].

HLT! (HLT!)

On the other hand, the **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 off-line. Since this is a software based layer and the decision time is not as critical, 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!** uses the so-called **trigger paths** in order to select specific particle topologies in the collisions. This selection imposes some thresholds on key quantities of the different objects, such as their transverse momentum p_T, to avoid passing the bandwidth limit of 1 kHz. In some cases, not all the events of a given category can be collected and only a fraction of them are therefore recorded. Trigger paths designed to behave like this are called pre-scaled trigger paths.

3.2.7 DAQ! (DAQ!)

In summary, the CMS **DAQ!**, whose global architecture is represented in Figure **??**, has been designed to collect and analyze the data information at the nominal collision rate of 40 MHz and is fed directly from the **L1!** trigger. This means that it has to be able to read a flux of data of the order of 100 kHz ($\sim$ 100Gb/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.

The **DAQ!** is indeed also in charge of performing additional tasks, such as the generation of the **DQM!** (**DQM!**) information resulting from on-line event processing in the **HLT!**, the transfer of the data from local storage at the **CMS!** site to mass storage in the **CERN!** data center at the Meyrin site and the operation of the **DCS!** (**DCS!**) system, ensuring the correct operation of the detector and a high quality data taking at all times.

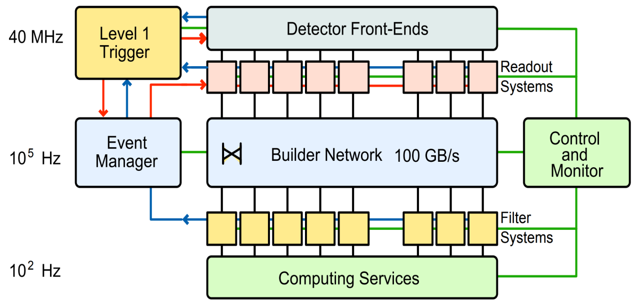


Figure 3.19: Architecture of the **CMS! DAQ!** system [**?**].

3.3 CMS! main goals

At the end of the day, the **CMS!** detector has been built in order to meet the goals of the **LHC!** physics program, which have been summarized in [**?**] as:

- A 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 momentum $p < 1$ TeV;
- A good charged-particle momentum resolution and reconstruction efficiency in the inner tracker, and an efficient triggering and off-line tagging of taus and b-jets, requiring pixel detectors close to the interaction region;
- A good electromagnetic energy resolution, good diphoton and dielectron mass resolution ($\sim$ 1% at 100 GeV), wide geometric coverage, π^0 rejection, and efficient photon and lepton isolation at high luminosities;
- And a good missing-transverse-energy and dijet-mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

CHAPTER 4

Event reconstruction

As we just saw, the **CMS!** detector is made out of different layers, each able to convert the interaction of the particles into electronic signals that can be measured and stored. However, such recorded signals contain only low-level, partial information about the particles, in such a way that an algorithmic strategy is needed in order to fully reconstruct physical information such as the number of particles and their charge, momentum and direction. Producing this kind of data is essential for all the off-line analyses performed, 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 compute and produce useful kinematic variables and physical objects (such as leptons and jets) is the so-called **PF!** (**PF!**) algorithm [**?**], which will be first of all described in Section **??**, along with the method used to identify unequivocally the primary vertex of the interaction in Section **??**. Then, a particular focus will be given to the definition and reconstruction of different objects of our particular analysis, such as electrons and muons (Section **??**), jets (Section **??**), the **MET!** (Section **??**) and the reconstruction of the $t\bar{t}$ system (in Section **??**).

4.1 PF! (PF!) algorithm

The **PF!** is an algorithm aiming to combine in the best way possible all the information coming from the different parts of the **CMS!** detector (mostly tracks and clusters of energy) 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, photos and hadrons. Then, more complex calculations are performed to identify eventual unstable particles, jets coming from the hadronization of quarks and gluons, and to compute complex variables such as the leptons isolation and the **MET!**.

The most basic elements used by this algorithm for the reconstruction of high-level physics objects are the hits left by charged particles in the tracker and in the muon chambers, and the clusters of energy left in the two calorimeters. For this algorithm to be as efficient as possible, the detector has been carefully designed, as described in Chapter **??**: a magnetic field as intense 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 muon system has been carefully designed as well and, in general, the whole detector is obviously as hermetic as possible.

The way the different particles produced in each collision are identified is quite easy to summarize and is represented in Figure **??**. Basically, the different kinds of particles produced interact with different parts of the detector and the combination of this information then allows to unequivocally identify each particle. The **PF!** algorithm starts from the collection of tracks and vertices reconstructed by the dedicated tracking and vertexing algorithms and from the local reconstruction of energy clusters in the calorimeters. The sequence follows a specific order to make this reconstruction process as efficient as possible:

1. First of all, the **PV! (PV!)** is identified by taking into account the **PU!** and by assigning the different tracks to the different *pp* collisions happening during a single bunch-crossing. All the tracks originating from less energetic vertices or from a secondary vertex of interaction are typically ignored and the corresponding hits in the tracker left by such particles can therefore be ignored later on, leaving less hits available for the clustering algorithm, 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 outer parts of the detector. Depending on the version of the algorithm used, each muon identified can be associated to its track in the tracker, where all the hits matching a muon can therefore be subsequently ignored to simplify the following reconstruction steps.

3. **Electrons** do have a charge, so they are visible by the tracker, and can interact electromagnetically with 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 measure efficiently and precisely their energy while avoiding any double counting. All the tracker hits corresponding to electrons are also ignored for the rest of the reconstruction 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 fourth step, using the last 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 therefore 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 more detail the reconstruction method applied in order to reconstruct the main objects of this analysis, i.e. the leptons, jets, **MET!** and bottom/top quarks.

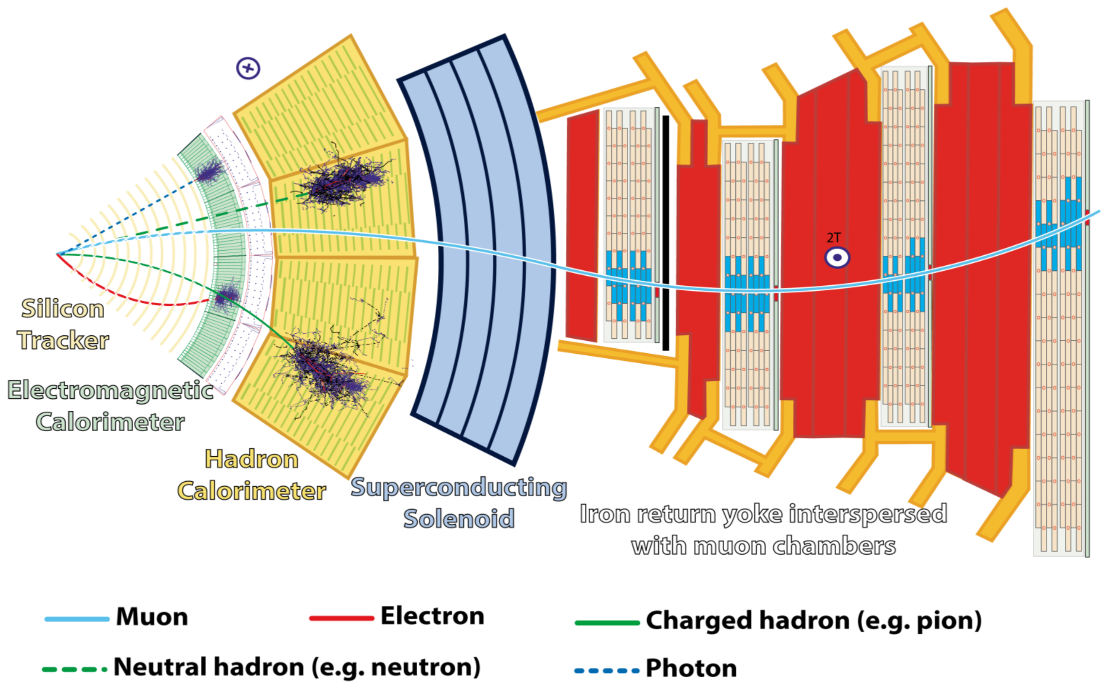


Figure 4.1: Transverse section of **CMS!** showing the different interactions expected by different kinds of particles in the detector.

4.2 Primary vertex definition

Different kinds of vertices originating from a single *pp* collision can usually be defined, as shown in Figure **??**. First of all, we have the set of **PU! vertices**, corresponding to the different simultaneous collisions of a single bunch-crossing of the **LHC!** or eventual remnants coming from a previous collision (the out-of-time **PU!**). The **PV! (PV!)** is then defined as the most energetic **PU!** vertex and the only vertex considered in most of the physics analyses and finally, we have the **secondary vertices**, mainly due to the eventual presence of hadrons living long enough to travel a significant distance in the detector, decay chains or jets.

The **PF!** algorithm is not able to reconstruct the vertices itself, but it can use the information coming from a dedicated vertexing algorithm able to identify all these kinds of vertices. This is done by considering all the tracker hits observed, 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 p_T^2 summed over all the physics objects of the event is then assumed to be the **PV!**, as it is considered to be the origin of the collision that actually fired the trigger, from which many different tracks are emitted.

4.3 Leptons reconstruction

Different kinds of leptons are typically produced by a *pp* collision. The muons, electrons and their respective anti-particles can be quite easily identified, mainly because their lifetime and velocities are high enough 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

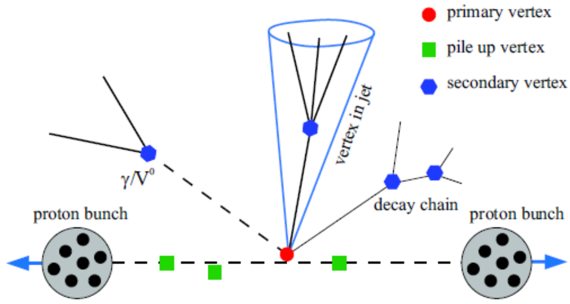


Figure 4.2: Different kinds of vertices typically observed in a pp collision in the **LHC!**.

the time to hadrons. However, our analysis does not consider taus directly but only the leptons originating from their decays, making the details regarding their reconstruction irrelevant.

4.3.1 Muons

Muons are the first leptons to be reconstructed by the **PF!** algorithm since, by design and at first order, they are the only particles expected to reach the muon chambers, resulting in an easier and more efficient identification.

The typical signature of a muon consists of several hits in the silicon tracker forming a track 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 to the **PF!** algorithm, then able to reconstruct different kinds of muons [**?**].

Standalone muons

The standalone muons 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 **DT!**s, **CSC!**s and **RPC!**s and tries to reconstruct a vector of trajectory in each case using a **KF!** (**KF!**) [**?**]. The segments obtained are then combined in the best statistical way in order to form a candidate track for each muon of the event, by extrapolating the innermost vectors obtained to the next chamber and by comparing it with the local track segment; the trajectory parameters are then updated and the process continues until reaching the outermost chamber. Finally, this process is repeated in the reverse order, from the outside chambers to the inside, to estimate the innermost track parameters as well.

It is important to note that cosmic rays reaching the Earth every day are an important source of contamination, since we estimate that around 10.000 muons per m^2 and per minute coming from such processes can be observed at sea level. Putting the detector underground removes part of this contamination but, to further limit the possibility of misidentification due to showering of cosmic rays, the tracks need to pass some quality criteria: for example, checking the extrapolation of the trajectory to the point of closest approach to the beam line allows to reduce this contamination. Additionally, at least two hits need to be measured for the fit to be performed, one of them coming from either the **DT!**s or **CSC!**s, in order to remove fake segments contamination due to combinatorics.

In any case, 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 the so-called tracker muons on the other hand is able to propagate tracks identified in the inner silicon tracker (having momenta $p > 2.5$ 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*). An extrapolated track and a segment are only matched if the difference between their positions in the x coordinate is smaller than 3 cm or if the pull, the ratio of this distance to its uncertainty, is smaller than 4.

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 muon 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 by 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 for $p_T > 200$ 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, as seen in Figure **??**.

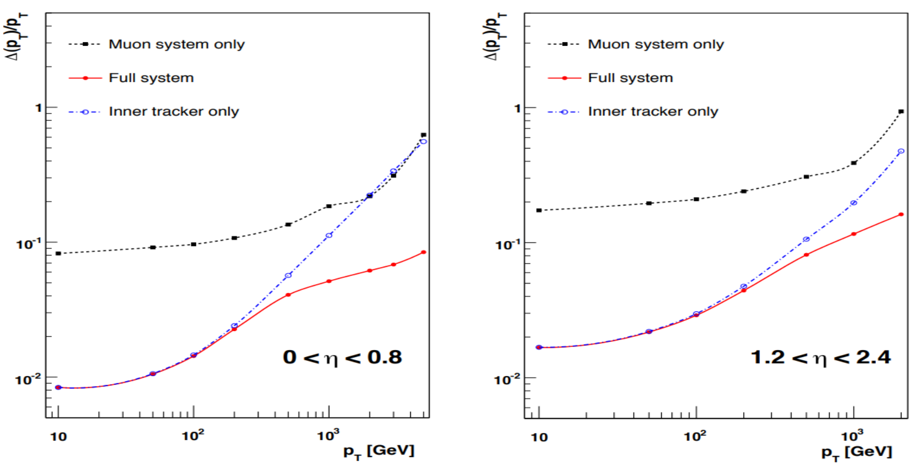


Figure 4.3: Muon p_T resolution obtained in simulation in the barrel (on the left) and endcap (on the right) for different muon reconstruction algorithms [**?**].

Once reconstructed, muon 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 in the (η, ϕ) plane, as shown in Figure **??**) criteria in order to enhance the purity of the reconstructed prompt muons (muons directly originating from the main collision taking place in the event), by rejecting muons coming from the decay of heavy flavour quarks, typically surrounded by a large amount of hadronic activity. The calculation of the lepton isolation is typically performed considering only the **PV!** since higher levels of **PU!** are expected to bias this measurement by artificially increasing the hadronic activity.

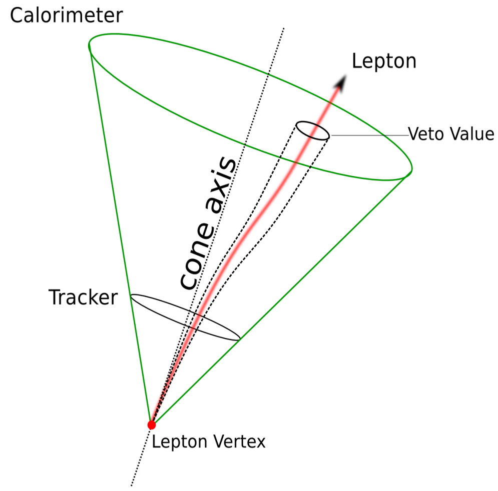


Figure 4.4: Lepton isolation cone typically used to enhance the prompt leptons purity.

Different identification **WP!**s (**WP!**s) can then be defined for the off-line analyses, from veto to tight, in order to reject more or less contamination from misidentified leptons, bearing in mind that a tighter selection is cleaner, but less efficient: the loose and tight **WP!** have then been defined in order to respectively achieve 95% and 98% efficiencies, respectively. These efficiencies are tuned using simulated $Z \to \mu^+\mu^-$ events with a $p_T > 20$ GeV, while the efficiency to reject muons in jets is done using simulated QCD and W+jets processes [**?**].

Our particular muon definition is based on the medium **WP!** provided centrally by the muon **POG!** (**POG!**), but has been slightly modified and made more robust against non-prompt leptons, according to the selection detailed in the Chapter **??**.

4.3.2 Electrons

Electrons and positrons 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 electrons themselves and the crystals composing this sub-detector.

It is usually a bit harder to reconstruct electrons than muons mainly because electrons do interact much more with the tracker and this interaction therefore needs to be modeled to understand the exact behaviour of such particles: it 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 [**?**]. In order to measure precisely the energy of an electron, all the photons emitted by bremsstrahlung before reaching the **ECAL!** (usually, along the ϕ plane 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 **SC! (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, as the energy deposited in the **ECAL!** is usually spread into several different crystals because of the electromagnetic shower effect 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 2σ of the electronic noise and matching some quality criteria ($E_{\text{seed}} > 230$ MeV in the barrel, $E_{\text{seed}} > 600$ MeV and $E^T_{\text{seed}} > 150$ 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, $E^T_{\text{cluster}} > 1$ GeV in the endcaps) [**?**]. 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!** is 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 [**?**] 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 is moving according to an helix in the magnetic field of the detector. This seeding is mostly optimized for isolated electrons 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 when considering electrons inside jets.

   Once the seeds have been identified, the identification of tracks can begin. First of all, the gathering of compatible hits from the set of available seeds is done using 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 the case of bremsstrahlung emission.

   Finally, once the hits are collected, a **GSF!** (**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 χ^2 penalty to each distribution, depending on several parameters, such as 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 next 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!**

SC!s previously built. This step is also designed to preserve the highest efficiency possible while keeping the misidentification probability low and eventual ambiguities related to single electron seeds which can often lead to several reconstructed tracks are also resolved at this stage.

4. Finally, a loose pre-selection 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) can be combined into a **MVA!** (**MVA!**) estimator allowing to define several electron **WP!**s as well.

This electron reconstruction workflow has been summarized in Figure **??**.

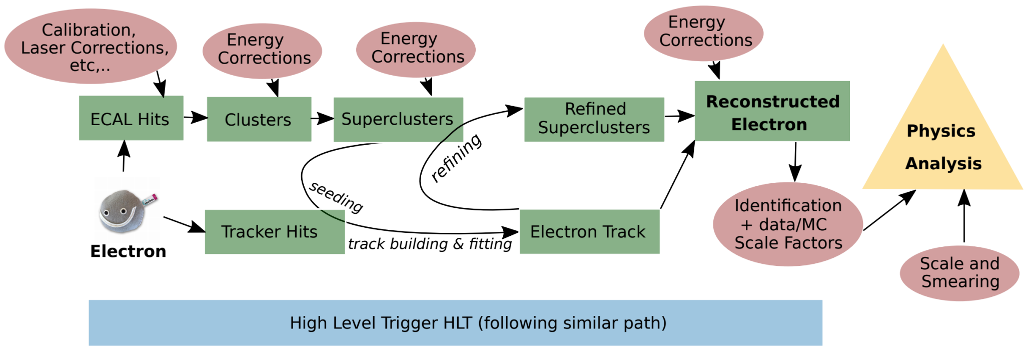


Figure 4.5: Schematic representation of the full electron reconstruction workflow in **CMS!** [?].

4.4 Jets reconstruction

Eventual quarks and/or gluons produced by pp collisions 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 experimental signature of the apparition of a quark is a shower 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 **??**.

Several algorithms can be used to reconstruct the jets by linking the information coming from the tracker and the calorimeters, but the most common 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 leptons produced and merge them into a single jet [?]. 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; in this case, other algorithms can be applied.

To perform such a job, sequential clustering algorithms 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 **??**, these distances can be computed using different variables such as $\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$, the distance

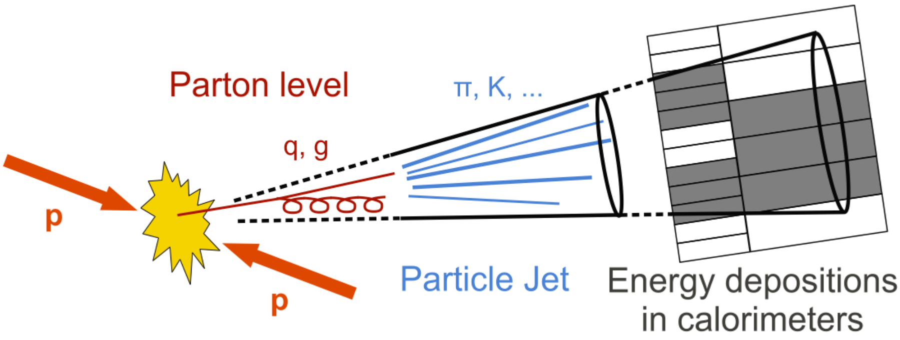


Figure 4.6: Schematic representation of the typical development of a jet within the **CMS!** detector.

between i and j in the (η, ϕ) space, the p_T^2 of each particle and the clustering algorithm radius parameter determining the final jet size and usually set to 0.4 by the **CMS!** collaboration. This distance parameter defines a cone in which the momenta of all the particles observed is summed to get the momentum of the jet itself.

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

The algorithm works by looking at all the i, j combinations, comparing the distances d_{ij} and d_{iB} until only jets and individual unclustered particles are present in the event:

- If d_{ij} is smaller than d_{iB}, 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 than d_{ij}, then i is considered to be the final jet and is therefore removed from the list of jet candidates as well.

Several corrections, mainly the so-called **JEC!**s (**JEC!**s) [**?**], 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 **PU!** effects and the dependence of the reconstruction on the jet flavor.

The efficiency of the **PF!** algorithm for jet identification and reconstruction has been checked using simulation, as shown in Figure **??**. 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 value of the jet p_T.

4.4.1 B-tagging

Jets coming from bottom quarks are usually interesting to study in many different fields of particles physics, such as 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 Chapters **??** and **??**.

This specific kind of jets can be distinguished from other jets because of the relatively long lifetime of the bottom quark that produces in the detector a secondary vertex displaced by up to a few

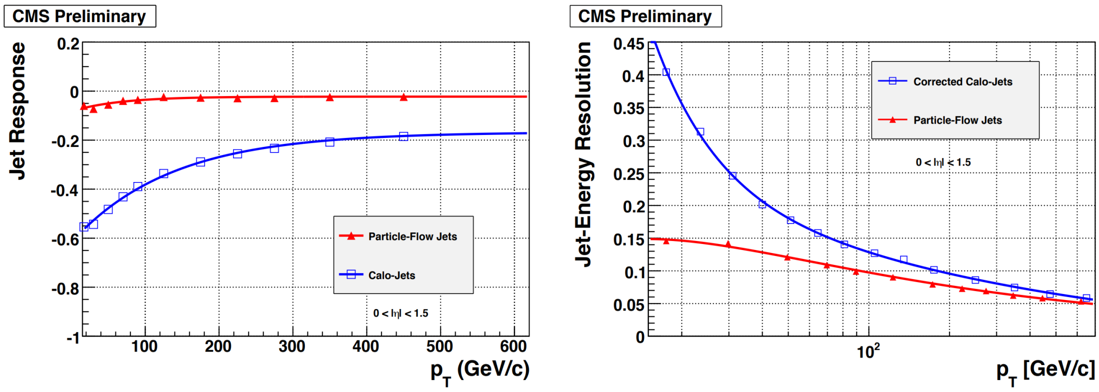


Figure 4.7: Comparison of the jet energy response (defined as the percentage of reconstructed energy, on the left) and jet energy resolution (on the right) for dijets simulated events in the barrel for jets reconstructed using only the calorimeters (in blue) and jet candidates from the **PF!** algorithm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

millimeters with respect to the **PV!**, as shown in Figure **??**; and this gives a perfect way to discriminate b-jets from other jets coming from light quarks. 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.

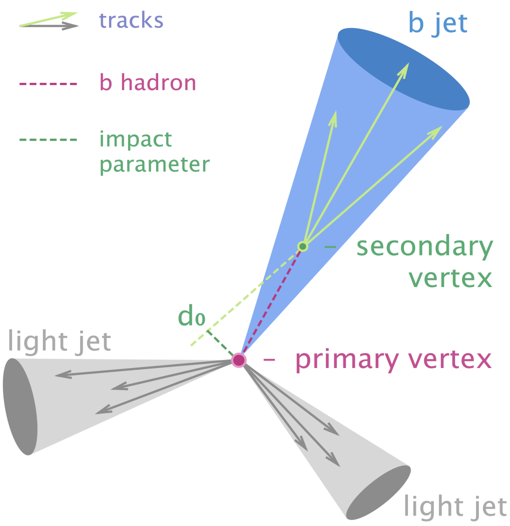


Figure 4.8: Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex.

Because of these specific properties, some algorithms can 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 keeping the risk of possible misidentification of a jet as low as possible.

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

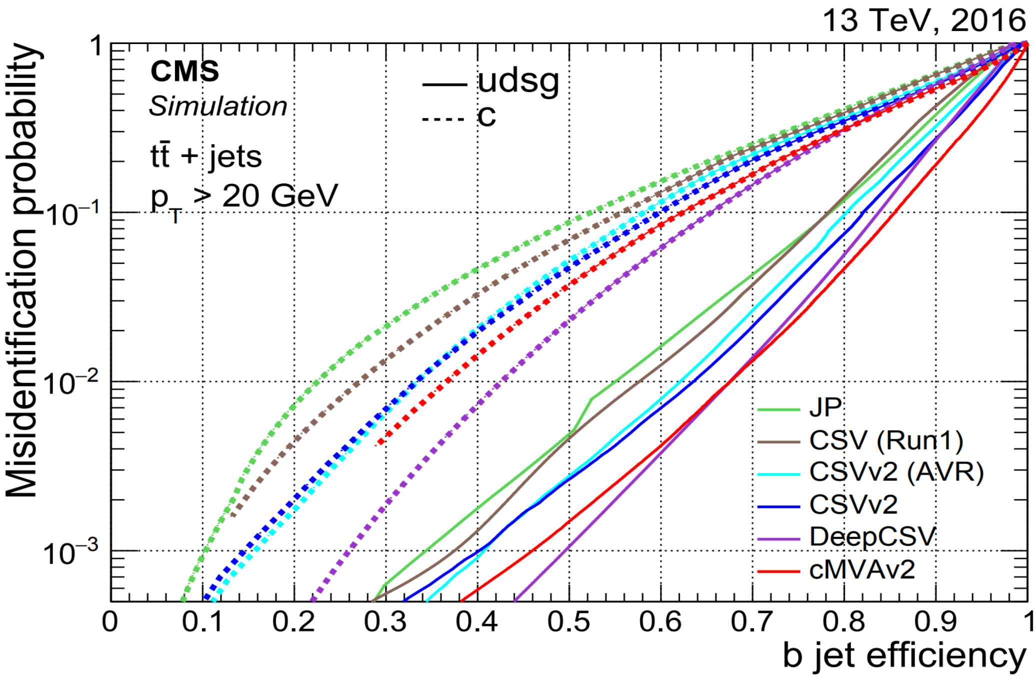


Figure 4.9: b-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [**?**].

Different **WP!**s are then also made available in order to obtain the desired combination between b-jet identification efficiency and misidentification rate. The loose, medium and tight b-jets **WP!**s have been defined in such a way that they limit this misidentification rate of a light jet as a b-jet to 10%, 1% and 0.1% respectively.

4.5 MET! (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 conversation principle. This statement is not totally true though since we are aware of several effects that could induce an imbalance in this transverse momentum:

- Even though the **CMS!** detector has been carefully designed to be as hermetic as possible, some particles could be created outside of its acceptance 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 residual energy while 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 and actually detected, leading to an instrumental contribution to the **MET!** in some cases;

- Some events also present an anomalous **MET!** measurement which can arise because of a variety of reconstruction failures or malfunctioning detectors [?];
- Finally, the eventual exotic weakly interacting particles produced, such as dark matter, are typically expected to leave some **MET!** in the detector as well.

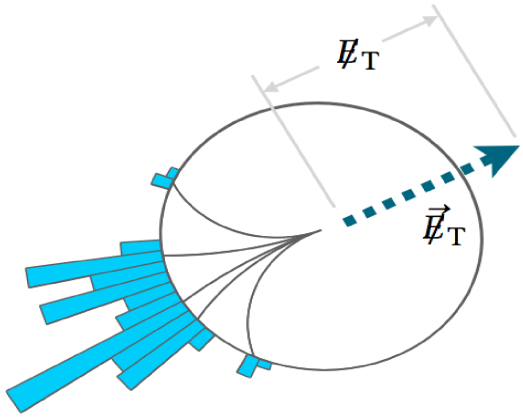


Figure 4.10: Schematic representation of the **MET!**.

The E_T^{miss} variable, defined in Equation **??** as the negative sum of the transverse momentum of all the particles j of the event, accounts for this eventual imbalance in the transverse momentum and is therefore a key variable in some of the analyses searching for new **BSM!** physics, typically producing new kinds of exotic particles not expected to interact with the detector.

$$E_T^{\text{miss}} = \vec{p}_T^{\,\text{miss}} = -\sum_j \vec{p}_{T,j} \tag{4.2}$$

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

- 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 performed nowadays, because of its simple, robust, yet very efficient estimate of the **MET!** spectrum;
- And the **PUPPI! (PUPPI!) MET!** [?], 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.

For this analysis, we decided to use the pfMET for simplicity after observing that using the **PUPPI! MET!** did not seem to give a better agreement between data and **MC!** (**MC!**) in most of the regions defined, as shown in Figure **??** in a 2018 $t\bar{t}$ region, especially in the most interesting regions for us, for high **MET!** values. In particular, the Type I corrected **MET!** is used in this work, replacing the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which **JEC!**s are applied [?].

Several corrections 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 due to several effects, such as the the electronic noise and eventual dead cells in the calorimeters or the presence of an eventual beam halo of particles from the **LHC!** itself, leading to a global miscalculation of the final energy of the event. These filters, detailed in Table **??**, are extremely important, especially in the tail of the **MET!** spectrum, as observed in Figure **??**.

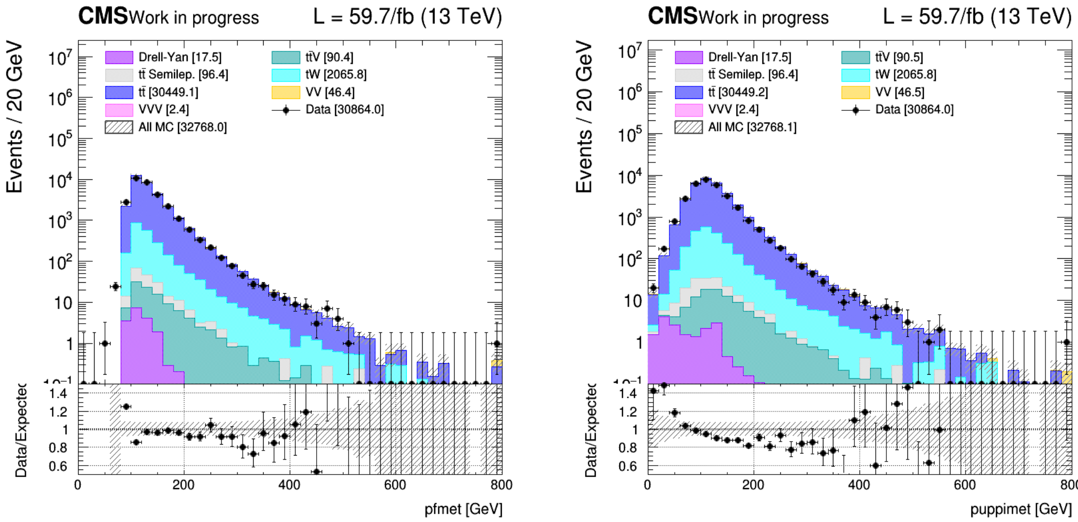


Figure 4.11: pfMET (on the left) and **PUPPI!** **MET!** (on the right) distributions observed in a general 2018 $t\bar{t}$ control region.

Filter name	Applied to data	Applied to **MC!**
Flag_goodVertices	✓	✓
Flag_globalSuperTightHalo2016Filter	✓	✓
Flag_HBHENoiseFilter	✓	✓
Flag_HBHENoiseIsoFilter	✓	✓
Flag_EcalDeadCellTriggerPrimitiveFilter	✓	✓
Flag_BadPFMuonFilter	✓	✓
Flag_ecalBadCalibFilterV2†	✓	✓
Flag_eeBadScFilter	✓	–

† applied only to 2017 and 2018.

Table 4.1: **MET!** filters applied to events selected in data and to simulated events, an hyphen (–) indicating the filter is not applied.

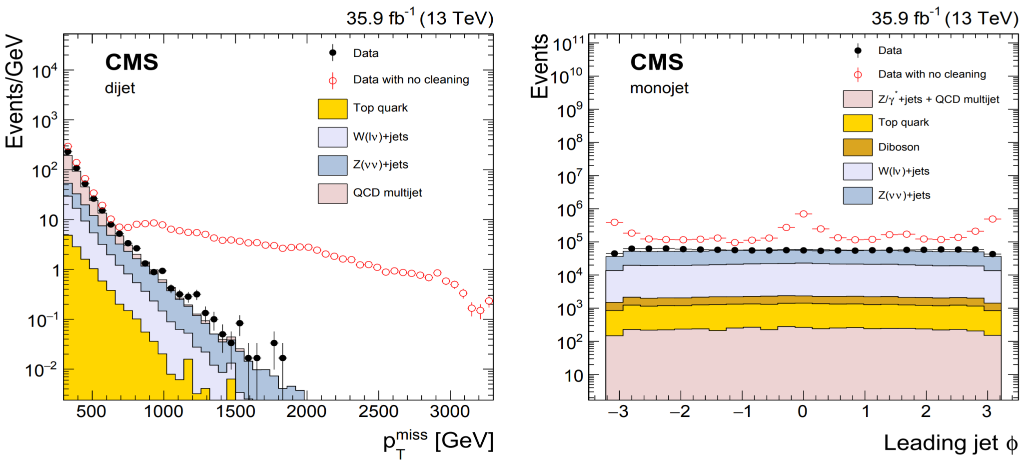


Figure 4.12: **MET!** (on the left) and jet (on the right) ϕ distributions with and without **MET!** filters applied [**?**].

4.6 Top reconstruction

Although not formerly a part of the **PF!** algorithm and done off-line, the kinematic reconstruction of the $t\bar{t}$ system is still an extremely important part of this analysis. Indeed, many variables allowing us to discriminate the signals we are interested in from the different background processes are spin correlated variables, sensitive to the nature of the dark matter mediator and which require knowledge of the top quark and anti-quark four-momenta to be computed, not immediately available without such a complete reconstruction of the $t\bar{t}$ system.

However, this reconstruction from channels containing two leptons is typically quite challenging, given the fact that the neutrinos produced at the end of the decay chain of the $t\bar{t}$ cannot be directly observed. This means that the only observable information we have regarding these neutrinos is the vectorial sum of their transverse momentum, which can be only inferred from the total momentum imbalance of the event and which frequently comes with a bad resolution. In this sense, the determination of the individual momenta of each neutrino requires advanced computation techniques, either numerical or analytical.

4.6.1 Numerical and analytical top reconstruction

Two main methods exist in order to reconstruct the $t\bar{t}$ system:

- The **Sonnenschein numerical method** [?] first of all relies on the kinematics of the system and on the expression of the four-momenta of the different particles involved in the top quark decay chains (in this case, we consider only the chain $t \to bW \to bl\nu$ for the two tops, as previously explained), as expressed in Equations **??** to **??**, if we assume that the **MET!** of the event is coming only from the two neutrinos produced.

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

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

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

  In this case, we therefore have 6 equations to solve and exactly 6 unknowns (since the energy of the neutrinos is considered equal to their momentum because of their extremely low mass, and if we assume the W boson and top quark masses to be known and fixed) corresponding to the three momentum components of each neutrino produced, a problem that can in principle be solved. This leads to a quartic equation in p_{ν_x}, analytically solvable but quite ambiguous given the variable number of solutions of such equation (typically, the solution giving the lowest invariant mass for the $t\bar{t}$ system is then chosen).

- The **Betchart analytical method** [?] on the other hand is able to describe the decay

$t \to bl\nu$ using this time a geometric approach. This method was chosen in this analysis because it offers the following advantages:

- With this method, the invariant mass constraints from the top quark and the W boson are both exact and do not suffer the same kind of ambiguity as observed previously;
- The solution set for each neutrino momentum in this case is an ellipse in the (x, y) plane that can be described precisely and from which the solutions for the neutrino momentum can be reduced to a discrete set of values, which will be helpful when trying to estimate the p_T of the mediator of the **DM!** signals considered, while also giving us information about the precision of this measurement, as we will see later;
- The results obtained here can be useful for other event topologies featuring similar kinematic constraints as well.

Basically, this method relies on two observations constraining the geometrical shape of the W boson momentum vector. First of all, the decay of a top quark constrains this vector to an ellipsoidal surface of revolution around an axis coincident with the bottom quark momentum. The decay of the W boson itself on the other hand additionally constrains this vector to another ellipsoidal surface of revolution around an axis matching the momentum of the resulting charged lepton. The W boson momentum vector will then be defined by the intersection of the surfaces given by these two constraints, resulting in an ellipse in the (x, y) plane. The neutrino momentum vector can then be expressed as a translation of this ellipse, using a parametric expression, as described in [**?**].

In the two neutrinos final state, it is then possible to show that the elliptical solution sets for the neutrino momenta $(\nu_\perp,\ \bar{\nu}_\perp)$ respective to the two top quarks decaying to leptons are given by Equation **??**, where $N_\perp$ and $\bar{N}_\perp$ are the solution ellipses of the (anti)neutrino in the transverse plane and expressed in the laboratory coordinate system.

$$\begin{cases} \nu_\perp^T N_\perp \nu_\perp = 0 \\ \bar{\nu}_\perp^T \bar{N}_\perp \bar{\nu}_\perp = 0 \end{cases} \tag{4.4}$$

Given the fact that the measured components $(\not{x},\ \not{y})$ of the **MET!** are the sum of the $\nu_\perp$ and $\bar{\nu}_\perp$ components, they can be related by Equation **??**.

$$\bar{\nu}_\perp = \begin{pmatrix} -1 & 0 & \not{x} \\ 0 & -1 & \not{y} \\ 0 & 0 & 1 \end{pmatrix} \nu_\perp \equiv \Gamma \nu_\perp \tag{4.5}$$

The solutions for the momenta of the neutrinos will then be given by the intersections of these two ellipses giving either zero, two or four solution pairs $(\boldsymbol{p}_\nu, \boldsymbol{p}_{\bar{\nu}})$, as shown in Figure **??**. If the two ellipses do not intersect, a χ^2 method can be used to check the compatibility between the best solution obtained and the standard $t\bar{t}$ process, the best solution being defined as the point of closest approach between the ellipses.

So far we have seen two methods able to reconstruct the individual momenta of both neutrinos. However, both these methods usually assume that the **MET!** is only coming from these neutrinos and this assumption is no longer verified for our signals, for which the **DM!** particles produced will contribute by a significant amount to the global **MET!**, resulting in a slightly lower reconstruction efficiency compared to the one obtained considering only the standard $t\bar{t}$ process. The method used then needs to be slightly adapted to our particular case.

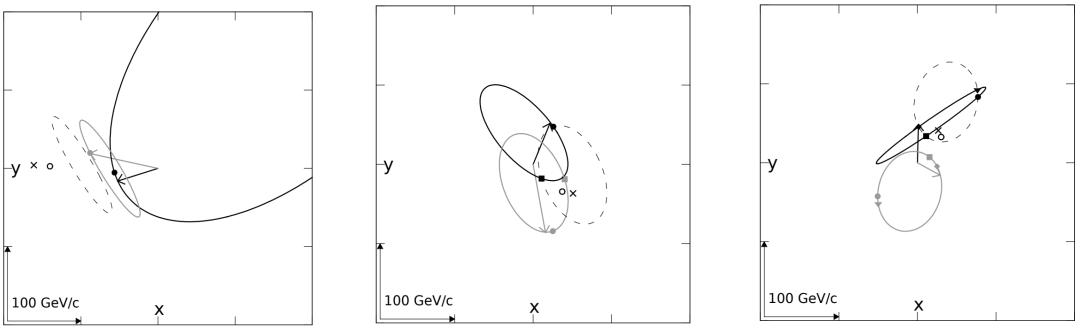


Figure 4.13: Three events constraining the neutrino and antineutrino momenta (black and grey arrows, respectively) resulting in 0 (on the left), 2 (on the center) or 4 (on the right) solutions. The dashed ellipse is obtained by using the additional constraint according to which the measured **MET!** is equal to the sum of neutrino transverse momenta [?].

4.6.2 Top reconstruction with additional dark matter

In this particular case, including an additional contribution $\boldsymbol{\phi}_\perp = (\phi_x, \phi_y, 0)^T$ to the **MET!** is equivalent to slightly modifying the Equation **??** to obtain Equation **??**.

$$\bar{\nu}_\perp = \begin{pmatrix} \bar{\nu}_x \\ \bar{\nu}_y \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & \cancel{x} \\ 0 & -1 & \cancel{y} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_x + \phi_x \\ \nu_y + \phi_y \\ 1 \end{pmatrix} \equiv \Gamma(\nu_\perp + \phi_\perp) \tag{4.6}$$

By substituting these values into Equation **??**, we then get a new relation between the two ellipses for the neutrino, $N_\perp$, and for the antineutrino, $\bar{N}_\perp$, given by Equation **??**.

$$\nu_\perp^T \Gamma^T \bar{N}_\perp \Gamma \nu_\perp + \nu_\perp^T \Gamma^T \bar{N}_\perp \Gamma \phi_\perp + \phi_\perp^T \Gamma^T \bar{N}_\perp \Gamma \nu_\perp + \phi_\perp^T \Gamma^T \bar{N}_\perp \Gamma \phi_\perp = 0 \tag{4.7}$$

This modification does make the reconstruction a bit more complicated by adding crossed terms between $\boldsymbol{\nu}_\perp$ and $\boldsymbol{\phi}$, which modify the phase space of solutions. Even though the reconstruction gets more complex in this case, performing it is extremely important because it provides us a way to determine two excellent discriminating variables:

- First of all, the so-called **dark** p_T is a variable estimating the value of the p_T of the eventual mediator of the interaction. If the system considered does not admit any solution, then there is no intersection between the ellipses. However, in this case, the distance between the ellipses gives us a way to approximate by how much we missed a perfect standard $t\bar{t}$ reconstruction, and it is actually possible to show that the distance between the center of both ellipses is related to the p_T of the mediator of the interaction [?]. This method being completely analytical, estimating the distance between the ellipses is trivial and is giving us the first background-signal discriminating variable that will be developed later in Section **??**;
- The second interesting variable is the so-called **overlapping factor** R, defined in Equation **??**, where l_1 and l_2 are the two sizes of the ellipses measured along the axis joining their centers, and d is the distance between these centers.

$$R = \frac{l_1 + l_2}{d} \tag{4.8}$$

By its definition, this factor is able to take into account not only the distance between these two ellipses but also their respective sizes and this is extremely important since, at the end of the day, we want an event having small ellipses far away from each other to have a much higher weight than an event featuring two incredibly large ellipses and therefore less significant when trying to distinguish the background and signal processes. Both these extreme cases are represented in Figure **??**.

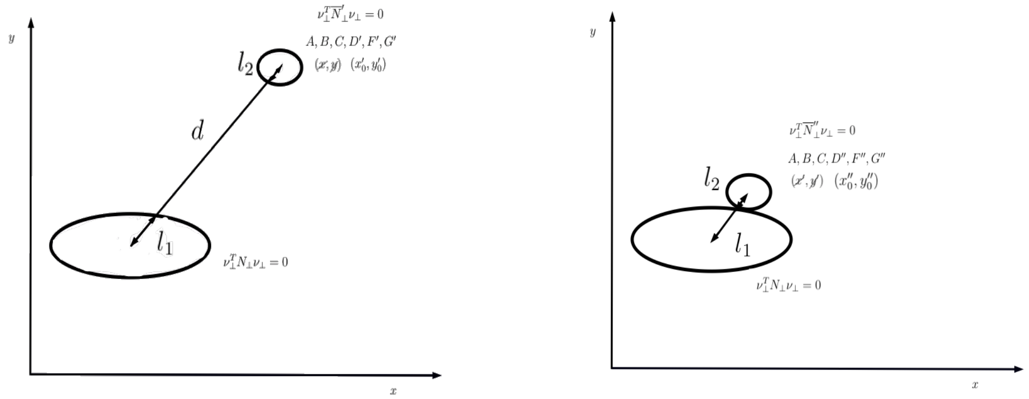


Figure 4.14: Schematic representation of the two extreme cases that can be observed when defining the overlapping factor: the reconstruction of a system with (on the left) and without (on the right) the presence of **DM!** [**?**].

4.6.3 Top reconstruction in practice

The idea behind the algorithm is quite simple: it basically take as input the 4-momenta of the two leptons, the two b-jets, and the two components of the transverse missing momentum and performs all the calculations just described while returning the optimal momenta value for the two neutrinos. However, several complications quickly appears when solving this problem in practice.

First of all, as we already saw, when the two ellipses intersect each other, we can observe either 2 or 4 solutions, even though only one actually corresponds to the physical momenta of the neutrinos. The solution giving the lowest possible invariant mass for the $t\bar{t}$ system is then simply chosen and taken in consideration for the analysis since this optimizes the probability of making the right choice: indeed, it has been shown [**?**] that in 85% of the cases, the solution satisfying this requisite actually matches the true simulated kinematics.

Then, even though the process studied is always the same and is supposed to give the same final state, each recorded event typically comes with a different number of leptons, jets and b-jets, mostly due to mismeasurements and b-tagging inefficiencies. This means that this algorithm needs to be applied several times, considering all the possible leptons and (b-)jets combinations of the event, taken two by two. In practice, the following categories are therefore defined:

- If exactly 0 b-jets are observed, the event is not considered in this analysis, according to the selection of the signal regions chosen;
- If exactly 1 b-jet is observed, then it is kept and used as input, while all the non b-tagged jets are considered as the second b-jet candidate. All the possible combinations between the two leptons and this set of jets are then tested one by one;
- Finally, if 2 or more b-jets are observed, all the combinations between the two leptons, the first b-jet (ordered in p_T) and the other b-jets of the event are considered.

When taking into account all these possible combinations, a reconstruction efficiency of $\sim 71\%$ has been achieved when considering standard $t\bar{t}$ **MC!** samples.

Finally, we know that any measurement made is not perfect and comes with uncertainties. The impact that imperfectly measured kinematic variables can have on the top reconstruction process can be estimated through the **smearing method**, by repeating the reconstruction 100 times for each combination previously defined, while updating in each iteration several parameters:

- The energy of the jets are updated within their respective uncertainties, using a random number drawn from a Gaussian distribution whose mean value μ is null and whose standard deviation σ has been estimated to 0.3 using the jet energy corrections information;
- The energy of the leptons is also updated using a random correction factor generated from the distribution shown in Figure **??**, computed in **MC!** and representing the energy deviation between the generation and reco information;

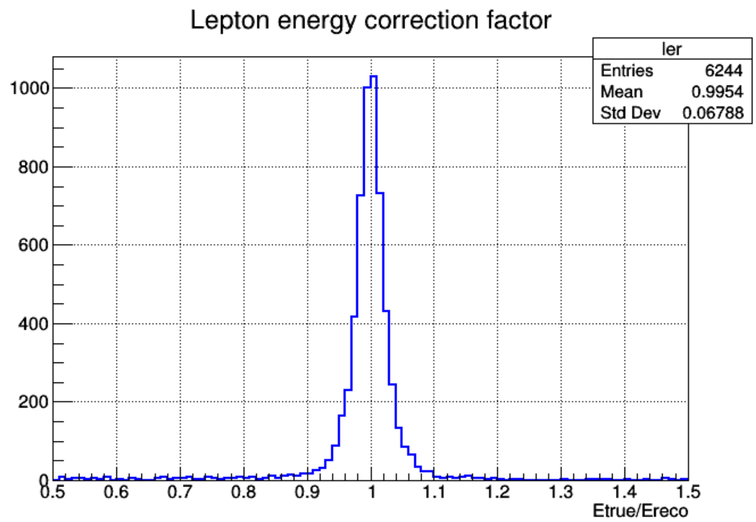


Figure 4.15: Definition of the correction factor used in the smearing of the leptons energy [**?**].

- The **MET!** is also recomputed each time using both these newly calculated parameters, according to Equation **??**;

$$\not{E}_{T_{x,y}}^{\text{updated}} = \not{E}_{T_{x,y}}^{\text{reco}} + \sum_{\text{jet}_i=1}^{2} \Delta(p_{x,y}^{\text{jet}_i}) + \sum_{\text{lep}_i=1}^{2} \Delta(p_{x,y}^{\text{lep}_i}) \tag{4.9}$$

- An angular smearing is also performed, using the angles α and ω shown in Figure **??**.

  The ω angle is taken from a simple uniform distribution $[0,\, 2\pi]$ while the angle α is taken as a random number generated from the distributions of the angle between the particle and detector level direction, as shown in Figure **??**. These distributions have been obtained in **MC!** using Equation **??**, relating the generation and reco normalized vectors $\hat{p}$ representing the direction of the momentum of a given object (lepton or b-jet);

$$\begin{cases} \alpha = \arccos(\hat{p}^{\text{reco}} \cdot \hat{p}^{\text{gen}}) \\ \hat{p} \equiv \dfrac{\overrightarrow{p}}{|\overrightarrow{p}|} = (\cos(\phi)\cdot\sin(\theta), \sin(\phi)\cdot\sin(\theta), \cos(\theta)) \end{cases} \tag{4.10}$$

- Finally, the mass of the W boson, entering the reconstruction as a parameter and not a

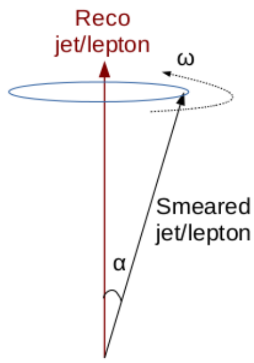


Figure 4.16: Graphical definition of the two angles α and ω used in the smearing of the leptons and jets directions [?].

universal constant in nature, is updated each time by generating a random number according to a Breit-Weigner distribution of mean 80.38 GeV and width 2.08 GeV.

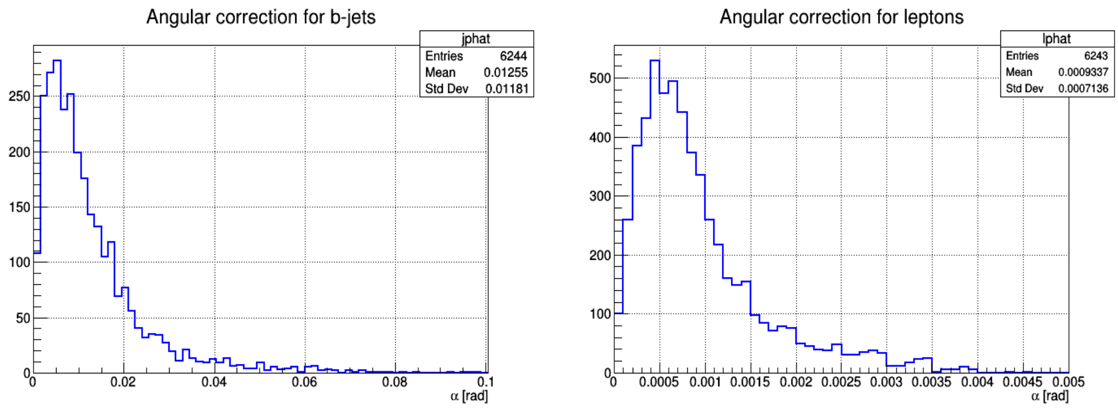


Figure 4.17: Simulated distributions of the α angle between the particle and detector level direction for b-jets (on the left) and leptons (on the right).

During each iteration of the smearing process, an individual weight w is computed from the true invariant mass m_{lb} distribution previously obtained using generation. Both the W mass and true m_{lb} distributions obtained are represented in Figure ??.

A global weight W is then assigned to each lepton/b-jet combination by summing the weight $\sum w(m_{l\bar{b}}) \cdot w(m_{\bar{l}b})$ obtained for each smearing iteration, and the combination with the largest weight is considered to be the solution of the event. This weight distribution will be another variable used to discriminate the signal from the backgrounds processes, since the higher this variable is, the larger the likelihood for the event to be an authentic $t\bar{t}$ event is.

Even though this smearing increases the efficiency of the reconstruction to $\sim 90\%$ when considering the $t\bar{t}$ process, this kinematic reconstruction algorithm does not allow us to find a solution to all the events. If an event does not present any solution, then all its kinematic variable which do not depend on the neutrino momenta are still considered for the analysis, while the variables which do require the reconstruction information are set to non-physical default values.

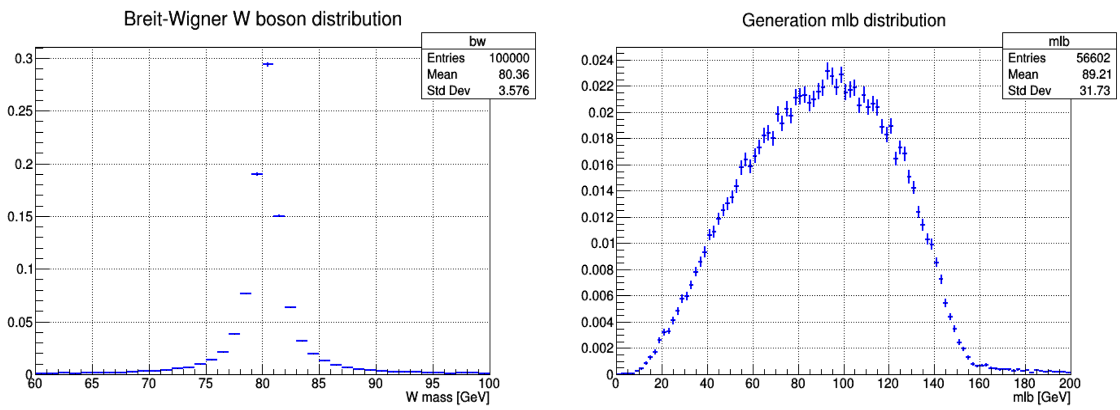


Figure 4.18: Breit-Wigner spectrum obtained when generating randomly the mass of the W boson (on the left) and true m_{lb} distribution obtained using the generation information from the standard $t\bar{t}$ process (on the right).

CHAPTER 5

Samples and objects definition

In order to find a possible hint for the production of dark matter in the **LHC!** collisions considering our signal models of interest, the data collected needs to be compared with **MC!** (**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 the samples and the code used for this particular search will therefore be thoroughly described in Section **??**.

Another crucial part of the analysis is definition of the objects that we want to consider. Section **??** will in this sense be devoted to the definition of the leptons, triggers and (b-)jets used in this work, while the different weights and correction factors that needs to be applied to the **MC!** simulation in order to be able to compare it to the data collected will finally be described in Section **??**.

5.1 Data and simulated samples

5.1.1 Data samples

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 ± 0.9) fb^{-1} (2016) [**?**], (41.5 ± 1.0) fb^{-1} (2017) [**?**] and (59.7 ± 1.5) fb^{-1} (2018) [**?**] has been collected, resulting in a total dataset of (137.1 ± 2.0) 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 **??** and by taking care of avoiding any eventual double counting due to eventual events which have been collected by several triggers at once. All the data samples actually considered for this analysis are listed in Appendix **??**.

5.1.2 The Monte-Carlo 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 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 interactions between the particles produced and the detector itself are extremely complex by nature.

The basic idea behind the **MC!** simulation generation process 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 algorithms called **event generators** and it is important to note that since we usually do not 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, as shown with the color code used in Figure **??**.

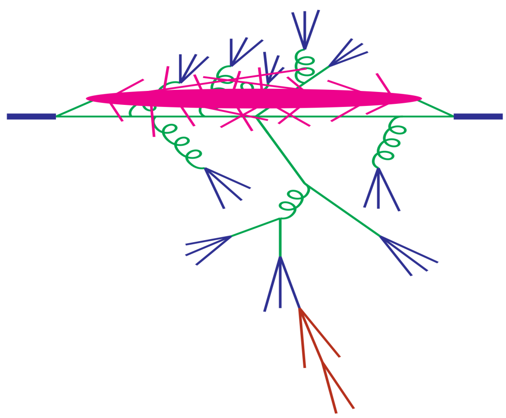


Figure 5.1: Structure of a pp collision and different steps of the **MC!** simulation used by the event generators, such as the parton shower (in green), the **UE!** (**UE!**) (in pink), the hadronization (in blue) and the decay of unstable particles (in red) [**?**].

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 **??** [**?**].

$$\sigma_A(s) = \sum_{i,j} \iint dx1\ dx2\ f_i(x_1, \mu^2)\ f_j(x_2, \mu^2)\ \hat{\sigma}_{ij \to A}(\hat{s}, \mu^2) \tag{5.1}$$

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

- The artificial parameter μ^2 used as the delimitation between short and long range physics;
- The **PDF!**s (**PDF!**s) $f_i(x, \mu^2)$ 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 $\hat{\sigma}_{ij\to A}$ 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 $ij \to A$ 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 [**?**] (at LO) and POWHEG [**?**] and MC@NLO [**?**] (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 **??**. 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 also be emitted by **ISR!** (**ISR!**) and by these remnants themselves.

UE! (UE!)

Once the hard scattering and all the possible gluon emissions simulated, the next step consists in considering the so-called **UE! (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 **MPI!**s (**MPI!**s). 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 accounted for and 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 fm^2: an increased **UE!** will be obtained when the so-called impact parameter, the distance between the parton and the center of this area, is decreased, making the collision mostly central and almost head-on [**?**]. The **UE!** is typically well simulated using algorithms such as Herwig [**?**] and PYTHIA [**?**]. The spectrum for the generation of some variables in a top enriched sample can be found in Figure **??**.

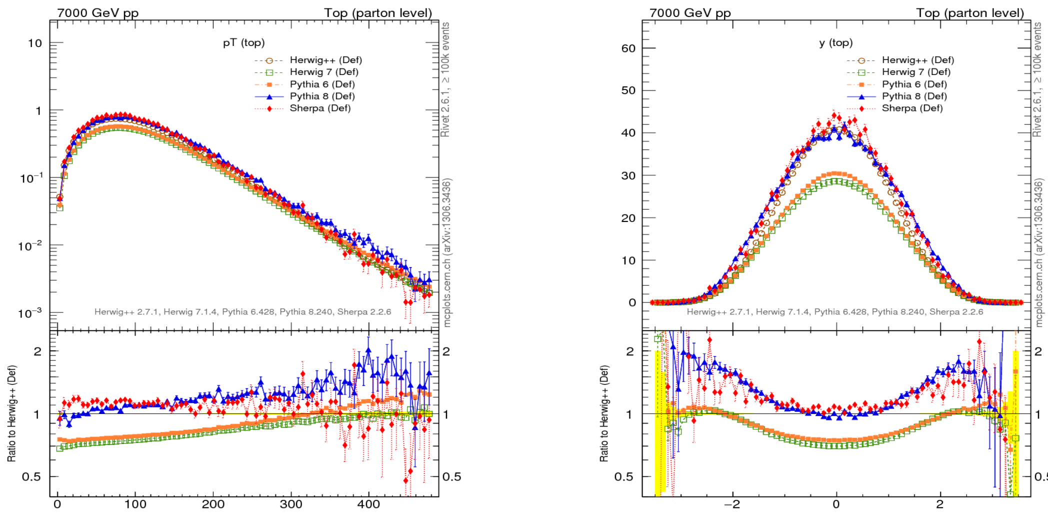


Figure 5.2: Top p_T (on the left) and rapidity (on the right) distributions obtained using different **MC!** generators [**?**].

Hadronization

Once all the primary and secondary collisions simulated, the event generators need to simulate the **hadronization** and binding processes of the different coloured partons emitted into colourless hadrons, following the process explained in Section **??**. 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 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 generated and the "imperfect" **CMS!** detector.

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

- The modeling of the interaction region;
- The modeling of the particle passage through the volumes that compose CMS detector and of the accompanying physics processes;
- The modeling of the effect of multiple interactions per beam crossing and/or the effect of events overlay (**PU!** simulation);

- The modeling of the detector electronic 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 **??**.

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 **??**.

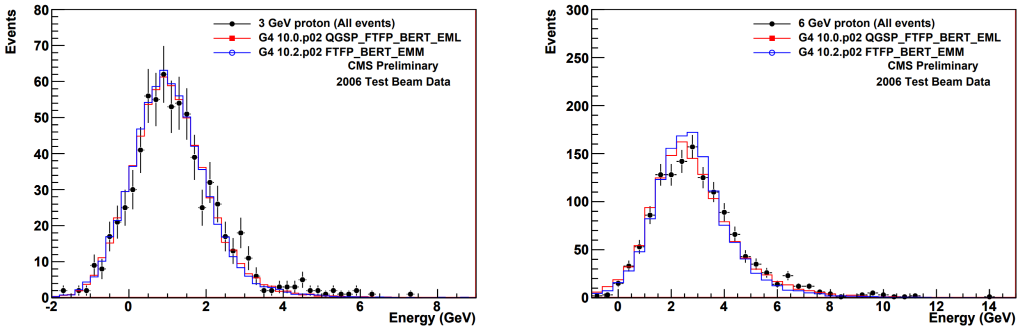


Figure 5.3: Proton energy distribution at 3 (on the left) and 6 (on the right) GeV compared for the test beam data (in black) and two different GEANT4 versions [**?**].

However, the modeling of the detector is not perfect and not all the inefficiencies can be accounted for. In some cases, **SF!** (**SF!**) are then used to precise the **MC!** simulations and correct some expected discrepancies between data and **MC!**.

5.1.3 Signal samples

Two different sets of **MC!** signal samples have been produced centrally with MADGRAPH and PYTHIA8 at **LO!** for this analysis, corresponding to the $t/\bar{t}$+DM and to the $t\bar{t}$+DM signals. Different mass points were produced first privately and then centrally (in the $t\bar{t}$+DM case), considering different dark matter masses, from 1 to 51 GeV, and different scalar or pseudoscalar mediator masses depending on the model considered, ranging from 10 to 1000 GeV. In this context, 13 (17) different mass points have been produced for each mediator considered for the $t/\bar{t}$+DM ($t\bar{t}$+DM, respectively) signals, as listed in Appendix **??**.

The impact on the kinematics (in this particular case, on the spectrum of the pf**MET!**) of these different mass points available can be observed in Figures **??** (scalar $t/\bar{t}$+DM), **??** (pseudoscalar $t/\bar{t}$+DM), **??** (scalar $t\bar{t}$+DM) and **??** (pseudoscalar $t\bar{t}$+DM). As expected from Table **??**, we can see first of all in these figures that the higher the mediator mass is, the lowest is its spectrum because of the lower cross section associated to the model. On the other hand, we can also observe that the mass of the **DM!** itself has little to no impact regarding to the kinematics of the event, as expected given the decay of such exotic matter to a pair of invisible particles. Finally, we can also observe in these plots the difference in kinematics between the background (in blue) and the different signals, which will be the basis of the analysis performed.

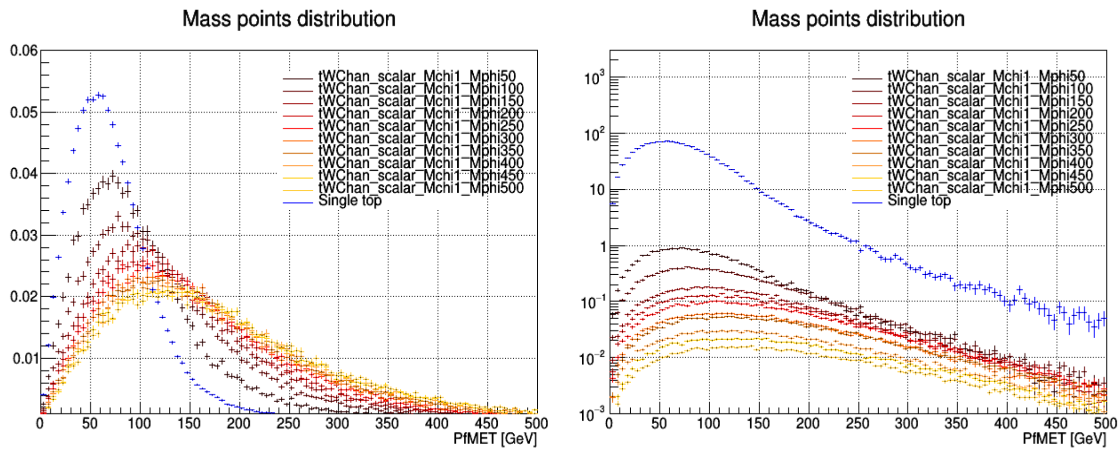


Figure 5.4: **MET!** spectrum for several $t/\bar{t}$+DM **scalar** mediators, with (on the left) and without unit normalization (on the right).

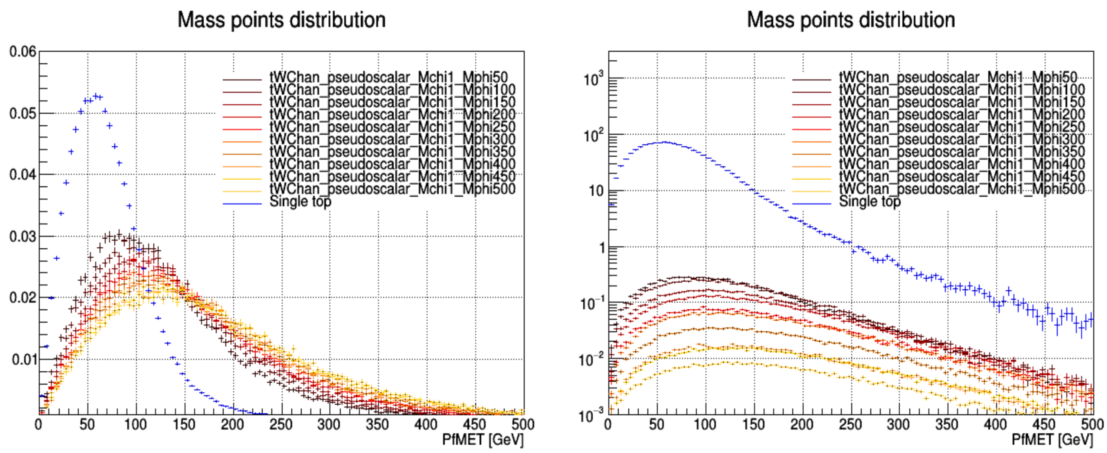


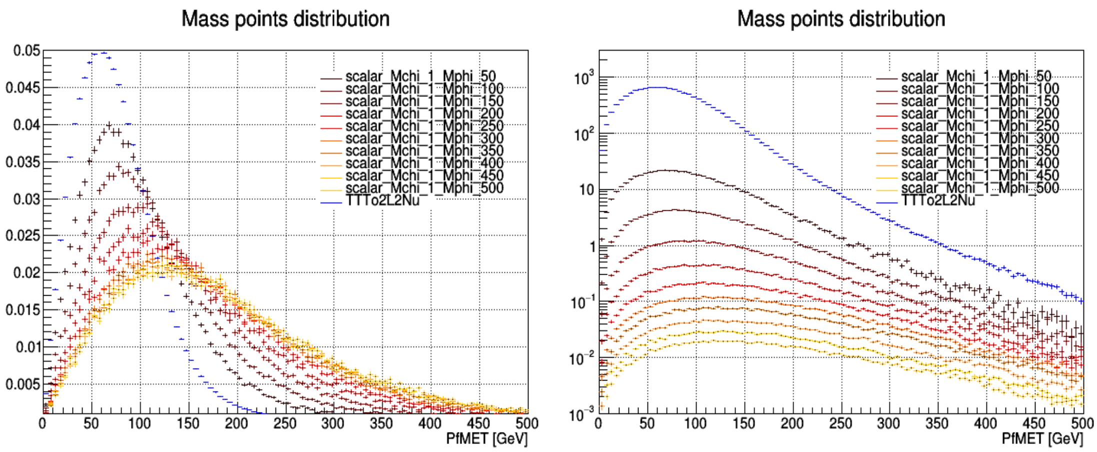
Figure 5.5: **MET!** spectrum for several $t/\bar{t}$+DM **pseudoscalar** mediators, with (on the left) and without unit normalization (on the right).

5.1.4 Files format

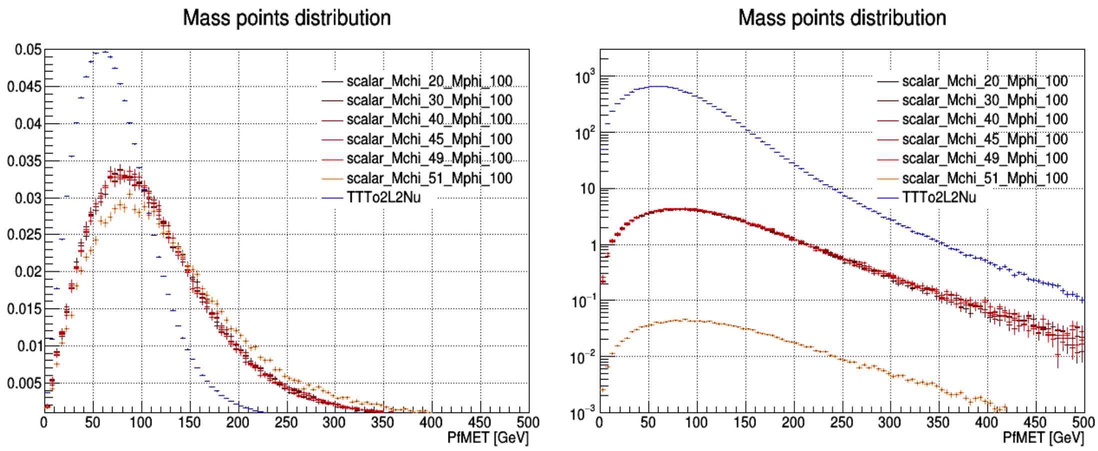
Once recorded (or simulated), the data (or **MC!**) still needs to go under a complete post-processing in order to change its format and reduce the total size of the samples to be considered in the different analyses. Different types of analyses are expected to need different levels of data reduction,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

- **Virgin-RAW**: used only in low rate runs with heavy ions collisions (10-15 Mb/event)
- **RAW** : standard raw data event content (1 Mb/event)
- **RECO**: detailed information on reconstructed physics objects (3 Mb/event)
- **AOD! (AOD!)**: physics objects used in analysis (400-500 kB/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. Then, 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 ~ 50 was recently introduced:



(a) Dark matter mass fixed to 1 GeV



(b) Mediator mass fixed to 100 GeV

Figure 5.6: **MET!** spectrum for several $t\bar{t}$+DM **scalar** mediators (on the top) and dark matter (on the bottom) masses, with (on the left) and without unit normalization (on the right).

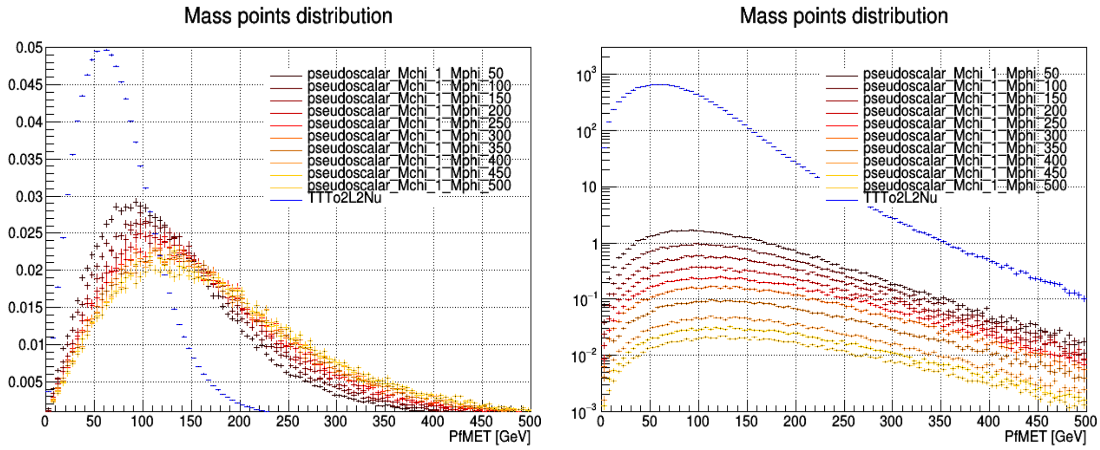
the nanoAOD, able to retain most of the information of each collision in around 1 kB of data per event only. This reduction in size was achieved by optimizing the floating point of the variables, by avoiding the storage of quantities that can be recomputed from the available information and by limiting 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 actually rely on such files in order for their work.

In this case, the 7th 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 (signals and backgrounds).

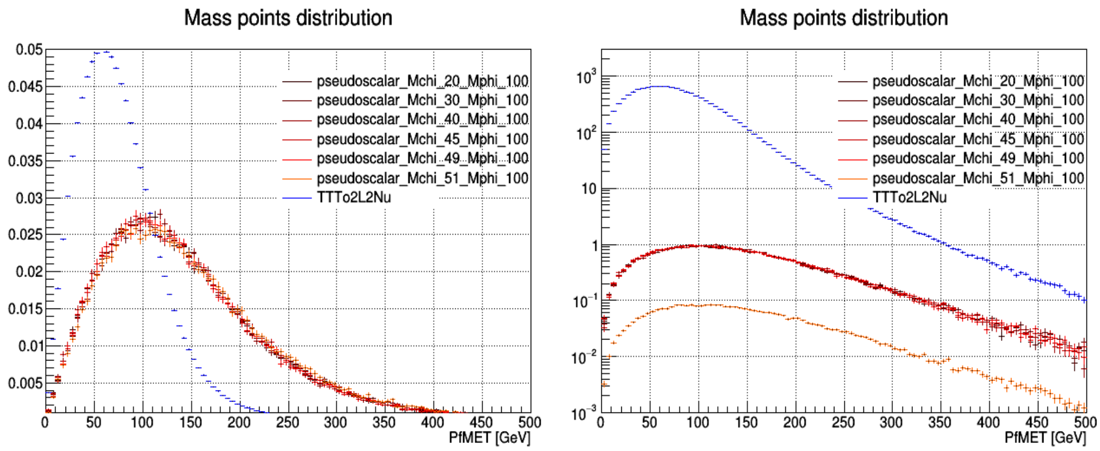
5.1.5 Analysis code

The code used for the event generation, simulation and reconstruction is the version 10_4_X of the official software of the **CMS!** collaboration, called CMSSW [**?**]. This software contains the **CMS! EDM!** (**EDM!**) which is able to describe every event as a C++ object containing all the RAW and reconstructed information related to the collision. These object are stored using the ROOT file format [**?**], an analysis package written in C++.

Once all the different samples produced centrally up to the nanoAOD stage, another framework



(a) Dark matter mass fixed to 1 GeV



(b) Mediator mass fixed to 100 GeV

Figure 5.7: **MET!** spectrum for several $t\bar{t}$+DM **pseudoscalar** mediators (on the top) and dark matter (on the bottom) masses, with (on the left) and without unit normalization (on the right).

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 framework, the so-called *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 correction factors to the **MC!** samples and produce the histograms and datacards needed to perform a search such as this one.

5.2 Objects definition

A typical $t/\bar{t}$ or $t\bar{t}$+DM event signature is made out of a certain number of (b-tagged) jets along with two leptons (electrons and/or muons) and some **MET!** coming from the two neutrinos and from the two **DM!** particles created along the way. It is extremely important to describe the **WP!** (**WP!**) picked and the selection applied in order to select the objects of the analysis, such as the leptons and the jets used, chosen in such a way to optimize the lepton reconstruction efficiency while reducing as much as possible the misidentification rates of the different objects.

Dataset	Run range	**HLT trigger path**
SingleMu	[297020,306462]	HLT_IsoMu27_v*
SingleEle	[297020,306462]	HLT_Ele35_WPTight_Gsf_v*
DoubleEG	[297020,306462]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[297020,299336]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
	[299337,306462]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*
MuonEG	[297020,306462]	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
	[297020,299336]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
	[299337,306462]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*

Table 5.1: 2016 trigger paths considered for this analysis.

Dataset	Run range	**HLT trigger path**
SingleMu	[297020,306462]	HLT_IsoMu27_v*
SingleEle	[297020,306462]	HLT_Ele35_WPTight_Gsf_v*
DoubleEG	[297020,306462]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[297020,299336]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
	[299337,306462]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*
MuonEG	[297020,306462]	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
	[297020,299336]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
	[299337,306462]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*

Table 5.2: 2017 trigger paths considered for this analysis.

5.2.1 Triggers selection

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

Our analysis relies on the dilepton final state, so 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 is considered. Eventual events passing several triggers are taken into account as well to make sure to avoid any double counting.

The trigger efficiencies, defined for each trigger as the ratio between the number of events passing our object selection and the trigger itself in the numerator and the number of events passing our selection in the denominator has been computed using orthogonal **MET!** datasets to avoid any bias. Studying this efficiency is important in the sense that we want to make sure that the triggers used are efficient enough in the p_T region of the leptons of the analysis to avoid any undesired effect due to the turn-on region of any trigger, and because we actually use the computed efficiency to reweight all the simulated samples used in this analysis. The efficiencies have been calculated using orthogonal **MET!** triggers and a **MET!** $>$ 100 GeV cut for the different data taking periods

Dataset	Run range	**HLT trigger path**
SingleMu	[315252,325175]	HLT_IsoMu24_v*
SingleEle	[315252,325175]	HLT_Ele32_WPTight_Gsf_v*
DoubleEG	[315252,325175]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[315252,325175]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*
MuonEG	[315252,325175]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
		HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 5.3: 2018 trigger paths considered for this analysis.

and for each channel individually, as shown in Figure **??**, following the **CMS!** recommendations [**?**]. At the end of the day though, 2D efficiencies (representing the leading lepton p_T versus η, as shown in Figure **??**) were actually estimated in both data and **MC!**. A scale factor was then estimated and used in the analysis to compute the weight of each event that needs to be applied to account for this particular effect.

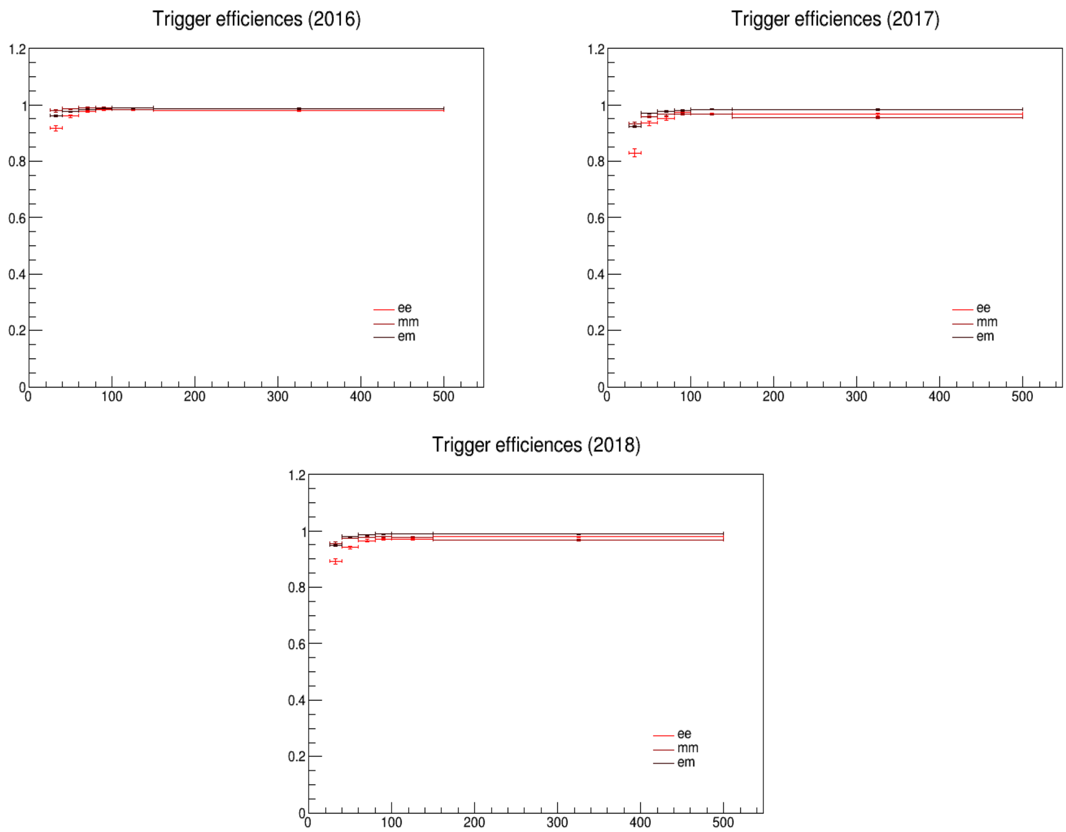
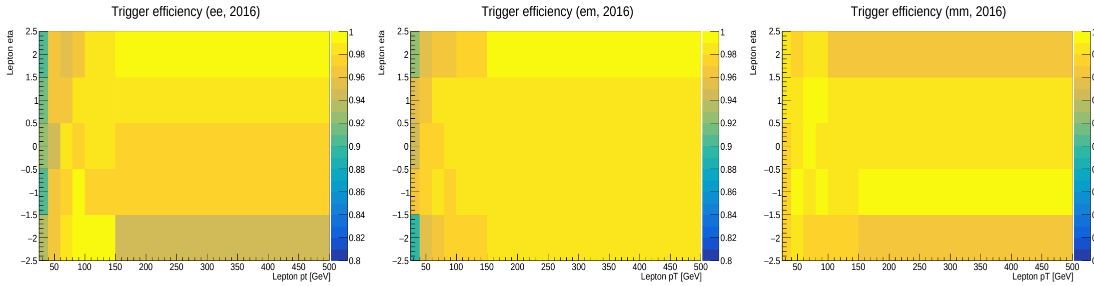
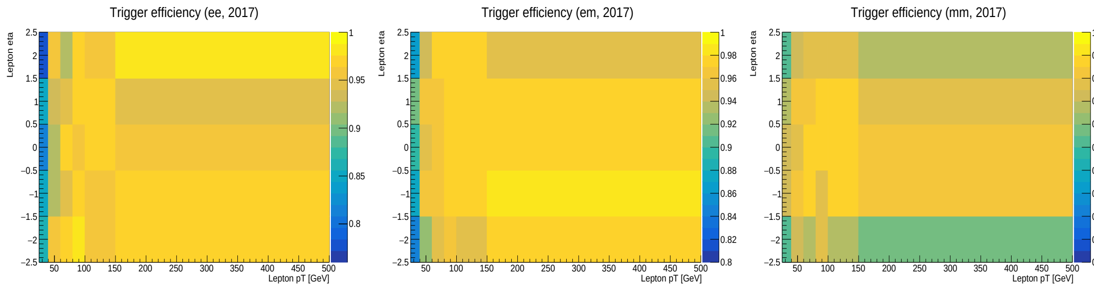


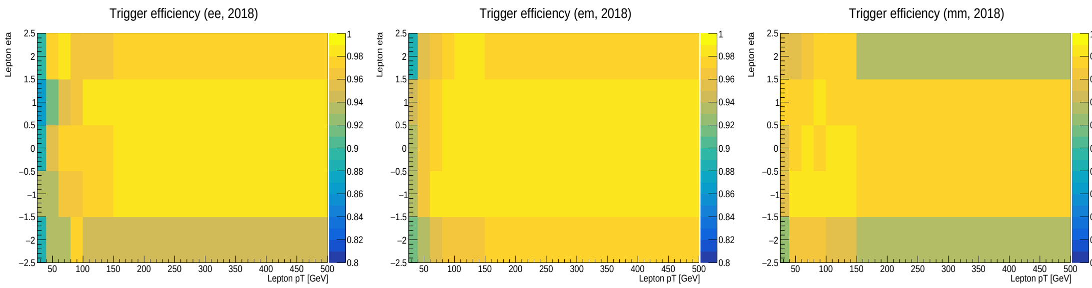
Figure 5.8: Trigger efficiencies with respect to the leading lepton p_T using orthogonal **MET!** datasets for each year in 2016 (on the top left), 2017 (on the top right) and 2018 (on the bottom).



(a) 2016



(b) 2017



(c) 2018

Figure 5.9: 2D trigger efficiencies obtained for the leading lepton p_T versus η, using orthogonal MET datasets for each dataset and for different channels (ee on the left, $e\mu$ on the middle and $\mu\mu$ on the right).

5.2.2 Electrons selection

Four different electron **WP!**s (veto, loose, medium and tight) are defined by the **CMS!** EGamma **POG!** [**?**] with slightly different quality cuts in the barrel and in the endcaps in order to select electrons with a given efficiency while trying to limit their misidentification rate. The tight **WP!** is then the one with the lowest electron selection efficiency (of the order of 70% for electrons with $p_T > 20$ GeV) but the best to reject misidentified electrons, and is to be used mostly when the backgrounds are expected to be quite large. On the other hand, the veto **WP!** corresponds to an average electron selection efficiency of the order of 95%.

For this analysis, we rely on the cut based medium identification **POG! WP!** to define our electrons [**?**]. Electrons originating from photon conversions are rejected by requiring that the electron track does not have missing hits in the innermost layers of the tracker, on top of the conversion veto included in the cut based ID definition.

5.2.3 Muons selection

The selection applied to muons is based on the Muon **POG!** as well, providing references efficiencies for standard selection and recommendations for the tight muon selection [?].

For our analysis, we decided to use directly the medium ID **WP!** provided by the **POG!** [?], along with a tighter isolation criterion (< 0.15) with $\Delta\beta$ correction and in a cone size $\Delta R < 0.4$, as defined in Equation **??**, in order to reduce the number of muons coming from the hadronization process of bottom and charm quarks.

$$\mathrm{ISO} = \frac{\sum p_T^{\text{ch. had. (PV)}} + \max\left(0, \sum E_T^{\text{neut. had.}} + \sum E_T^{\gamma} - 0.5 \times \sum p_T^{\text{ch. had. (PU)}}\right)}{p_T(\mu)} \tag{5.2}$$

5.2.4 Leptons selection

A few additional quality cuts are applied to both the electrons and muons. We ask them to have a $p_T > 8$ GeV and a pseudo-rapidity $|\eta| < 2.4$ to be pre-selected for the analysis.

Furthermore, lepton candidates trajectories are required to be compatible with the primary interaction vertex by imposing constraints on their transverse ($|d_0| < 0.05$ cm) and longitudinal ($|d_z| < 0.10$ cm) impact parameters, and on the three-dimensional impact parameter significance ($S_{3D}^{d} < 4$), computed as the ratio of the three-dimensional impact parameter and its uncertainty. This helps in removing some of the non-prompt contamination, whose misunderstanding might be a problem, mostly in the tail of the **MET!** spectrum.

Veto WP! As we will see later on, events featuring more than two leptons are rejected, and this is achieved by defining a looser lepton selection. For electrons, the cut based veto ID definition is used, and the requirement of not having missing hits is removed while for muons, candidates satisfying the loose ID and very loose isolation criteria are selected to define the veto **WP!**.

5.2.5 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 **??**, the jets are clustered from the **PF!** candidates using the anti-kT algorithm (with a typical distance parameter $R = 0.4$).

In this analysis, jets are selected following the tight **WP!** definition given by the **CMS!** JET/MET **POG!** [?], whose selection depends on the pseudorapidity of the jet and on the data taking period [?]. The tight **POG! 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 $|\eta| < 3.0$. Jets are further required to have a $p_T > 30$ GeV Sand a $|\eta| < 2.4$ to be considered in this analysis, and jets overlapping with any selected lepton within a cone of radius $R < 0.4$ are removed to prevent signal leptons clustered as jets from entering the jet counting. The recommended jet smearing method is further applied to jets, which consists in using data rather than simulation in order to estimate the amount of **MET!** to be expected from multijet production.

Finally, a slightly different additional selection (tight **PU!** jet ID) is applied to jets having a $p_T < 50$ GeV in order to reject jets coming from **PU!** interactions. Following the official **POG!** recommendations, jet energy correction and resolution are also taken into account [?, ?].

b-jets The b-jets are selected among all the jets using the recommendations given by the B-Tagging and Vertexing **POG!** [?]. The medium deepCSV b-tagging **WP!** is used in this case, as explained in Section **??**, for which the rate for misidentifying a light jet as a b-jet is around 1%. A jet is therefore in particular considered to be a b-jet in this analysis if it passes the jet requirements and if its deep CSV b-tagging weight is larger than 0.6321, 0.4941 or 0.4184 for the year 2016, 2017 and 2018, respectively.

5.3 Weights and corrections applied

Several different weights and **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!** simulations, such as the efficiency of the selection of the different objects, usually directly provided by the different **POG!**s (**POG!**s) for their own default objects definitions.

The **MC!** samples are additionally typically reweighed to match the distribution of true interactions observed in data due to multiple collisions collisions happening in the same bunch-crossings (the **PU!**, as defined in Section **??**). To illustrate the effect of these corrections, a few particular weights that are applied to the **MC!** samples will now be detailed.

MET! corrections

The **MET!** distribution is expected to be independent of ϕ because of the rotational symmetry of the collisions around the beam axis but we do observe that the reconstructed **MET!** does depend on ϕ because of possible anisotropic detector responses, inactive calorimeter cells or tracking regions, detector misalignment, or eventual displacements of the beam spot. XY-shift corrections have therefore been applied to reduce this modulation. In 2017, large level of noise were also observed in the data collected by the **ECAL!** at high pseudorapidities. To mitigate this effect, especially on the **MET!** tails, the correction consisting in completely excluding jets with raw transverse momentum < 50 GeV and $2.650 < |\eta| < 3.139$ from the calculation of the **MET!** was considered, as recommended by the JET/MET **POG!** [?]. The effect of these corrections is shown in Figure **??**.

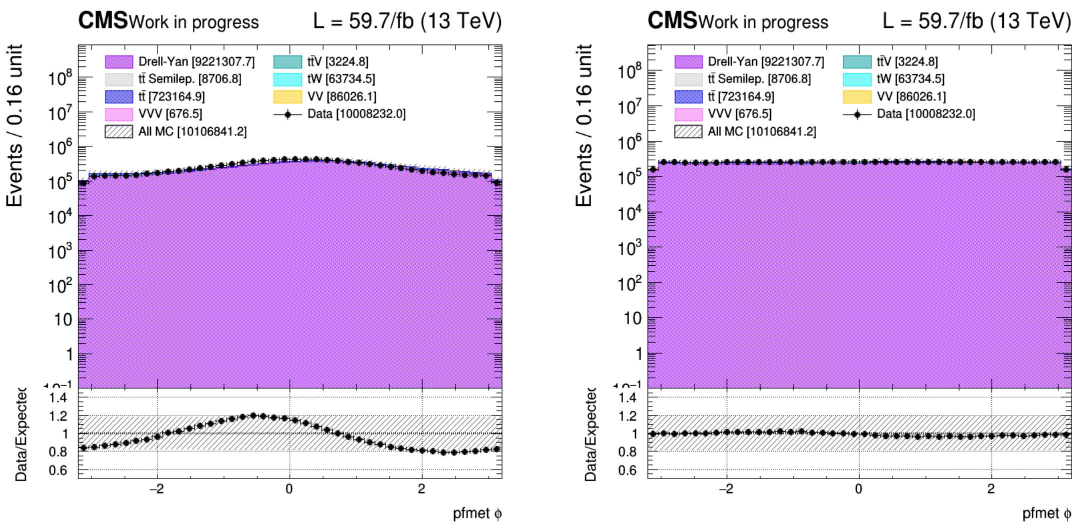
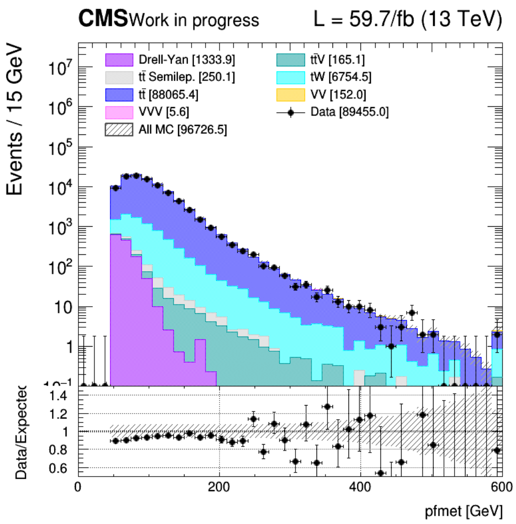


Figure 5.10: Uncorrected MET (on the left) and MET ϕ after applying the EE noise and XY-shift corrections (on the right) distributions observed in a 2018 **DY!** (**DY!**) inclusive control region.

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 [?]. This known mismodeling of the top quark can be corrected using a general reweighting recipe [?], and its effect on the p_T spectrum can be observed in Figure ??.



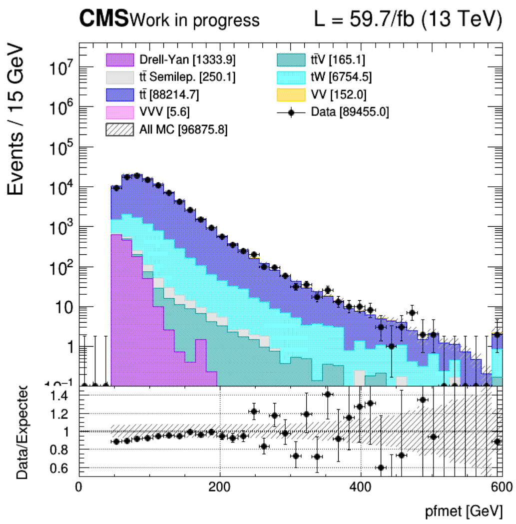


Figure 5.11: Data/**MC!** agreement without (on the left) and with (on the right) top p_T reweighting corrections in a 2018 top control region.

Even though we initially planed on applying this correction factor, as it seems to slightly improve the agreement between data and simulation, the most recent recommendation is not to apply this factor for our kind of **BSM!** searches where $t\bar{t}$ **MC!** is used to model the background, as this correction factor might suppress a potential **BSM!** signal and bias the final results. In this case, the effect of the top p_T mismodeling can be considered covered by the existing uncertainties and no additional correction or uncertainty is needed [?].

Other factors

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 **L1!** trigger 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 prefired [?]. The effect this correction has on the data/**MC!** agreement in a 2018 inclusive enriched control region is shown in Figure ??.

In 2018, another issue affecting this time two endcaps of the **HCAL!** and known as the HEM15/16 issue was reported [?], resulting in a loss of around 2% of **HCAL!** coverage. This issue has a small but measurable effect on the **MET!** spectrum, as shown in Figure ?? and is therefore taken into account and corrected by introducing another weight to the simulation.

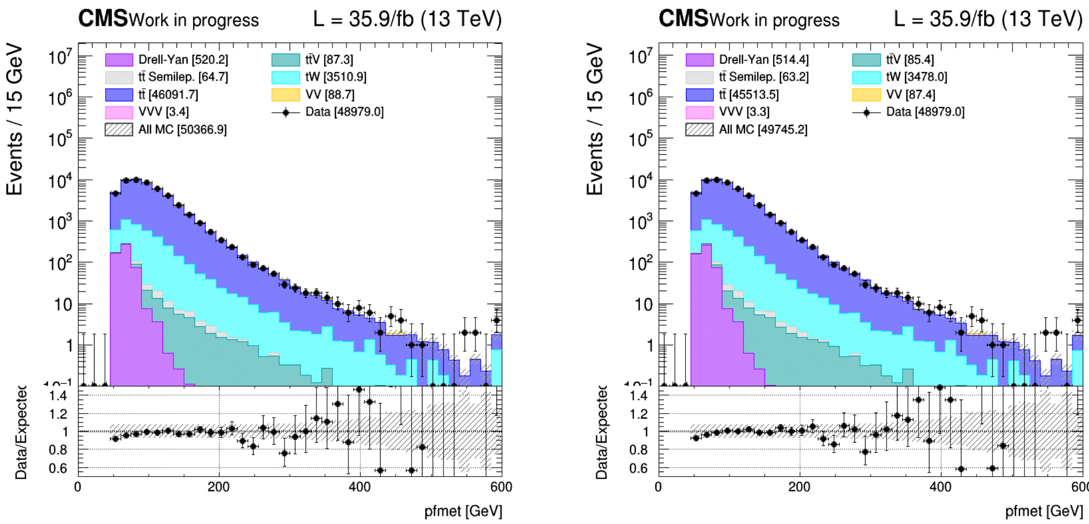


Figure 5.12: Data/**MC!** agreement without (on the left) and with (on the right) prefiring correction factor in a 2016 top control region.

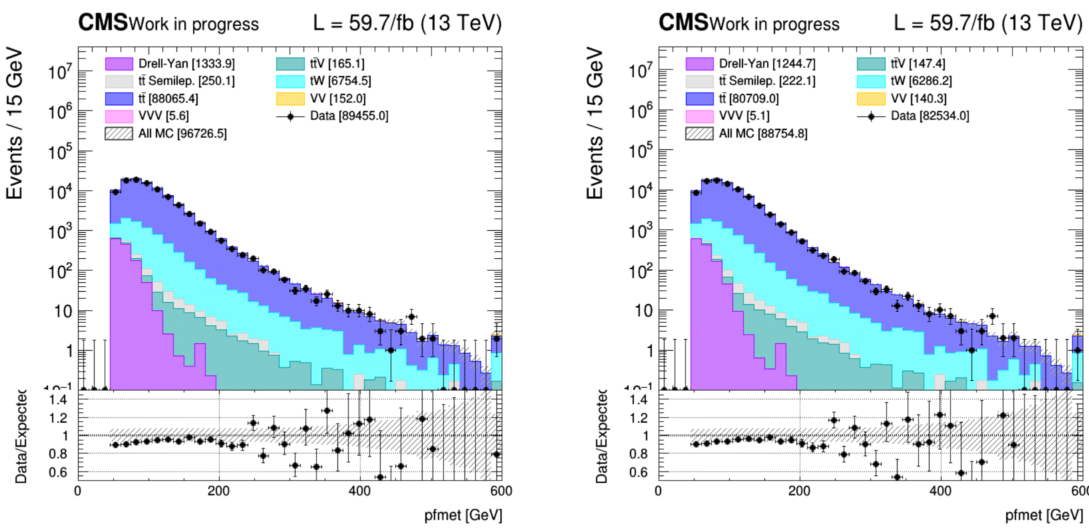


Figure 5.13: Data/**MC!** agreement without (on the left) and with (on the right) HEM correction factor in a 2018 top control region.

CHAPTER 6

Event selection and backgrounds prediction

In order to first of all define the control and signal regions for the analysis, the set of cuts defining the criteria for our selection of events needs to be introduced, as will be done in Section **??**.

Several **SM!** background processes have been considered for this analysis, all listed in Appendix **??** and mostly estimated directly from **MC!**. In Section **??**, the main backgrounds to consider for this particular analysis and the ways we have in order to check that we can trust the **MC!** simulation will be reviewed, starting with the **SM!** top production.

6.1 Event selection

6.1.1 Inclusive region

For this analysis, only events passing the so-called *inclusive region* were considered. This region, as inclusive as technically possible, is mostly used in order to spot initial issues and problems with the simulation of the major backgrounds of the analysis, and was therefore extensively studied. It is defined from the following cuts:

- Two opposite sign good leptons ($p_T > 25$ (20) GeV for the leading (trailing) lepton, both having $|\eta| < 2.4$) are first of all required;
- Then, events with a third lepton having a veto type and a $p_T > 10$ GeV are rejected;
- At least 1 jet ($p_T > 30$ GeV) is required;
- Low mass resonances are removed by asking for $m_{ll} > 20$ GeV.

Some of the distributions obtained in this particular region are shown in Figure **??**. A slight **MC!** mismodeling mostly due to the **DY!** process can be observed in these plots, but this feature is known in **CMS!** and mitigated anyway with the analysis cuts applied, and compensated for with the computation of a semi data-driven scale factor and the introduction of a large systematic uncertainty associated to this particular process.

6.1.2 Pre-selection region

A global pre-selection region was then defined, by applying on top of the inclusive region a few cuts:

- A 76-106 GeV Z-window veto is applied in the ee and $\mu\mu$ channels;
- The **MET!** has to be > 100 GeV;
- The stransverse mass (defined in Section **??**) $M_{T2}(ll)$ should be > 80 GeV;
- And at 1 medium deep CSV b-jet is required.

This selection serves as the basis for our signal regions and allows to keep this particular region orthogonal to the $t\bar{t}$ control regions used by the semileptonic channel, so that both channels can be combined, while removing more background than signal anyway.

Some distributions obtained in this pre-selection region are shown in Figure **??**. It is important to note at this point that, unless stated otherwise, all the relevant systematic uncertainties are applied on all the following data/**MC!** distributions shown.

6.1.3 Signal regions

Our signal regions are then defined on top of the pre-selection region. Two signal regions are defined, each enriched in one signal of interest:

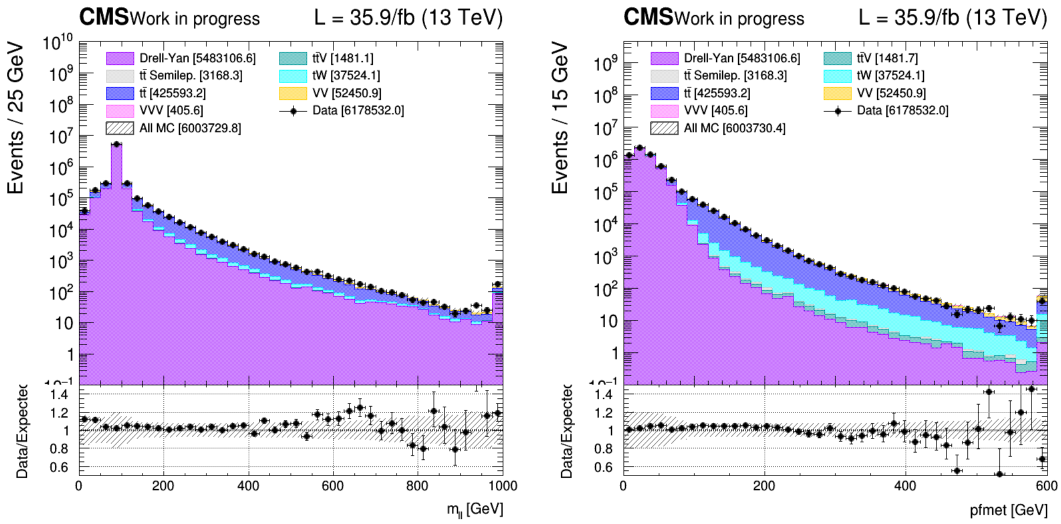
- The first one targeting the $t/\bar{t}$+DM signal, by considering events having exactly 1 jet, or exactly 2 jets and 1 b-jet;
- And another one targeting the $t\bar{t}$+DM signal, by considering this time only the events having exactly 2 jets and more than 1 b-jet, or more than 2 jets.

6.2 Backgrounds prediction

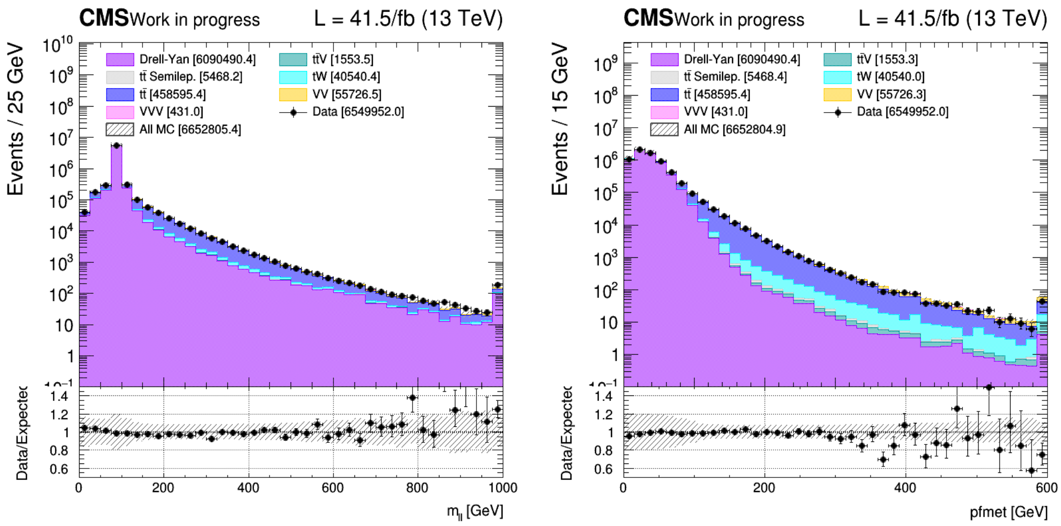
6.2.1 Top production

Because of the relatively high production cross section of top quarks at 13 TeV, the production of one or two top quarks, but without the production of associated **DM!**, is hard to suppress and is obviously the dominant background in both searches.

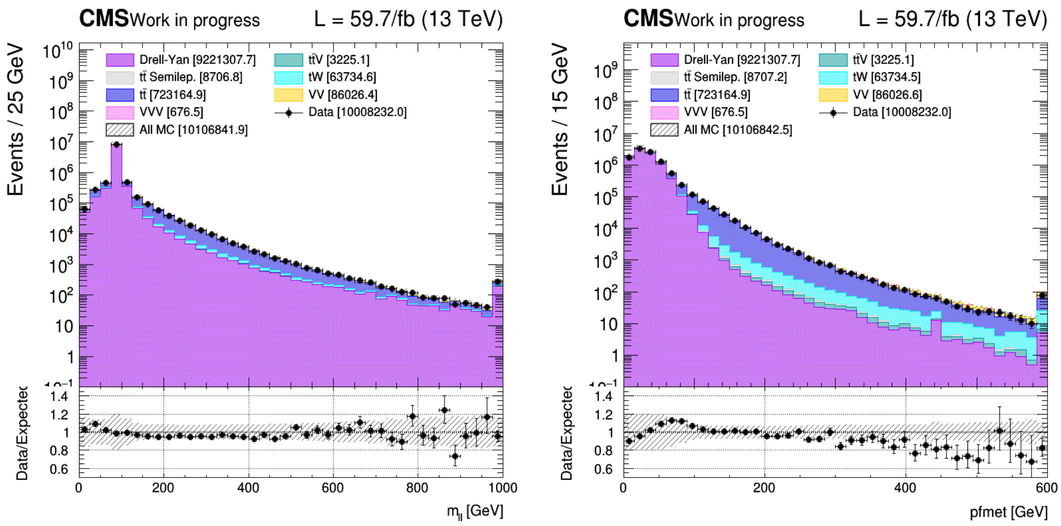
Such backgrounds actually have kinematics quite close to the ones expected for our signals. However, the additional **MET!** expected because of the production of a pair of **DM!** particles when considering the signals typically leads to a few differences in kinematics, allowing for some discrimination between this kind of background and the signal processes.



(a) 2016 inclusive region

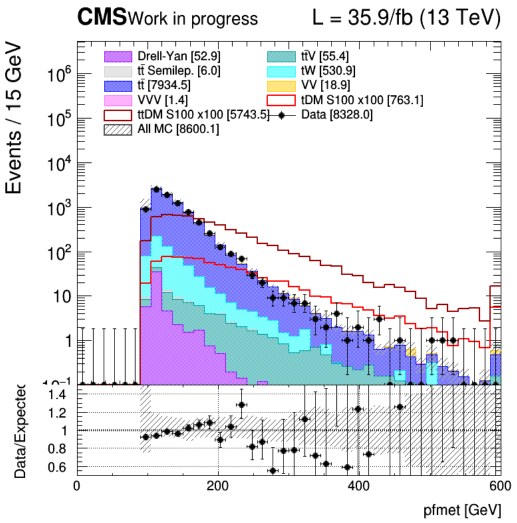


(b) 2017 inclusive region

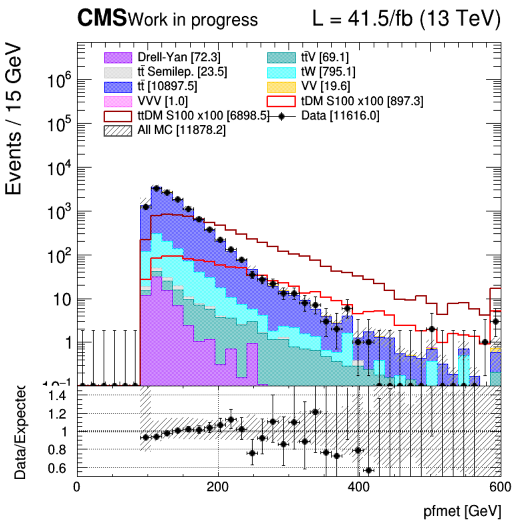


(c) 2018 inclusive region

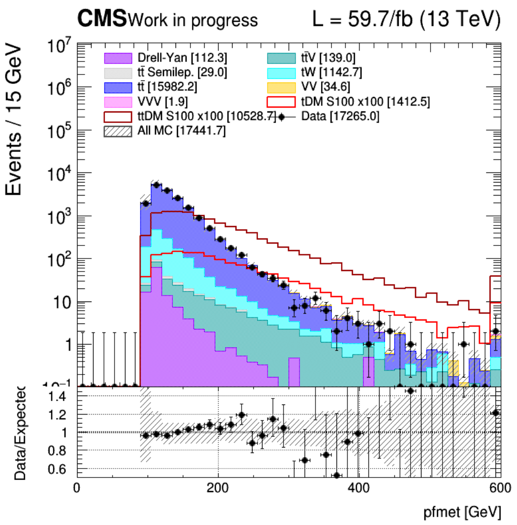
Figure 6.1: Two different variables (m_{ll} on the left and pf**MET!** on the right) represented in the inclusive control region defined.



(a) 2016 pre-selection region



(b) 2017 pre-selection region



(c) 2018 pre-selection region

Figure 6.2: Two different variables (pf**MET!**, on the left, and the stransverse mass M_{T2}^{ll}, on the right) represented in our pre-selection region.

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

This background is the most relevant for this analysis because of its large cross section and kinematics close to our signals of interest.

Different Feynman diagrams contribute to this process at **LO!** in a hadron collider, as shown in Figure **??**. NNPDF3.0 [**?**] was used as the default **PDF!**, while the default POWHEG setup with the PYTHIA8 CUETP8M2 (for 2016) and CP5 (for 2017 and 2018) tunes were used for the generation of this particular sample [**?**].

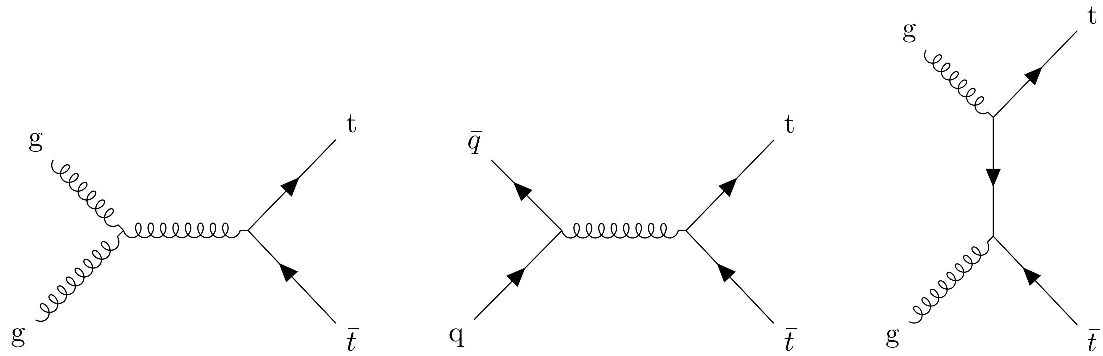


Figure 6.3: Main feynman diagrams for the production of the **SM!** $t\bar{t}$ process.

Single top

Different Feynman diagrams account for this process in the s-channel (Figure **??**), t-channel and tW production modes (Figure **??**), the latter being the dominant contribution in this analysis given its kinematics and cross-section. This process is also simulated using POWHEG (or MC@NLO, depending on the year and on the channel) with NNPDF3.0 and PYTHIA8, considering the CUETP8M1 (for 2016) and CP5 (for 2017 and 2018) tunes.

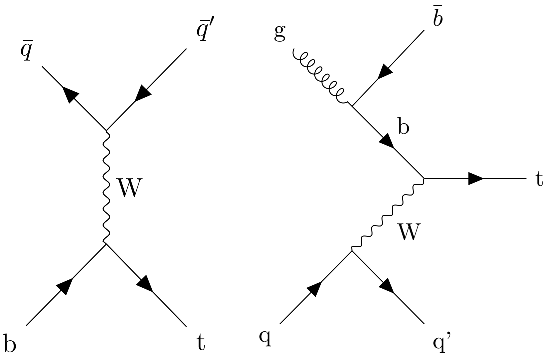


Figure 6.4: Feynman diagrams for the s-channel production mode of a single top quark.

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 then decays itself into different channels even though only its leptonic decay is usually considered in this particular case. The decay considered for the top/anti-top produced is represented in Figure **??**.

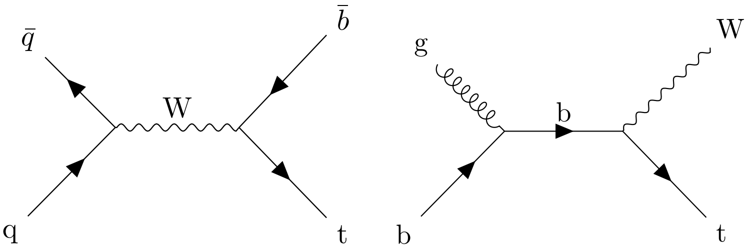


Figure 6.5: Feynman diagrams for the t-channel (on the left) and tW (on the right) production modes of a single top quark.

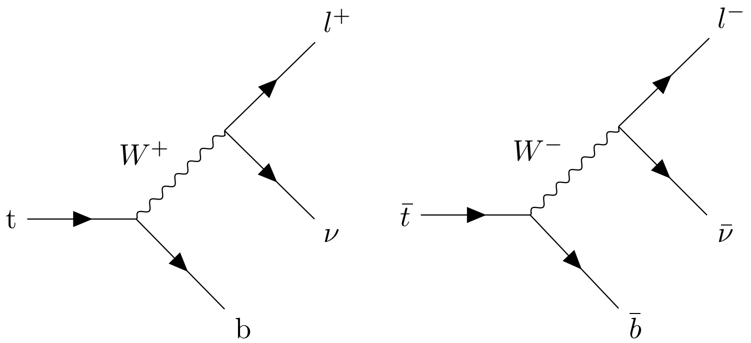


Figure 6.6: Feynman diagrams for the leptonic decay of the top (on the left) and anti-top (on the right) quarks.

The dilepton decay of the top quarks produced is obviously expected to be dominant in this analysis, but its semileptonic decay is also considered to make sure to cover possible events in which the jet emitted is misreconstructed by the detector.

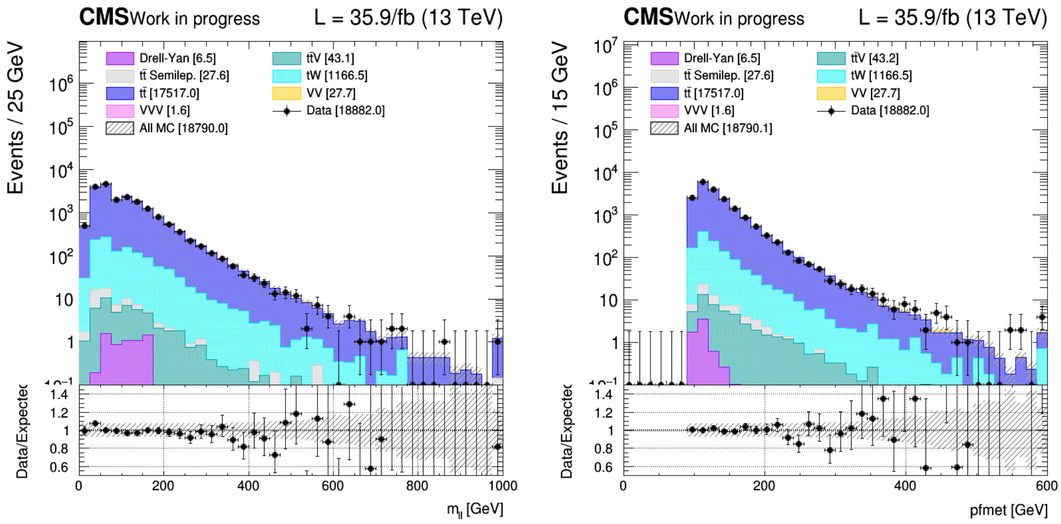
Top control region

Both these backgrounds are estimated directly from **MC!** and checked in a dedicated control region, defined from the previous pre-selection control region, but selecting only events having a **MET!** > 50 GeV and stranverse mass $60 < M_{T2}^{ll} < 80$ GeV, making this region perfectly orthogonal to our signal regions by definition. Some of the distributions obtained in this region are shown in Figure **??**. The small data-**MC!** discrepancies observed in this region are within the allocated uncertainties so the agreement is considered to be good enough for the analysis.

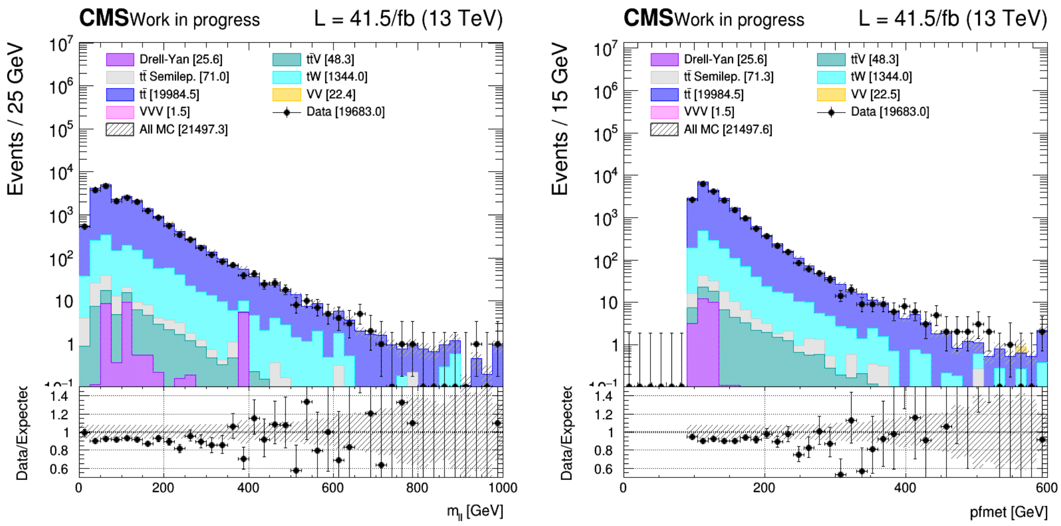
6.2.2 Drell-Yan

As previously mentioned, most of the **DY!** background, produced through the Feynman diagram represented in Figure **??**, is not expected to survive the selection applied to the analysis. However, because of the huge cross section of this process, two to three orders of magnitude larger than the production of a top quark, a thorough description of such process is still extremely important. Expected to be found mostly in the ee and $\mu\mu$ channels, it does survive even in the $e\mu$ channel because of possible tau decays ($Z\gamma^* \to \tau\tau \to e\mu\nu_e\nu_\mu\nu_\tau\nu_\tau$).

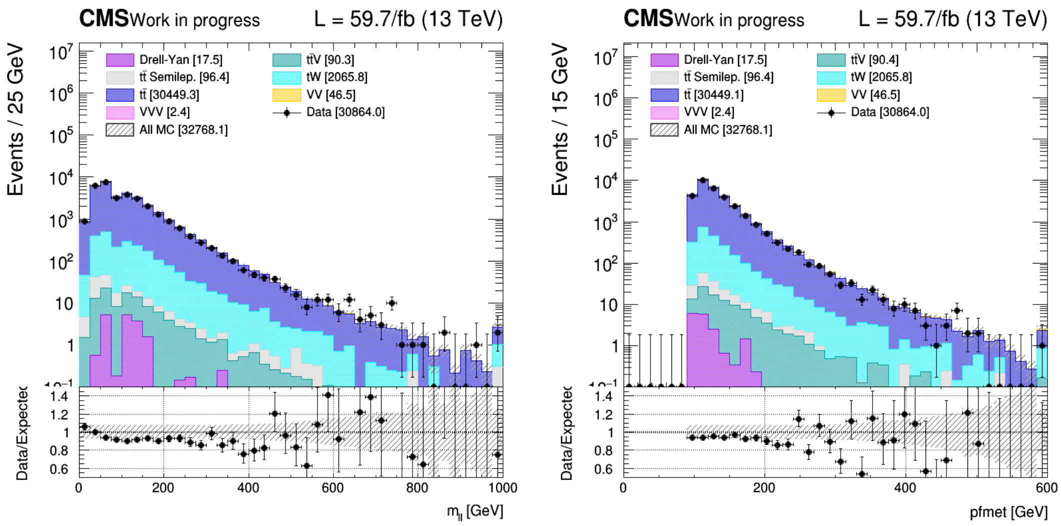
The samples for such background are generated with the MADGRAPH generator, MLM matching and interfaced to PYTHIA8 with the CUETP8M1 (for 2016) and CP5 (for 2017 and 2018) tunes for hadronization. The generation is then split into two distinct Z invariant mass ranges: 10-50



(a) 2016 top **CR!** (**CR!**)



(b) 2017 top **CR!**



(c) 2018 top **CR!**

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

GeV and >50 GeV. The HT-binned[1] samples have been chosen for the high mass sample over the inclusive ones in order to increase the statistics available.

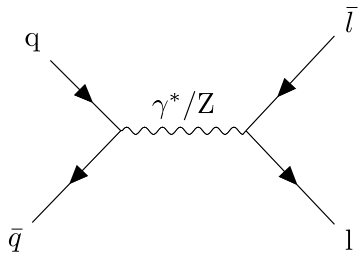


Figure 6.8: Feynman diagram for the **DY!** process involving a virtual γ^* or Z boson.

Apart from the impact it has on the different control and signal regions, the shape of this background is also estimated from **MC!** but a correction factor to its normalization is obtained using data, by using the a general method called $R_{\text{in-out}}$.

R_{in-out} method

The idea behind this semi data-driven method is simple: since a 15 GeV veto in the dilepton invariant mass m_{ll} is applied to both the ee and $\mu\mu$ channels to reduce the **DY!** contamination in the signal region, then we could use such vetoed events in order to estimate the **DY!** contribution outside of the Z mass window defined, referred to as M_Z in the following discussion. Mathematically, we can start by expressing in Equation **??** the total number of **DY!** events N_{DY}^{total} as the sum of events observed inside N_{DY}^{in} and outside N_{DY}^{out} of M_Z.

$$N_{DY}^{total} = N_{DY}^{in} + N_{DY}^{out} \tag{6.1}$$

The parameter $R_{out/in,}$ is then defined as the ratio between the number of **MC!** or data events outside and inside this Z mass window. This ratio is then simply used to estimate the number of DY events outside the veto region in data from the number of observed events in the veto region, as shown in Equation **??**.

$$N_{DY}^{out} = N_{DY,data}^{in} \cdot \left(\frac{N_{DY,MC}^{out}}{N_{DY,MC}^{in}} \right) \equiv N_{DY,data}^{in} \cdot R_{out/in,\ MC} \tag{6.2}$$

This equation relies on several assumptions. First of all, we have to assume that M_Z is dominated by actual **DY!** events and peaking backgrounds leading to a similar $R_{out/in}$ factor because of the prompt Z boson present in such backgrounds. Since this is typically a strong assumption to make, we actually decided to remove the contribution of non-peaking backgrounds from the data yields to take this effect into account properly. Additionally, we also have to assume that the $R_{out/in}$ factors measured in data and **MC!** are similar, assumption which might not be fully verified if there are data/**MC!** discrepancies in the m_{ll} distribution. To account for this effect, we defined a region close to the signal region by reversing the b-jet requirement. If we assume that this mismodeling does not have a dependence in the number of b-jets, as verified in Figure **??**, then we can simply correct the transfer factor previously obtained by a factor κ defined in Equation **??**, and whose value has been determined experimentally to be of the order of $\sim 5\%$.

[1] H_T is a variable corresponding to the scalar sum of the transverse momenta of the jets in an event

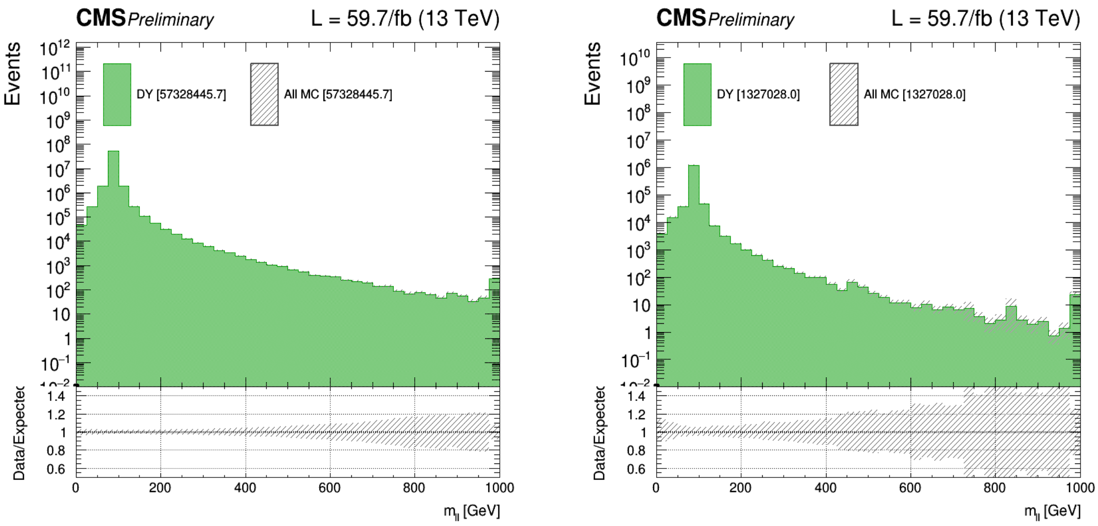


Figure 6.9: Normalized m_{ll} **DY!** distributions obtained using 2018 **MC!** simulations in the 0 (on the left) and 1+ b-jet (on the right) bins.

$$R^{corr}_{out/in,\ MC} = \kappa \cdot R_{out/in,\ MC} = \frac{R^{0bj}_{out/in,\ MC}}{R^{0bj}_{out/in,\ data}} \cdot R_{out/in,\ MC} \tag{6.3}$$

Additionally, we checked the behaviour of the transfer factor $R_{out/in,\ MC}$ across a large range of **MET!** values to make sure it is stable enough, as shown in Figure **??**. Given the relative flatness of the distributions obtained, we decided to apply a single flat scale factor to the **DY!** process, considering additionally a systematic uncertainty simply taken as the observed maximum observed deviation between the central value and the different bins (around 20% in this case).

DY! control region

Given the importance of this process, a dedicated control region was also defined in order to check the behaviour of the **DY! MC!** simulations as well. This region is defined from the pre-selection region, but by simply reversing the Z-veto requirement and relaxing the **MET!** cut down to 30 GeV, making it perfectly orthogonal and allowing us to use the $R_{\text{in-out}}$ method described in Section **??**, as shown in the distributions of Figure **??**.

6.2.3 Top quark production in association with vector bosons

These backgrounds, usually labeled as $t\bar{t}+W$ and $t\bar{t}+Z$, and generically grouped together under the label $t\bar{t}+V$, are coming from an usual $t\bar{t}$ production along with an **ISR!** (**ISR!**) or **FSR!** (**FSR!**) production of a W or Z boson, as shown in Figure **??**. The contribution of both these backgrounds is also taken directly from **MC!**, generated using either MADGRAPH or MC@NLO, depending on the sample, with the PYTHIA8 CUETP8M1 (for 2016) and CP5 (for 2017 and 2018) tunes for the hadronization process.

The resulting cross section of such processes 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 some neutrinos, leading to some actual **MET!**.

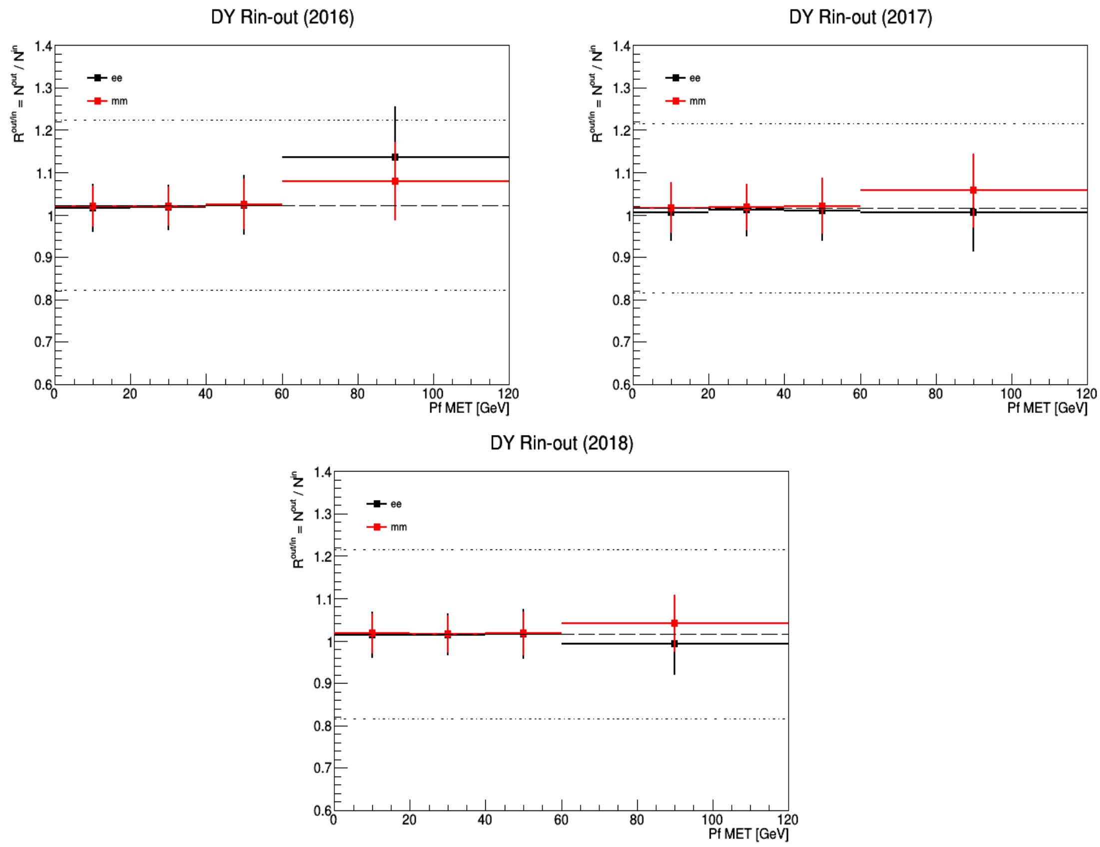
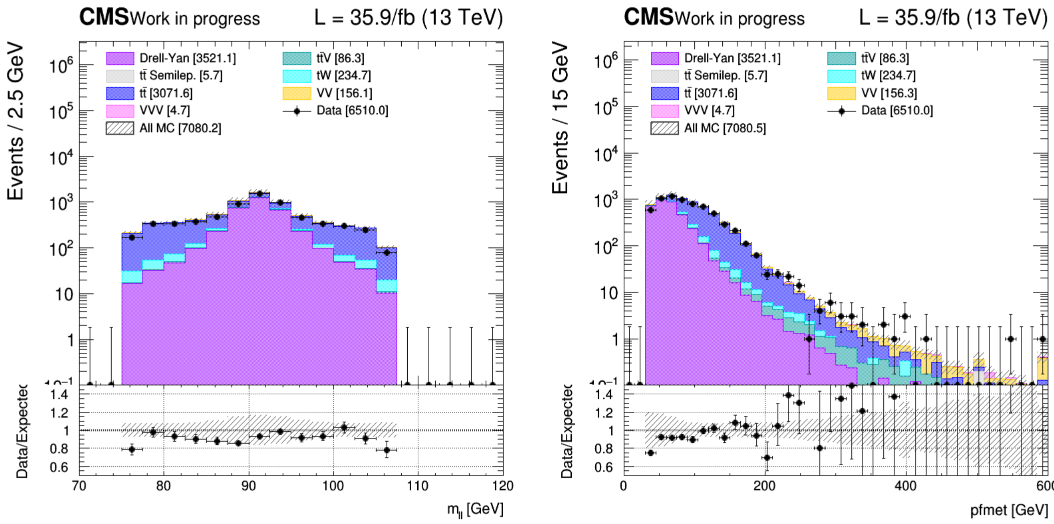


Figure 6.10: $R_{out/in}$ transfer factor obtained with the Rin-out data-driven method in bins of **MET!** in 2016 (on the top left), 2017 (on the top right) and 2018 (on the bottom).

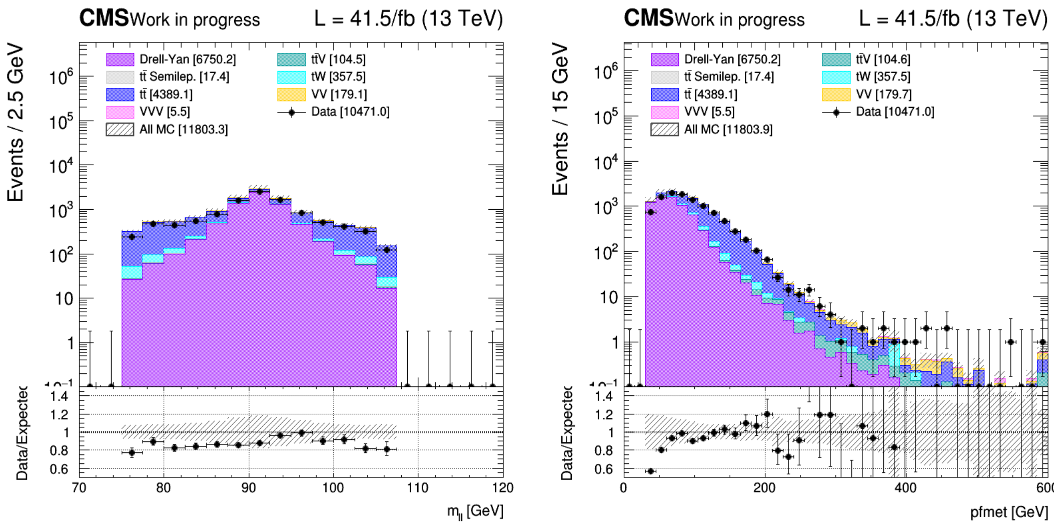
6.2.4 Smaller backgrounds

One background which can be of high importance in many analyses is the contribution from the so-called non prompt leptons, which can be for example jets misidentified by the detector as leptons or real leptons coming from heavy flavour decays. Because of its complexity, **MC!** simulations can usually not describe precisely this background and, in many cases, data-driven methods such as the one described in Appendix **??**, can be used in order to infer the kinematics of this process. However, the non-prompt contamination of this particular analysis is strongly reduced by the **IP!** (**IP!**) cuts applied on the leptons of the analysis, and is expected to be dominated by the semileptonic decay of the **SM!** $t\bar{t}$ process, when one of the bottom quarks produced leads to a fake lepton. This particular process, contrary to the W+jets and Z+jets processes, is expected to be modeled quite well by the **MC!** simulations. We therefore decided to rely entirely on the **MC!** simulations for this analysis, mainly because they typically come with a smaller systematic uncertainty than the data-driven estimation for this process.

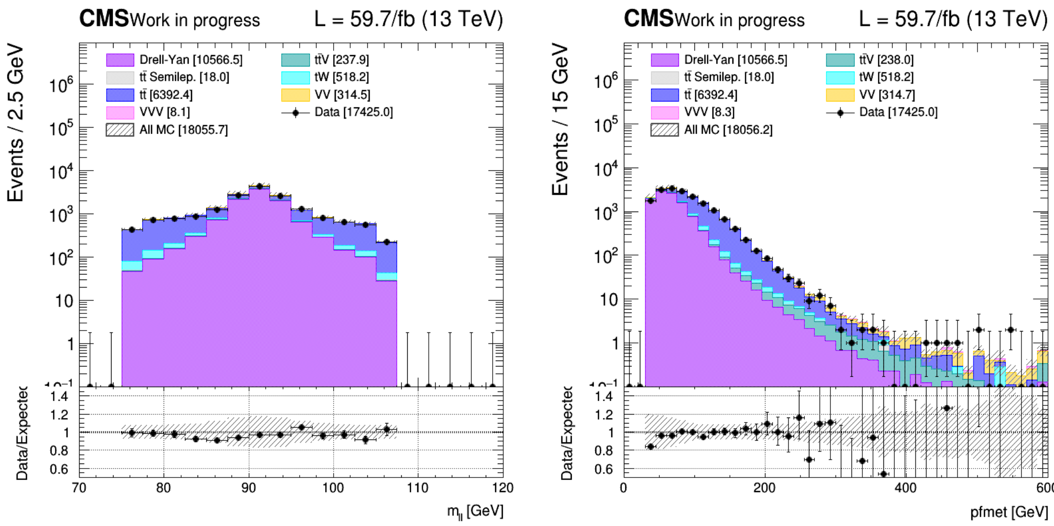
Finally, even though quite negligible and reducible, some additional backgrounds still need to be considered, such as the dibosons (WW, WZ and ZZ) and tribosons (WWW, WWZ, WZZ, ZZZ, WWγ) productions, as shown in Figure **??**. All these smaller backgrounds are taken directly from **MC!** and account in total for less than 1% of the total backgrounds in the signal regions.



(a) 2016 **DY! CR!**



(b) 2017 **DY! CR!**



(c) 2018 **DY! CR!**

Figure 6.11: Two different variables (m_{ll} on the left and pf**MET!** on the right) represented in the **DY!** control region defined.

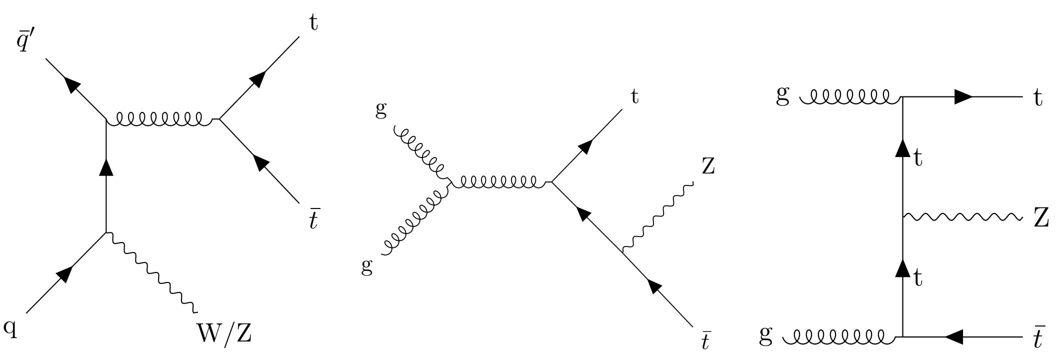


Figure 6.12: Possible Feynman diagrams for the **ISR!** $t\bar{t}$ with a W/Z boson (on the left) and for the production of an **FSR!** ttZ (on the center and right).

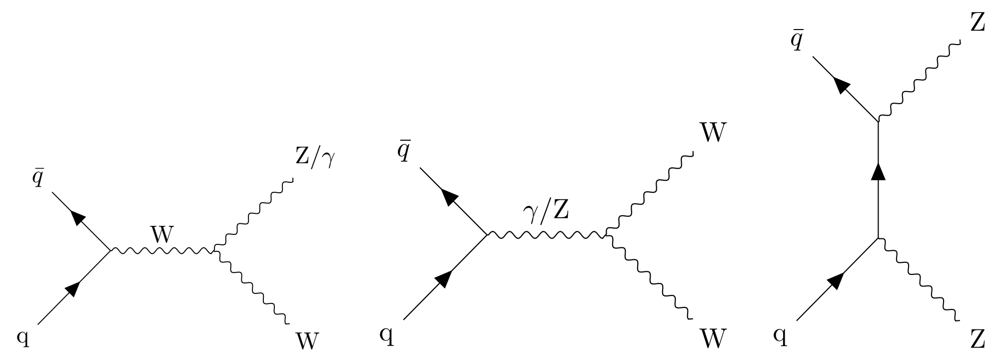


Figure 6.13: Possible Feynman diagrams for smaller backgrounds of this analysis: WW (on the left), Wγ and WZ (on the center) and ZZ (on the right).

CHAPTER 7

Signal extraction

In this analysis, two different signal regions are being used, targeting each one of our signals of interest and each based on the number of (b-)jets observed in an event. Many different discriminating variables are then first of all defined in Section **??**, allowing us to separate efficiently our signals from the different background processes that can show up, even after applying the full selection corresponding to both signal regions. The discriminating power of all these variables is then combined into a single variable through the use of a **BDT!** (**BDT!**) and a **ANN!** (**ANN!**) in each of these signal regions, and the output shape of the **MVA!** method is then simply used in order to perform a shape analysis and extract upper limits on the signal strength. This process will be described in Section **??**.

7.1 Discriminating variables

Several variables can be used in order to separate our signals of interest from the different background processes surviving our event selection. Some of the variables now presented did not feature a degree of discrimination as high as expected though and have been left behind in the actual analysis, but the ones featuring a discriminating power sufficiently high have actually been used as input to the **MVA!** technique developed. This allowed us to combine their discriminating power into a single output variable, using methods described in Section **??**.

It is also important to note that some of these variables rely on the information obtained from the performed top reconstruction of the $t\bar{t}$ system, which might fail in some cases, as described in Section **??**. When it happens, a default negative value is then assigned to the variables that need the neutrinos 4-momenta information in order to be computed, making it easy to select only events which actually pass this reconstruction algorithm. Note that such non-physical negative values were simply removed from the following distributions.

Number of b-jets and m_{bl}^t

Let's start with the two variables that we found to be the most useful in order to introduce some discrimination between the $t/\bar{t}$+DM and $t\bar{t}$+DM signals. Since both signatures can have a different number of top quarks in their final state, they are also expected to lead to a different number of observed b-jets, so this quantity is an obvious choice for a good discriminating variable. However, simply rejecting events featuring two b-jets or more to define a region enriched in $t/\bar{t}$+DM signal turns out in practice to be an ineffective strategy, mainly since the b-tagger efficiency for our chosen working point is of the order of 85% only, as discussed in Section **??**, which can result in a large surviving $t\bar{t}$+DM background in this hypothetical $t/\bar{t}$+DM enriched region, especially given the difference in cross-section between the two processes, as shown in Tables **??** to **??**.

A slightly more effective strategy than vetoing events with multiple b-jets consists in observing that if a b-jet is produced in a top-quark decay, its invariant mass is bounded from above by $\sqrt{m_t^2 - m_W^2} = 153$ GeV. Events compatible with two semileptonic top-quark decays can then be selected or rejected by introducing the observable m_{bl}^t, defined in Equation **??** [?], where the minimization is performed either over all the possible combinations of jets j_a, j_b among the b-jets of the events if three or more b-jets are observed, or otherwise over the b-jet(s) observed plus the non b-tagged jet having the highest b-tag weight of the event. The final distributions obtained for both variables in the pre-selection region are shown in Figure **??**.

$$m_{bl}^t = \min\left(\max(m_{l_1 j_a}, m_{l_2 j_b})\right) \tag{7.1}$$

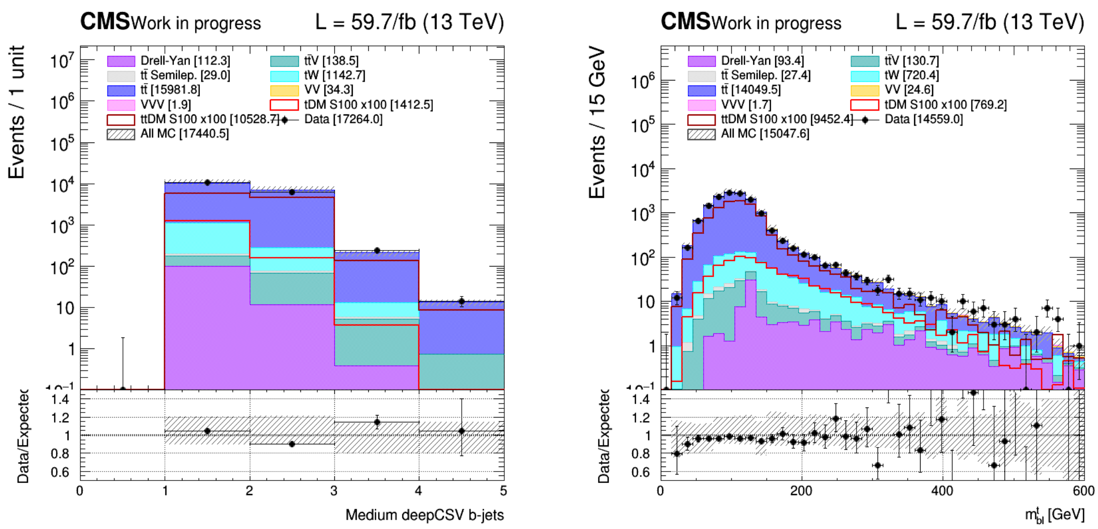


Figure 7.1: Number of b-jets (on the left) and m_{bl}^t (on the right) distributions in the 2018 pre-selection region.

Missing transverse momentum

Already defined in Section **??**, this variable corresponds to the imbalance in transverse momentum which can be left by different phenomena, such as the apparition of **SM!** neutrinos or the existence of **DM!** particles, able to escape the detector without being detected.

This is one of the most important variables of this analysis, expected to induce some discrimination between the signals and the backgrounds processes since, even tough some **SM!** processes such as the **SM!** $t\bar{t}$ production in the dilepton final state is also expected to produce some neutrinos and therefore some **MET!**, both the $t/\bar{t}$+DM and $t\bar{t}$+DM signal model are expected to have mostly the same contribution to the **MET!** from their own neutrinos, plus an additional contribution from the pair $\chi\bar{\chi}$ produced. The **MET!** distribution is in this sense expected to reach higher values for the signals than the backgrounds, as shown in Figure **??**.

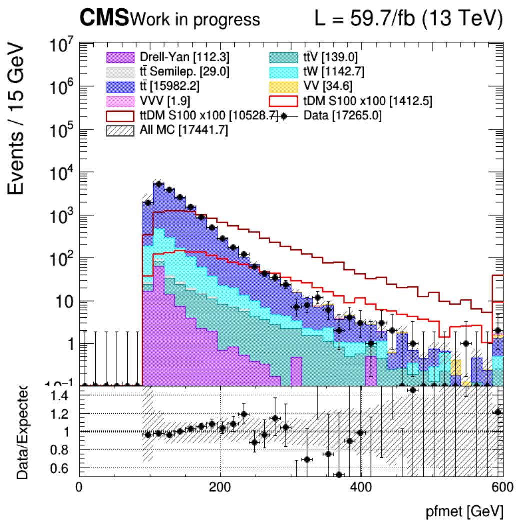


Figure 7.2: pf**MET!** distribution obtained in the 2018 pre-selection region.

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 flavour decay into one visible and one invisible particle, such as what happens in the double $W \to l\nu$ decay, for example.

For the signals considered in this work, two neutrinos contribute to the presence of **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 **??**, where $\boldsymbol{p_{T_i}} = \overrightarrow{p_{T_i}}$ is the (visible) transverse momentum of the particle i and α is the angle between the visible and invisible p_T of the particles involved in the decay considered [**?**].

$$\begin{cases} M_{T2}^2 = \min\limits_{\not{p}_{T_1}+\not{p}_{T_2}=\not{p}_{T_{\text{tot}}}} \left(\max\left(m_T^2(\boldsymbol{p_{T_1}}, \not{p}_{T_1}), m_T^2(\boldsymbol{p_{T_2}}, \not{p}_{T_2}) \right) \right) \\ m_T^2(\boldsymbol{p_T}, \not{p}_T) = 4\ |\boldsymbol{p_T}||\not{p}_T|\ \sin^2\left(\dfrac{\alpha}{2}\right) \end{cases} \tag{7.2}$$

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 possible value for the transverse mass.

In this particular analysis, M_{T2}^{ll} is calculated from a general algorithm described in [**?**], since the role of the visible particles is played by the two final state leptons. This variable is expected to introduce some discrimination because 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 **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. In practice however, 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.

A second variable, M_{T2}^{bl}, the stranverse mass of the b-jet lepton/pairs can also be defined in a similar way, except that in this case, the lepton is paired with a b-jet. The b-jet/lepton permutation giving the smallest value of M_{T2}^{bl} is kept, as shown in Figure **??**.

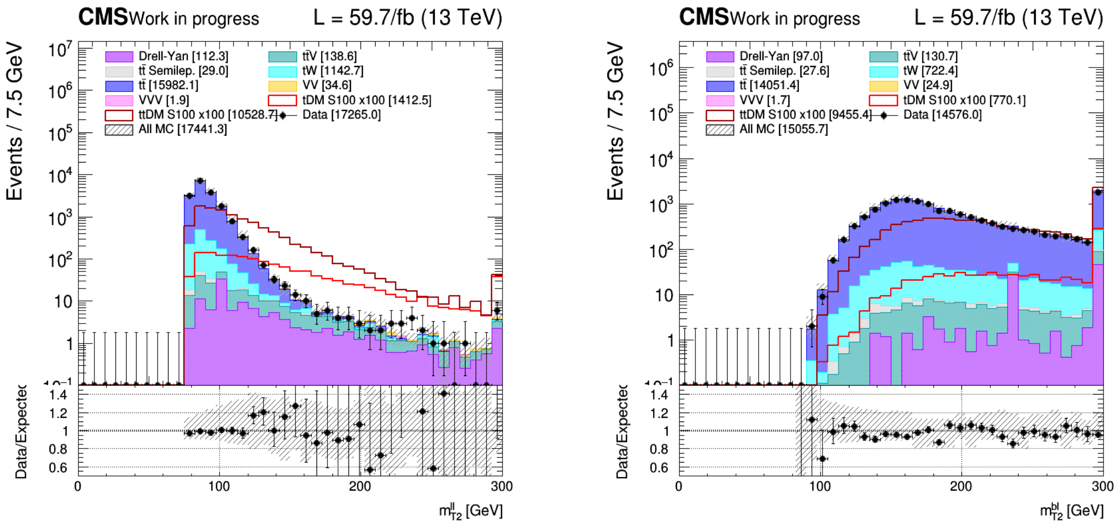


Figure 7.3: M_{T2}^{ll} (on the left) and M_{T2}^{bl} (on the right) distributions obtained in the 2018 pre-selection region.

Dark p_T and overlapping factor R

Already introduced with the top reconstruction method in Section **??**, these two variables are by construction expected to give some discrimination between the signals and the different background processes, mainly the dominant **SM!** $t\bar{t}$.

On one hand, the dark p_T is for example expected to take slightly higher values for the signals, where we expect the ellipses to be further apart from each other due to the production of **DM!**. On the other hand, the overlapping factor R defined in Equation **??** is expected to peak around 1 for the standard $t\bar{t}$ process (since in this case, we expect to have $d = l_1 + l_2$ for most of the events, as seen in Figure **??**), and to take slightly lower values for the different signal mass points considered. This intuition is confirmed from the distributions shown in Figure **??**.

Additional variables

A few additional variables which are expected to introduce some discrimination between the signals and the backgrounds have been considered as well. Among them, we can for example quote:

- $\Delta\Phi(E_T^{\mathrm{miss}}, ll)$: the difference in Φ between the two leptons and the **MET!**, expected to change depending on the eventual production of **DM!**, as shown in Figure **??**;
- The kinematic reconstruction weight w obtained from the $t\bar{t}$ reconstruction is also expected to give us some discrimination, for reasons already explained in Section **??**;

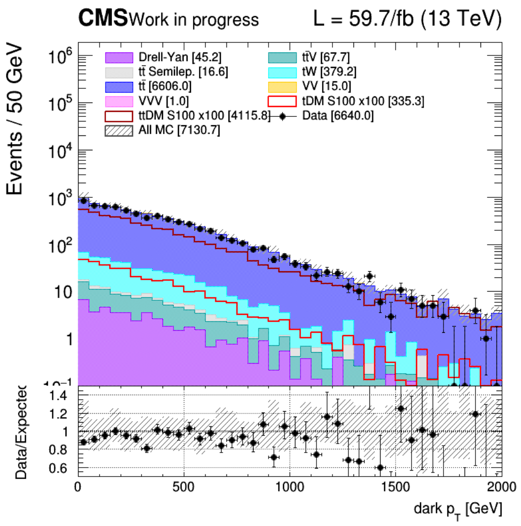


Figure 7.4: Dark p_T (on the left) and overlapping factor R (on the right) distributions obtained in the 2018 pre-selection region.

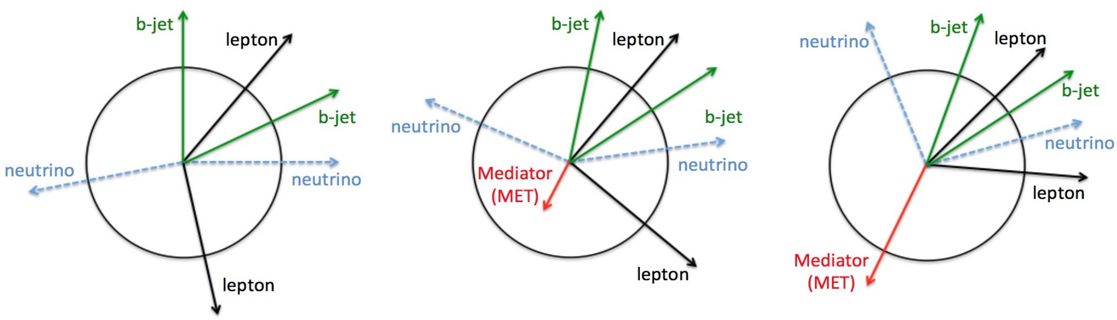


Figure 7.5: Schematic representation in the Φ plane of the distribution of the particles for the **SM!** $t\bar{t}$ process (on the left) and for the $t\bar{t}$+DM signal (on the center and right).

- Two other interesting variables previously used by the **ATLAS!** collaboration [**?**] are the so-called $r2l$, defined as the ratio between the pf**MET!** and the p_T of the two leptons observed and $r2l4j$ variables, defined in a similar way but considering additionally the p_T of the first 4 jets (if they exist) in the sum in the denominator;

- We also observed that the variable massT which corresponds to the scalar sum of the transverse component of the pf**MET!**, the two leptons and the two b-jets obtained by the top reconstruction process helps with the discrimination process;

- Finally, we know that spin correlation in a $t\bar{t}$-like event is expected to be conserved, because of the short lifetime of the top quark, and can actually be inferred from the top quark decay products, made accessible to us by the top reconstruction method. These so-called spin correlated variables are therefore interesting because the spin correlation in such events depend on the production mechanism and will be influenced by the additional coupling to a scalar or pseudoscalar mediator. Several spin correlated variables have been considered in this case, such as $\xi = \cos(\theta_l)\cos(\theta_{\bar{l}})$ and c_{hel}, defined as the full opening angle between the two leptons in their parent mass frame.

The distributions corresponding to all these variables in the 2018 pre-selection region of the analysis can be found in Figure **??**.

7.2 Multivariate analysis

As already seen in the previous sections, we know that our signal regions are going to be contaminated with some background processes, given the fact that such backgrounds have similar kinematics. A general **MVA!** (**MVA!**) technique has therefore been used in order to perform the actual signal extraction by combining the discriminating power of several of the variables previously presented in order to find a way to categorize reliably any single event as either more likely to be a signal (either $t/\bar{t}$+DM or $t\bar{t}$+DM) or a background event.

The idea behind this kind of techniques is quite simple. We have at our disposal on one hand **MC!** simulations of the most relevant background and signal processes and on the other hand, we have data collected by the detector and which needs to be categorized and labeled based on the value of several variables or *features*, by assigning a value of probability of belonging to any of these categories to every single event. By definition, these **MC!** simulations do come with a label already assigned, making this a simple supervised learning problem that we need to solve. The full set of events at our disposal can be divided into two subsets: one used for the adjustment of different internal parameters of the model used (the **training process**), while the other subset is used in order to evaluate the performance of the classifier that has been defined (the **testing process**). Finally, the model developed will be used by making predictions allowing us to categorize unlabeled data (the **evaluation process**).

Methods considered

Two different **MVA!** methods among the most popular nowadays have been studied in this work: several **BDT!**s (**BDT!**s) and **ANN!**s (**ANN!**s) have been defined and optimized separately in order to get the best discriminating power possible. Details regarding this optimization process can be found in Appendix **??**.

BDT!s (BDT!s)

This method relies on the definition of different decision trees, able to split the data recursively from a set of input variables or **features**. A typical **BDT!**, as shown in Figure **??**, is then made out of several **nodes**, allowing to make a binary decision (yes/no) and split the data based on a given feature until reaching a terminal condition and therefore a **leaf**, terminal node representing a class label or a probability. The split performed is obviously not chosen randomly, as a thorough training process needs to take place in order to define the optimal splitting based on the features given in such a way to maximize information gain. The **boosting** process then consists in training several trees and then to combine all of these trees together into a single strong classifier, by giving a weight to each tree depending on their overall accuracy. In our case, a classification is performed, by defining a **BDT!** able to read several input variables at once, create hundreds of trees made out of different layers of nodes, train them and apply the training to uncategorized events, labeling every single event as either signal or background-like.

The structure and complexity of a **BDT!** usually depends on the problem considered and can be

BDT parameter	Optimized value
Maximum depth	4
Minimum samples per leaf	2%
Loss function	Quadratic
Boost algorithm	Gradient descent
Shrinkage	0.3
Grid points n_{cut}	1000
Number of trees	250

Table 7.1: Summary of the parameters used for the training of the **BDT!**s in this analysis.

characterized from several different parameters which need to be tuned in order to get the best possible signal extraction accuracy while avoiding under- and overfitting:

- First of all, the **features**, or input variables, allowed to be used by the **BDT!** are extremely important. The use of a **maximum number of features** can be applied when building a tree, forcing a random selection of input variables if too many variables are given as input;
- The **maximum depth** characterizes the maximum number of vertical levels allowed and is usually chosen as a small number (which depends on the number of training events available);
- The **minimum number of samples per leaf** puts a requirement on the minimum number of samples required in order to create a new leaf;
- Regarding the boosting process itself, which consists in training several different **BDT!**s at once and to combine their individual predictions in order to enhance the accuracy of the overall predictions being made, comes with several additional parameters that can be optimized:
  - First, the actual **number of trees** used to define the ensemble of **BDT!**s is obviously an important parameter that needs to be optimized;
  - The **loss function** is helpful when trying to estimate the distance between the prediction made by the **BDT!** and the actual value expected;
  - The **learning rate**, or **shrinkage**, is another important parameter in the sense that it tells how much the weights of the tree should be adjusted after each training iteration. A small shrinkage typically demands more trees to be grown but can significantly improve the accuracy of the predictions made;
  - Finally, the **number of grid points** n_{cut}, defined as the granularity used to find the optimal cut when determining the node splitting is also important.

The actual parameters used for this particular analysis have been obtained after a thorough optimization process and can be found in Table **??**.

ANN!s (ANN!s)

Another large family of **MVA!** methods rely on the so-called **neural networks**, sets of algorithms designed to be able to learn how to recognize patterns. Neural networks help us cluster any unknown dataset, allowing us to group and classify unlabeled data according to similarities among the inputs or according to the information available in a training dataset.

The principle at the basis of neural networks relies on the human brain itself, a complex system of thousands of billions of neurons interacting with each other and able to perform complex calculations and generate ideas. In this case, highly connected mathematical units called **neurons** and grouped into **layers** are defined (as shown in Figure **??**), connected with each other with connections (also called **edges**), each characterized by a **weight**, denoting their significance. At the end of the day, these neurons are simply mathematical tools able to receive an input, modify it slightly and fire the computed signal to the next layer of neurons. The weight of this connection to each of the neurons in the next layer will then either be increased or decreased based on the overall accuracy of the prediction that has been made, through an extensive training process.

Mathematically, each neuron i of an intermediate (also called *hidden*) layer receives an input x_j from the j-th neuron of the previous layer, and combines them by multiplying them with the weight ω_{ij} of the corresponding connection between the two neurons involved. This process is repeated for each of the n neurons of the previous layer by summing all the actual inputs and, once done, the neuron then applies an internal non-linear **activation function** f to the previous value, and fires the computed final output y_i to the next layer of neurons, as shown in Equation **??**. This process is repeated until reaching the output layer.

$$y_i = f\left(\sum_{j=0}^{n} \omega_{ij} x_j\right) \tag{7.3}$$

The objective of the training process in this case is to find the optimal weights $\boldsymbol{\omega}$ for each connection that are able to deliver an output as faithful to the known event category as possible. This faithfulness of a given set of weights can be obtained from the so-called accuracy in such a classification problem, defined as a metric used in order to measure how often the algorithm classifies a data point correctly and which needs to be maximized. Another to characterize the goodness of a model is through the definition of the loss function which, unlike the accuracy, is not a percentage but a summation of the errors made for each instance in the training set. In this sense, this function can take different forms but usually simply computes the normalized sum of the difference between the expected $\hat{y_i^p}$ and obtained y_i^p values for each neuron i and each of the p events found in the training dataset, as shown in Equation **??**.

$$E(\boldsymbol{\omega}) = \sum_{i,p} \left(y_i^p(\boldsymbol{\omega}) - \hat{y_i^p}\right)^2 \tag{7.4}$$

This loss function allows us to determine the goodness of the output obtained from a given set of weights, but we still need a way to update these weights at any given training iteration to make this method useful. This can be done using different methods, such as by applying the back-propagation or performing a gradient descent method, which consists in starting from a random set of weights $\boldsymbol{\omega}_0$ and modifying them in the direction of the gradient of the loss function E at each iteration k, as shown in Equation **??**, with a given learning rate η.

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k - \eta \nabla_{\boldsymbol{\omega}} E(\boldsymbol{\omega}) \tag{7.5}$$

In this work, we used models with several hidden layers, therefore sometimes referring it as deep neural networks, where the **depth** parameter is equivalent to the number of hidden layers. The use of such methods with lots of hidden layers is helpful because it allows to solve the feature

DNN parameter	Optimized value
Hidden layers neurons	20, 15, 15, 10
Activation functions	Relu (x3), softmax (output)
Error function	Mean square error
Optimizer	Adam
Learning rate	0.001
Training epochs	250
Batch size	250

Table 7.2: Summary of the parameters used for the training of the **ANN!** in this analysis.

engineering problem, which states that to get the most of a certain dataset one should previously extract by hand the representations/features helping to solve the problem. However, with deep learning, the features are learned on the fly, without the need of doing a previous transformation by hand. In summary, the **ANN!**s allow to study phenomena a bit more disrupt, but are usually more sensitive to the overfitting even though some options, such as the dropout, dropping randomly some weights in the network at each training iteration, have been used in this work to mitigate this issue.

Several parameters are important to optimize during the training of a neural network, such as the **number of neurons in each layer**, the **activation** and **loss functions**, the **learning rate** (giving an idea of how much the weights should be updated after each training iteration, to be optimized in order to reach the minimum of the loss function faster while avoiding getting stuck in an eventual local minimum), the number of **training epochs** and the **batch size**, allowing to reduce the time required for the training by dividing the dataset into mini-batches and updating the weights for each mini-batch instead of doing it for all the features of every single event. After a complete optimization process extensively described in Appendix **??**, the actual parameters used for this particular analysis can be found in Table **??**.

Training process

One training has been done for each year and each mediator mass, considering separately the scalar and pseudoscalar mediators in order to enhance the sensitivity of the analysis. Each **BDT!** and **ANN!** has been trained in order to reduce as much as possible the dependence on the sample itself, by weighting the samples according to the specific cross-section of each mass point.

The two signals for a given mass point have been mixed together and trained against the backgrounds which tend to look more like our signals (typically, the **SM!** $t\bar{t}$ and single top), in the pre-selection region defined in Section **??**. Only events passing this pre-selection cuts were actually considered for the training, and two different trainings were performed for each mass point, in order to match our two signal regions:

- One targeting the $t/\bar{t}$+DM signal, by considering events having exactly 1 jet, or exactly 2 jets and 1 b-jet;
- And another one targeting the $t\bar{t}$+DM signal, by considering this time only the events having exactly 2 jets and more than 1 b-jet, or more than 2 jets.

The TMVA package, a toolkit for Multivariate Data Analysis with ROOT [?], was used in order to define both the **BDT!** and the **ANN!** previously defined and to perform the optimization

	$t/\bar{t}$+DM region		$t\bar{t}$+DM region	
Rank	Variable	Importance	Variable	Importance
1	M_{T2}^{ll}	$4.17 \cdot 10^{-1}$	M_{T2}^{ll}	$3.92 \cdot 10^{-1}$
2	E_T^{miss}	$3.41 \cdot 10^{-1}$	E_T^{miss}	$3.14 \cdot 10^{-1}$
3	massT	$1.28 \cdot 10^{-1}$	massT	$2.28 \cdot 10^{-1}$
4	$r2l4j$	$1.14 \cdot 10^{-1}$	$r2l4j$	$1.91 \cdot 10^{-1}$
5	m_{bl}^t	$6.12 \cdot 10^{-2}$	m_{bl}^t	$9.35 \cdot 10^{-2}$
6	Dark p_T	$1.59 \cdot 10^{-2}$	nbJet	$1.69 \cdot 10^{-2}$
7	Factor R	$1.32 \cdot 10^{-2}$	$r2l$	$1.22 \cdot 10^{-2}$
8	Weight w	$1.27 \cdot 10^{-2}$	Dark p_T	$9.20 \cdot 10^{-3}$

Table 7.3: Ranking for the importance of the main variables used as input for the **BDT!** and **ANN!**, considering the scalar 100 GeV training, in both signal regions.

	$t/\bar{t}$+DM region		$t\bar{t}$+DM region	
Rank	Variable	Importance	Variable	Importance
1	M_{T2}^{ll}	$5.86 \cdot 10^{-1}$	M_{T2}^{ll}	$5.74 \cdot 10^{-1}$
2	E_T^{miss}	$5.48 \cdot 10^{-1}$	E_T^{miss}	$5.29 \cdot 10^{-1}$
3	massT	$3.38 \cdot 10^{-1}$	massT	$4.09 \cdot 10^{-1}$
4	$r2l4j$	$1.92 \cdot 10^{-1}$	$r2l4j$	$3.21 \cdot 10^{-1}$
5	m_{bl}^t	$6.97 \cdot 10^{-2}$	m_{bl}^t	$1.43 \cdot 10^{-1}$
6	$r2l$	$2.79 \cdot 10^{-2}$	Dark p_T	$3.83 \cdot 10^{-2}$
7	Dark p_T	$1.57 \cdot 10^{-2}$	$r2l$	$3.54 \cdot 10^{-2}$
8	Factor R	$1.37 \cdot 10^{-2}$	$\Delta\Phi(E_T^{\text{miss}}, ll)$	$1.12 \cdot 10^{-2}$

Table 7.4: Ranking for the importance of the main variables used as input for the **BDT!** and **ANN!**, considering the scalar 500 GeV training, in both signal regions.

of the different hyperparameters of both methods. Such package is extremely helpful because it automatically gives us plenty of information when training a specific model for a given mass point, such as a ranking of the importance of the input variables, given in Tables **??** and **??** for each signal region, a plot representing the correlation between these input variables for the backgrounds and the signals, in Figure **??**, and the **ROC!** (**ROC!**) curves showing the background rejection achievable for a given selected signal efficiency, in Figures **??** to **??**.

All these plots allowed us to conclude that the **BDT!** seems to be able to achieve a better signal efficiency for a given background rejection value for all the mass points and mediators considered in this analysis, which is what we are interested in. We therefore decided to use this method for the rest of the actual analysis instead of the **ANN!**.

Finally, the output distributions obtained for the **BDT!** have been obtained in Figures **??** and Figures **??** for both training regions. All these results have been obtained by splitting the dataset available as 70% for the training process and the remaining 30% to test the classifier obtained, giving us between 50.000 and 60.000 training events for each mass point.

	$t/\bar{t}$+DM region		$t\bar{t}$+DM region	
Process	Yields	Percentage	Yields	Percentage
$t\bar{t} \rightarrow 2l$	7344.6	87.9%	8637.4	95.0%
Single top	846.4	10.2%	296.3	3.3%
DY	78.4	1.0%	33.9	0.4%
$t\bar{t} + V$	41.7	0.5%	96.9	1.0%
$t\bar{t} \rightarrow 1l$	9.6	0.1%	19.5	0.2%
VV	26.6	0.3%	7.9	0.1%
VVV	1.0	0.0%	0.9	0.0%
Total	8348.2	///	9092.9	///
Data	8808	///	8457	///

Table 7.5: Number of events of all the **MC!** processes in the 2018 signal regions.

Evaluation process

Once the training performed and the weights of the **BDT!** computed, we have been able to use them in order to to estimate the probability of each unknown data and **MC!** event to be a background or a signal event. This is simply done by reading the weights previously obtained by the training process, considering either the training that was performed in the $t/\bar{t}$+DM or $t\bar{t}$+DM enriched region, depending on the number of (b-)jets observed in the event, by defining similar categories than the ones used for the training process.

Once the weights were applied to our trees, the actual **BDT!** output distributions for each mediator mass point an each signal category have been obtained, as shown in Figures **??** to **??**. The yields of the different processes observed in both the 2018 signals regions can also be found in Table **??**, in order to get a deeper understanding of the backgrounds we are dealing with.

Shape analysis

A general shape analysis has finally been performed in order to extract the signals of interest from the shape of the output of the **MVA!**. In order to perform such an analysis, the binning of **BDT!** distributions in the signal regions was first of all optimized in order to make sure that each bin of the distribution contains a significant number of **MC!** events. Then, once this optimal binning has been defined, a maximum likelihood fit has been simply performed in order to extract the actual upper limits on both the signals production cross-sections, as explained in Chapter **??**.

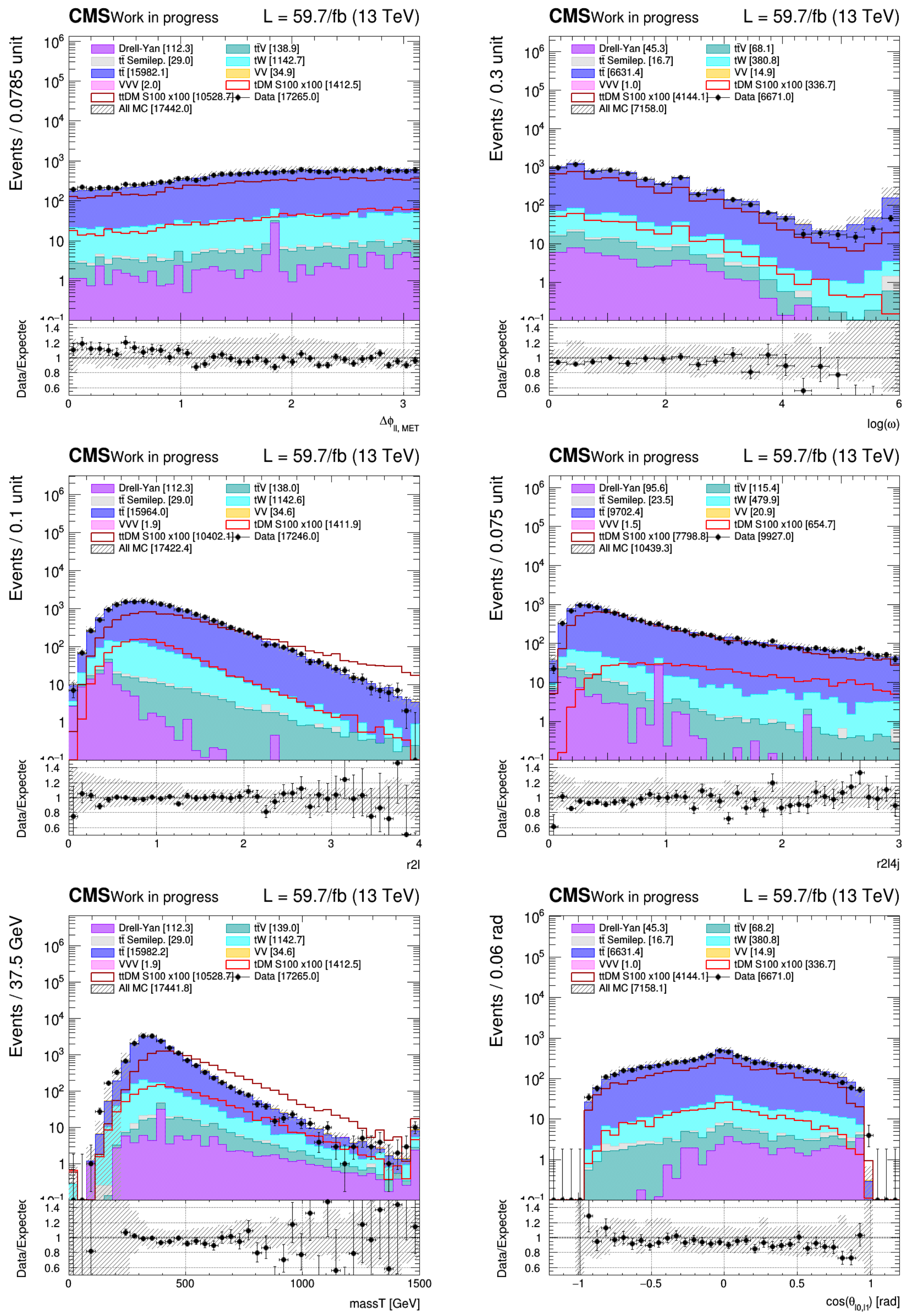


Figure 7.6: Some additional discriminating variables represented in the 2018 pre-selection region.

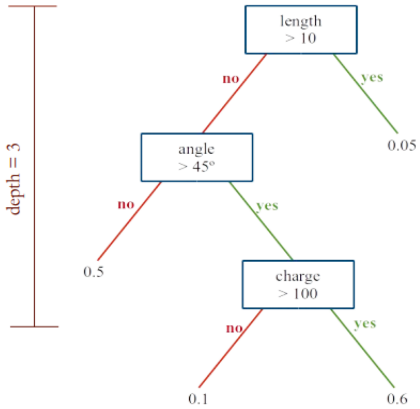


Figure 7.7: Schematic representation of a typical **BDT!**, with its nodes (represented by the blue boxes) and leaves [**?**].

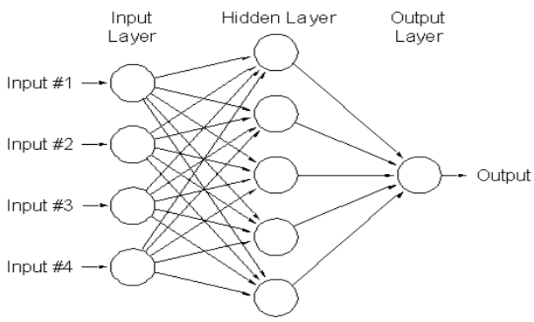
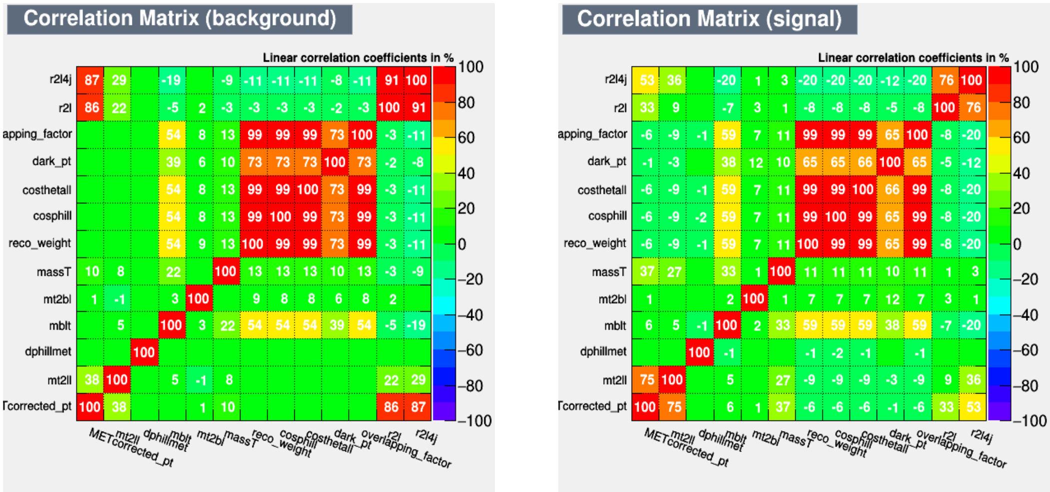
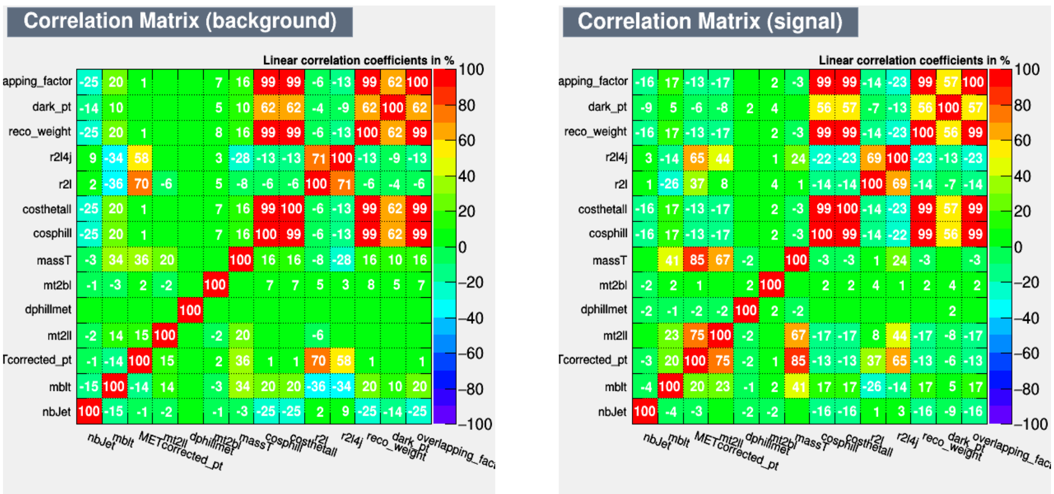


Figure 7.8: Schematic representation of a typical neural network with a single hidden layer [**?**].

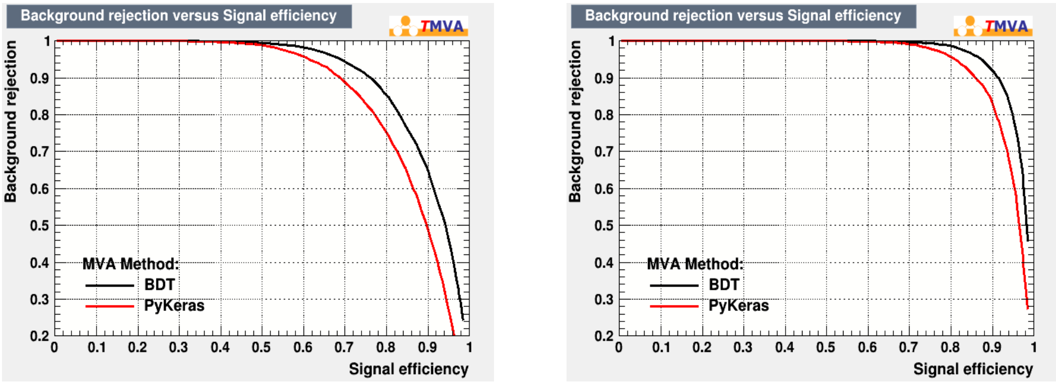


(a) $t/\bar{t}$+DM training region

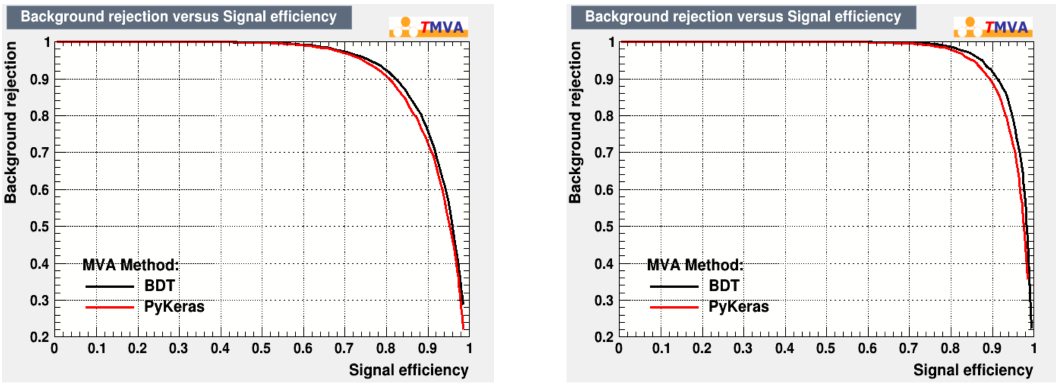


(b) $t\bar{t}$+DM training region

Figure 7.9: Correlation coefficients between the different variables used as input for the **MVA!** for the backgrounds (on the left) and the signals (on the right), in both training regions.

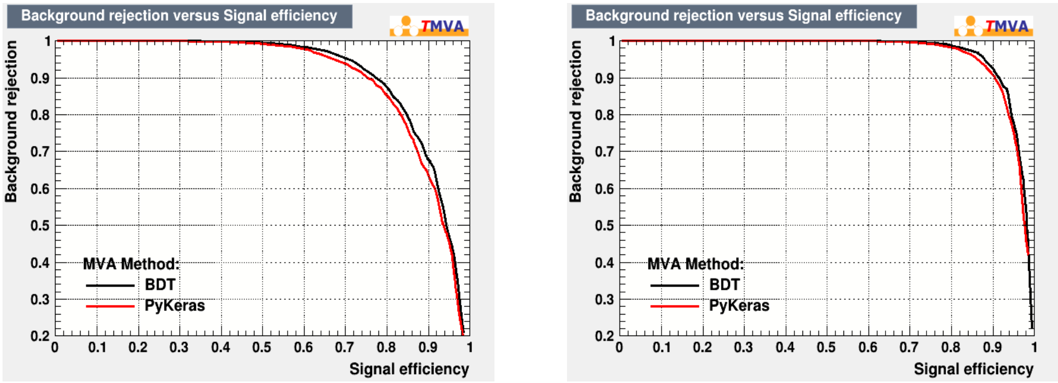


(a) Scalar mediators

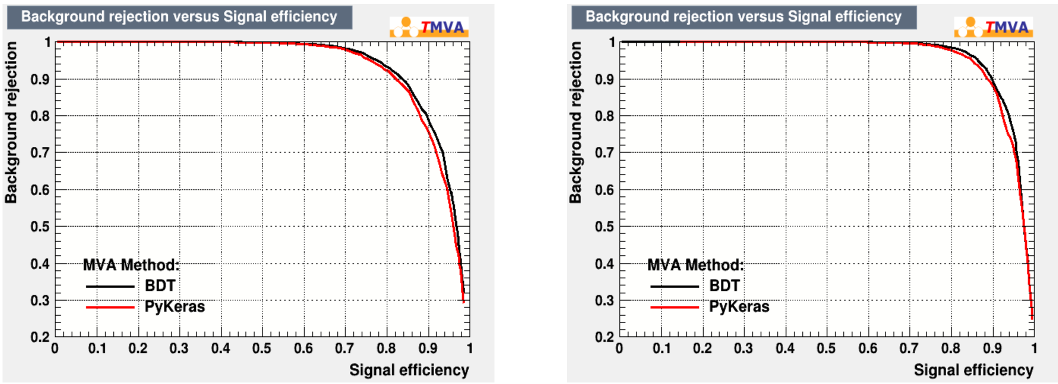


(b) Pseudoscalar mediators

Figure 7.10: Signal versus background **ROC!** curves for the **BDT!** and the **ANN!** obtained after the training performed, considering 100 (on the left) and 500 GeV (on the right) mediators, in the $t/\bar{t}$+DM training region.

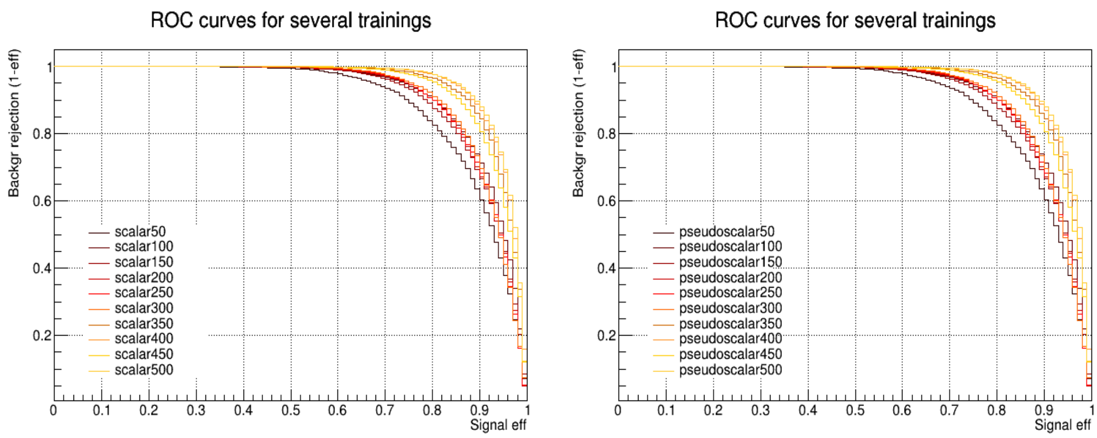


(a) Scalar mediators

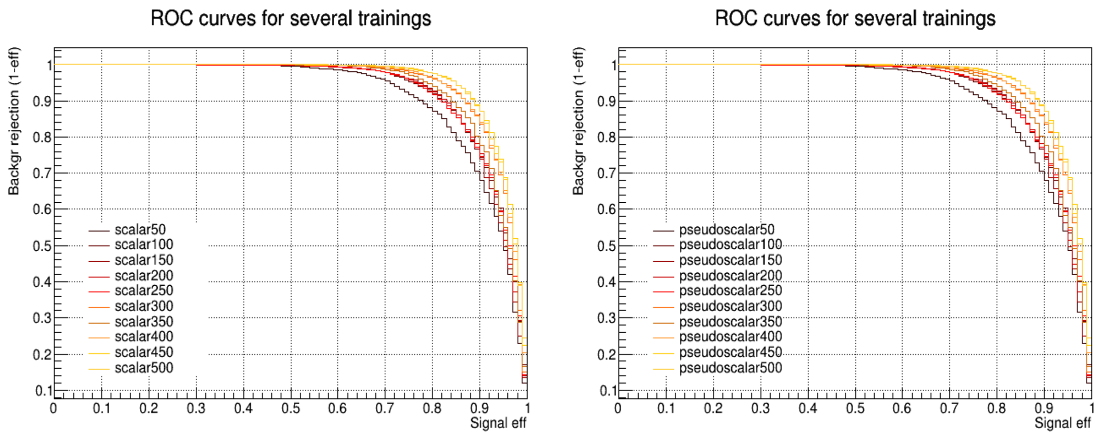


(b) Pseudoscalar mediators

Figure 7.11: Signal versus background **ROC!** curves for the **BDT!** and the **ANN!** obtained after the training performed, considering 100 (on the left) and 500 GeV (on the right) mediators, in the $t\bar{t}$+DM training region.

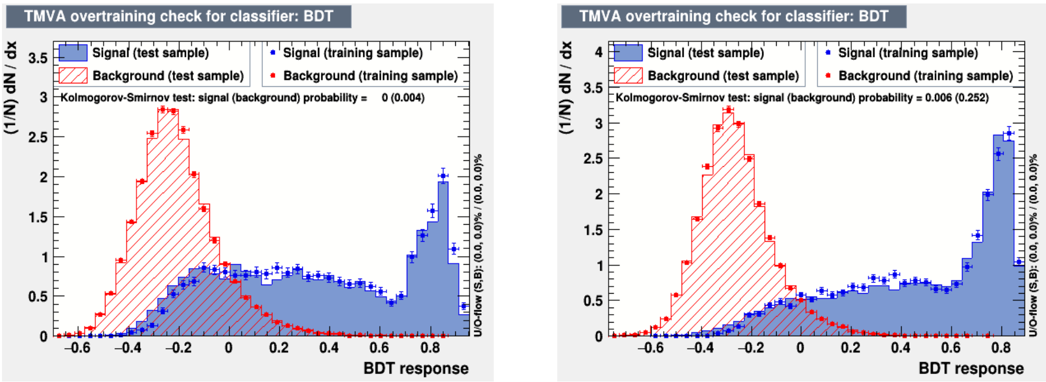


(a) $t/\bar{t}$+DM signal region

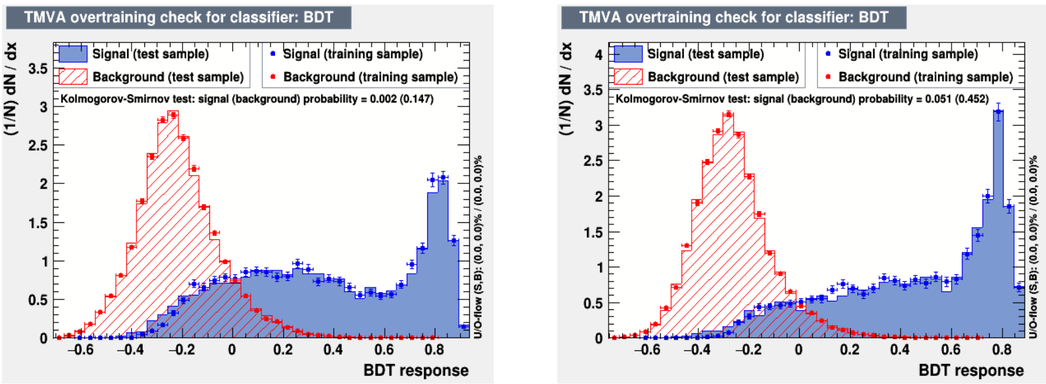


(b) $t\bar{t}$+DM signal region

Figure 7.12: Signal versus background **ROC!** curves obtained after the training performed with the **BDT!**, considering all the scalar (on the left) and pseudoscalar (on the right) mediators available, in both signal regions.

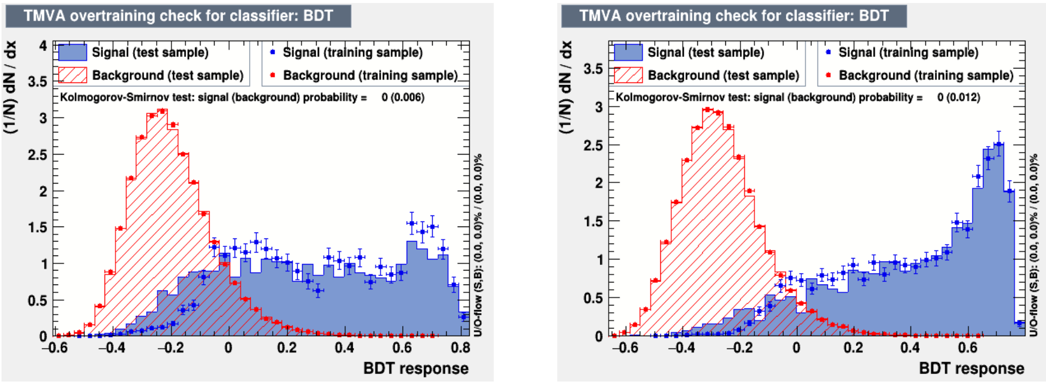


(a) Scalar output distribution

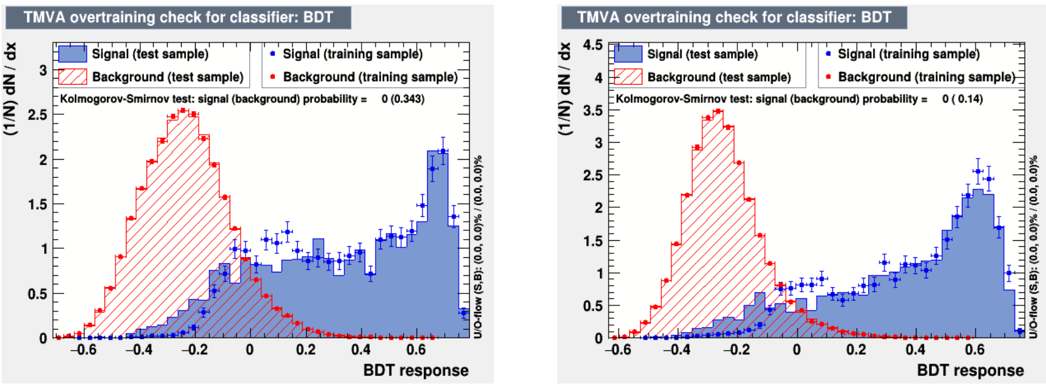


(b) Pseudoscalar output distribution

Figure 7.13: Output distributions obtained in the $t/\bar{t}$+DM training region for the **BDT!**, considering the 100 (on the left) and 500 GeV (on the right) mediators.

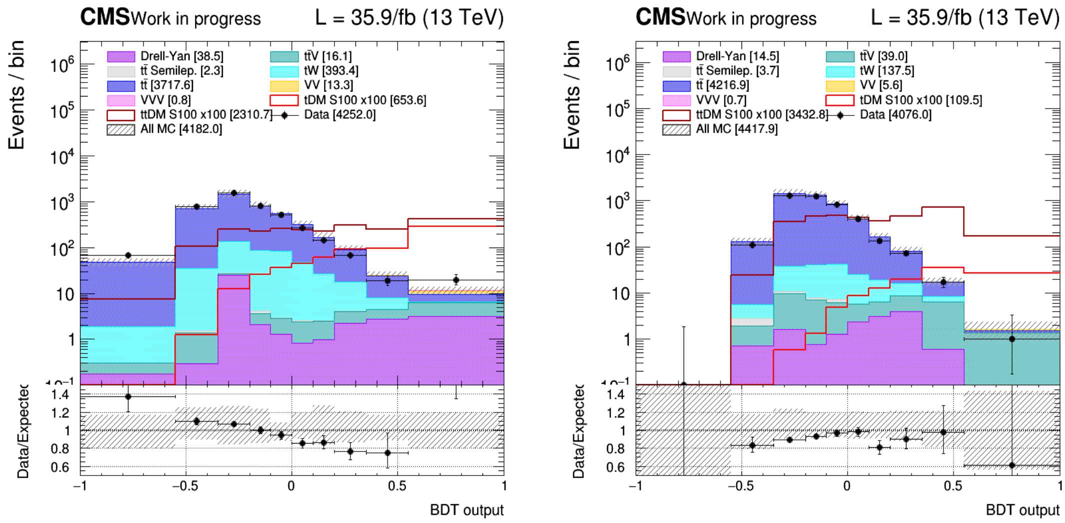


(a) Scalar output distribution

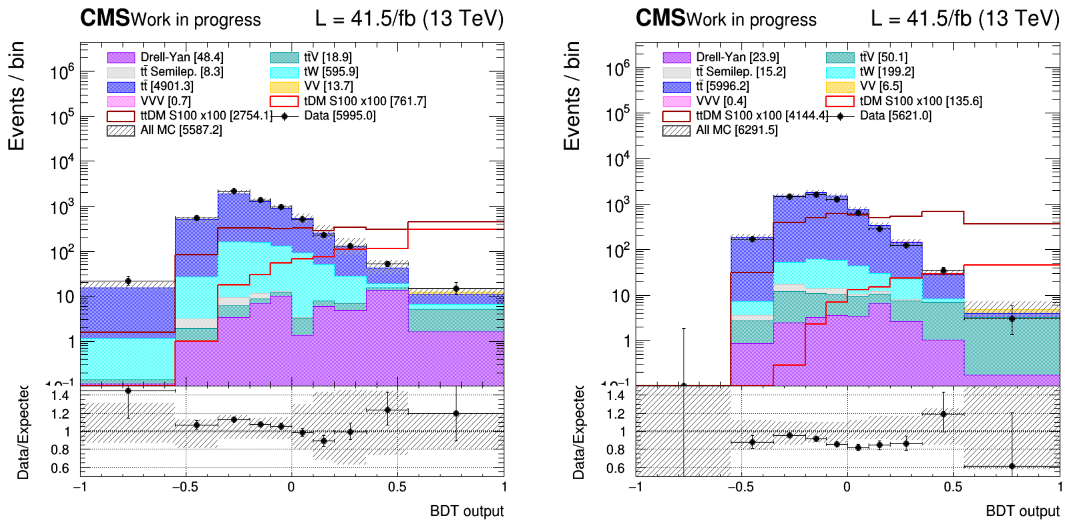


(b) Pseudoscalar output distribution

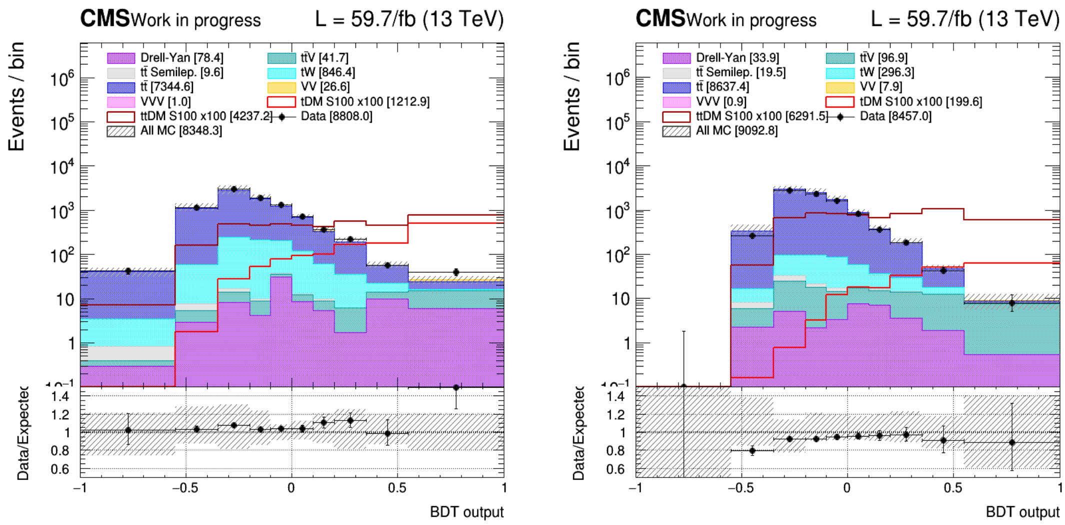
Figure 7.14: Output distributions obtained in the $t\bar{t}$+DM training region for the **BDT!**, considering the 100 (on the left) and 500 GeV (on the right) mediators.



(a) 2016 distributions

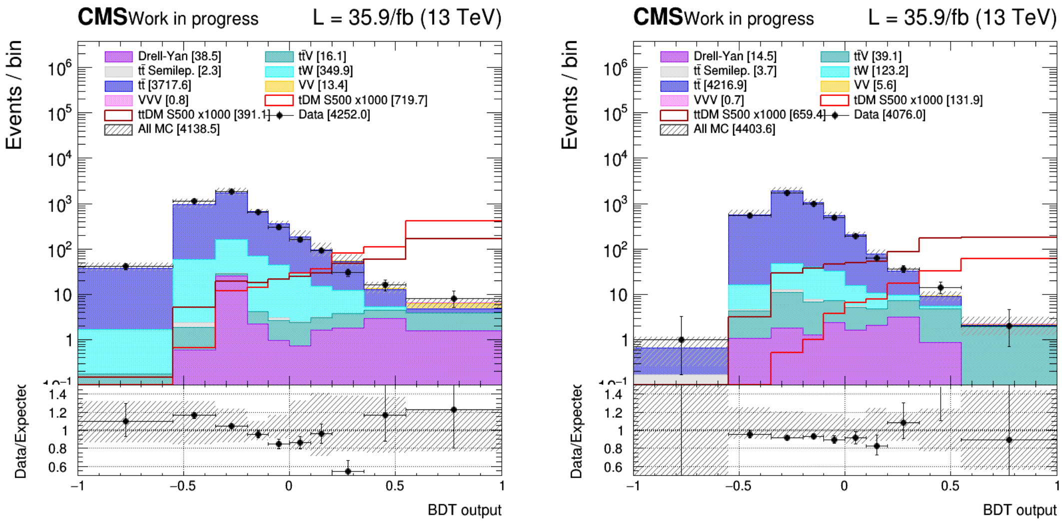


(b) 2017 distributions

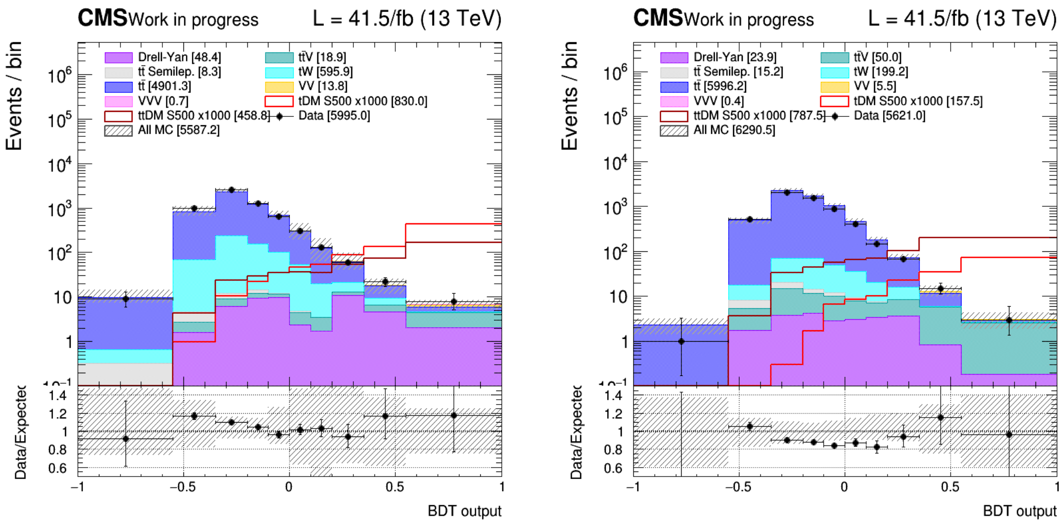


(c) 2018 distributions

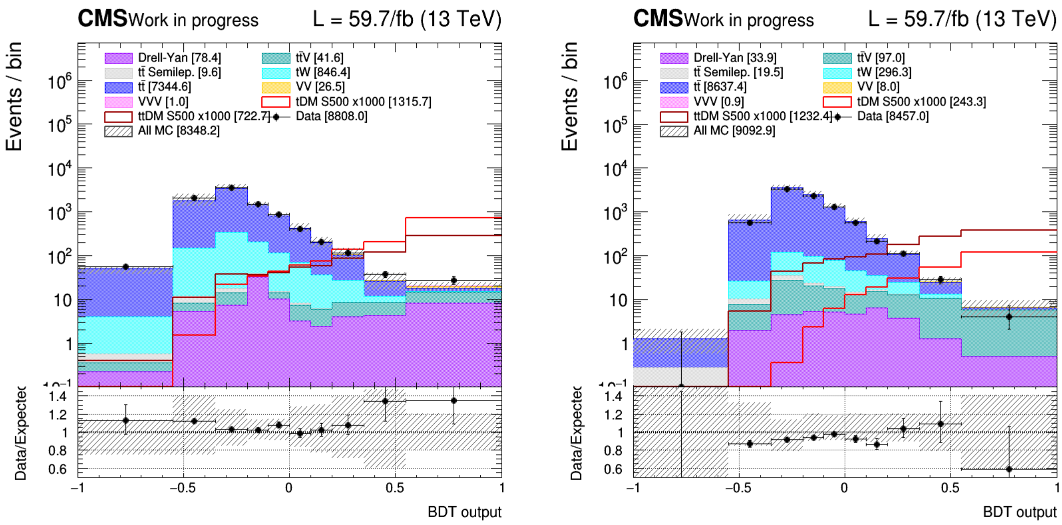
Figure 7.15: **BDT!** output distribution for the scalar 100 GeV training in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.



(a) 2016 distributions

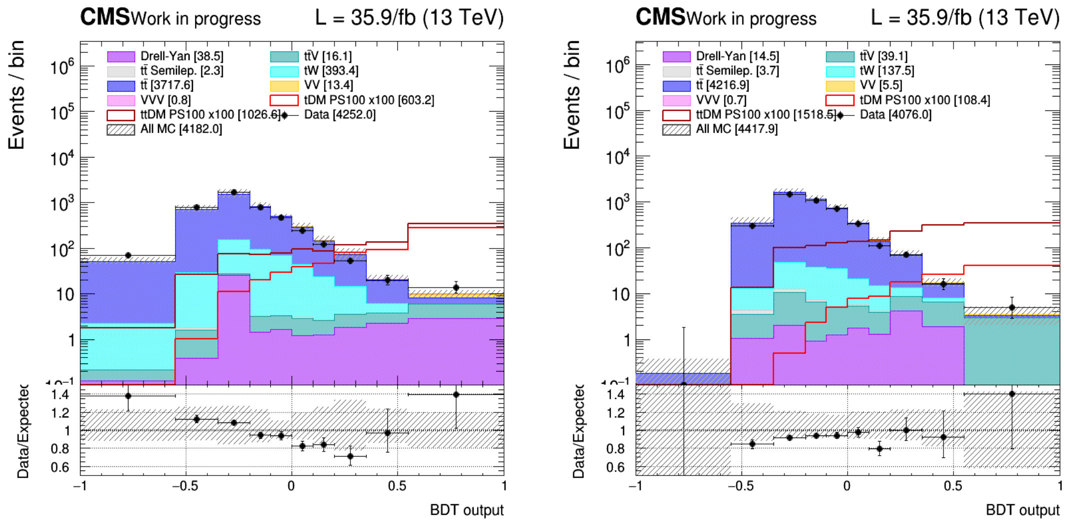


(b) 2017 distributions

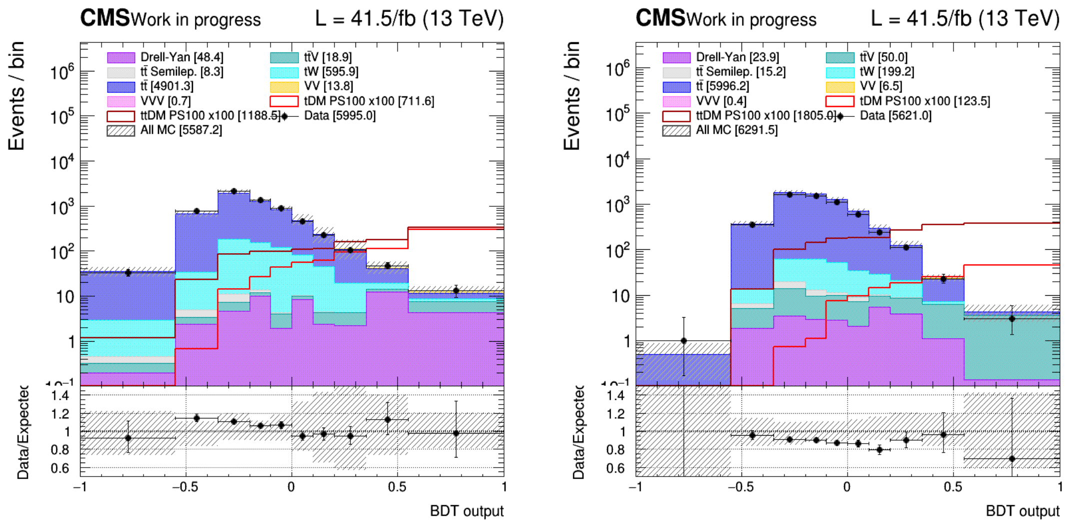


(c) 2018 distributions

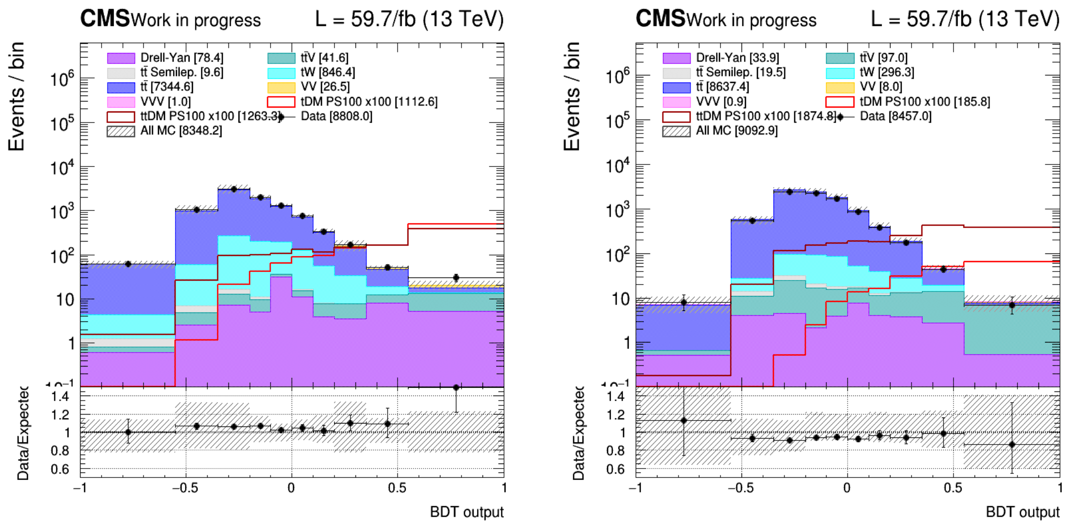
Figure 7.16: **BDT!** output distribution for the scalar 500 GeV training in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.



(a) 2016 distributions

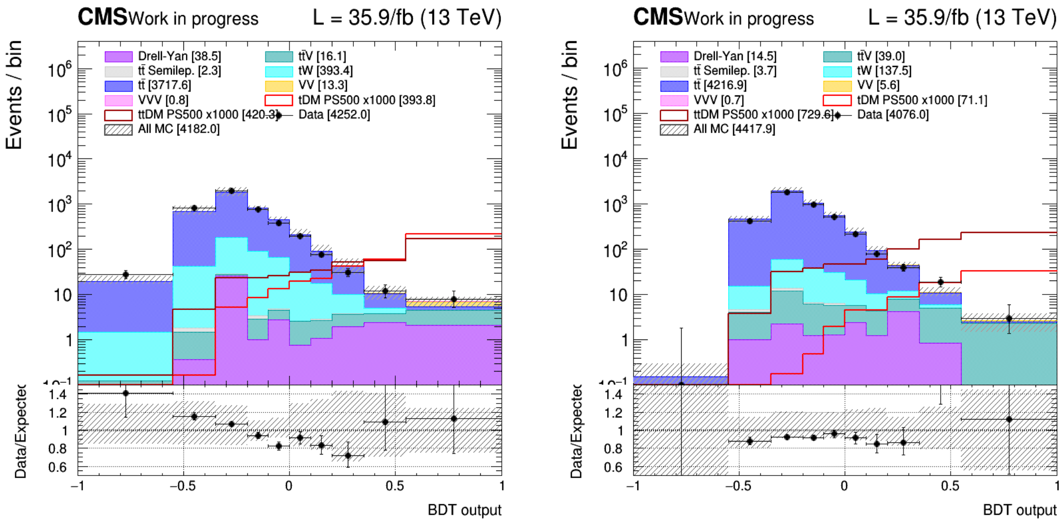


(b) 2017 distributions

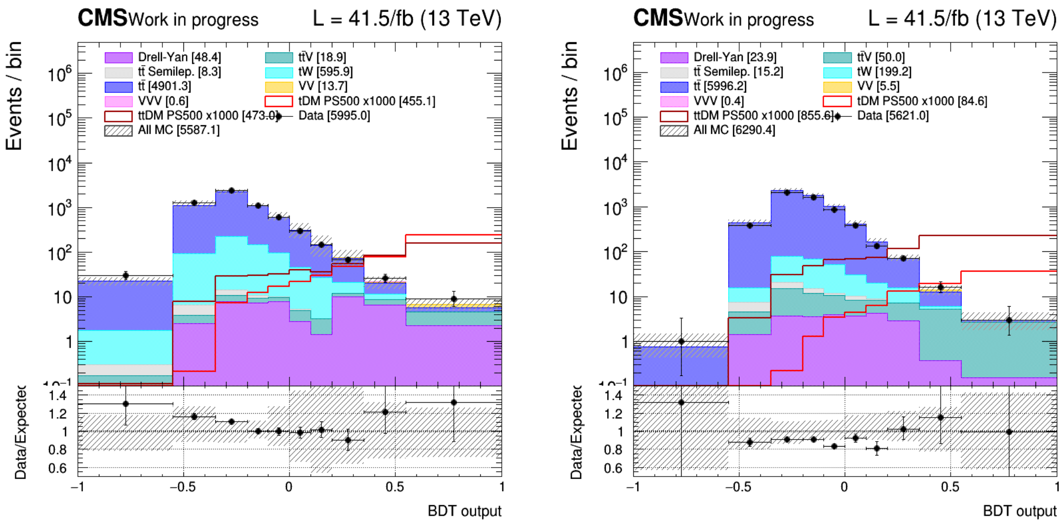


(c) 2018 distributions

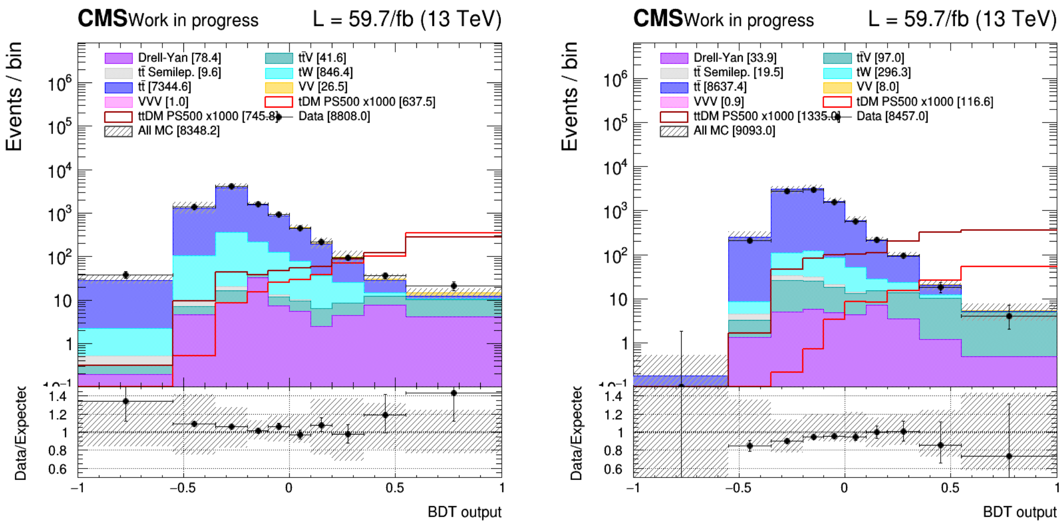
Figure 7.17: **BDT!** output distribution for the pseudoscalar 100 GeV training in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.



(a) 2016 distributions



(b) 2017 distributions



(c) 2018 distributions

Figure 7.18: **BDT!** output distribution for the pseudoscalar 500 GeV training in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.

CHAPTER 8

Results and interpretations

We now have all the ingredients needed to proceed with the interpretation of the data in terms of exclusion limits on the signal strength, for the different mediator categories, mass points and signal models considered. In this chapter, a short but necessary theoretical introduction to the statistical concepts used will be performed in Section **??**, before discussing about the impact of the systematics and uncertainties on the final results in Section **??**. Finally, the results and final limits obtained for this particular analysis will be shown in Section **??**.

8.1 Statistical interpretation

The results presented in this chapter follow the recommendations and use the tools given directly by the **LHC!** Higgs Combination Group [**?**], result of a collaboration between teams of **ATLAS!** and **CMS!**. The exclusion limits on the signal strength presented here have been calculated according to the asymptotic formulas described in this section, allowing us to quantify how sensitive the experiment is to the new physics searched for by calculating both the expected and observed significance obtained for a variety of signal hypotheses, corresponding to different mediator categories and mass points, as described previously.

Setting limits is typically performed using a frequentist statistical test, which relies on the definition of two hypotheses [**?**, **?**]:

- The **null hypothesis** H_0, or b, according to which the signal searched for does not exist and does not contribute to the data observed in a given part of the phase space.
- And the **alternative hypothesis** H_1 tested against H_0 and stating exactly the opposite,

mainly that our signal does actually exist. Since the typical search is not free of background, this hypothesis is usually called $s + b$.

In practice, we can usually not give a definite true or false answer to the question whether the alternative hypothesis is correct or not: we actually need to compute and quote a **CL! (CL!)** and a **confidence limit**, corresponding to the value of a population parameter excluded at a specified **CL!**, for the exclusion. The objective is in this sense to quantify the level of agreement of the observed data with a given hypothesis H by computing the so-called **p-values**, corresponding to the probability of finding data of equal or greater incompatibility with the predictions of H, under the assumption of H. This hypothesis H can then be disregarded if its p-value is observed below a specified threshold, typically set to 0.05, resulting in a **CL!** common for the exclusion limits of most of the **BSM!** searches performed at **CERN!**.

In order to establish an eventual discovery or exclude certain models, this analysis then uses a quite typical frequentist significance test using a likelihood ratio, considering parameters of interest, such as the cross-section of the signal process and nuisance parameters, such as the background normalizations, as discussed in Section **??**.

Mathematically, as shown in [**?**], this method can be understood by considering the x variable corresponding to any of the shape of the **BDT!** output distributions previously obtained, containing each a certain expected number of events per bin $E[n_i]$, given by Equation **??**, where b_i (s_i) is the number of background (signal) events in the *i-th* bin and μ is the so-called **signal strength**, a positive parameter characterizing the normalization of the signal under consideration and independent on the bin considered ($\mu = 0$ therefore corresponds to the background-only hypothesis and $\mu = 1$ is the nominal signal hypothesis).

$$E[n_i] = b_i + \mu s_i \tag{8.1}$$

The number of background and signal events in each bin can be expressed through Equation **??**, where b_{tot} (s_{tot}) correspond to the total number of background (signal) events, the function $f_b(x; \boldsymbol{\theta}_b)$ ($f_s(x; \boldsymbol{\theta}_s)$) are the **PDF!**s of the variable x considered for signal and background events and $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, b_{\text{tot}})$ represent all the nuisance parameters considered.

$$\begin{cases} b_i = b_{\text{tot}} \displaystyle\int_{\text{bin } i} f_b(x; \boldsymbol{\theta}_b) dx \\ s_i = s_{\text{tot}} \displaystyle\int_{\text{bin } i} f_s(x; \boldsymbol{\theta}_s) dx \end{cases} \tag{8.2}$$

To help constraining even more the nuisance parameters considered, control regions enriched in background where we expect to observe a low number of signal events if it were to exist in nature are usually defined. The expected number of events in each bin $E[m_j]$ can then be expressed in Equation **??** in a similar way, where u_j is a calculable quantities depending on the set of nuisance parameters $\boldsymbol{\theta}$. The definition of such control region is particularly helpful when constraining the background normalization parameters introduced in Section **??**.

$$E[m_j] = u_j(\boldsymbol{\theta}) \tag{8.3}$$

If we assume that the number of events in each bin is large enough, then we can define in Equation **??** the likelihood function as a product of Poisson probabilities for each of the N and M bins, by using the mean expected number of events ($E[n_i]$ and $E[m_j]$) and the actual number of events measured (n_i and m_j) in the signal and control regions.

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \left(\prod_{i=1}^{N} \frac{E[n_i]^{n_i}}{n_i!} e^{-E[n_i]} \right) \left(\prod_{j=1}^{M} \frac{E[m_j]^{m_j}}{m_j!} e^{-E[m_j]} \right) \tag{8.4}$$

This last relation is helpful because it allows us to test any value of μ, by considering the profile likelihood ratio $\lambda(\mu)$ defined in Equation **??**, where $\hat{\boldsymbol{\theta}}_\mu$ is defined as the value of $\boldsymbol{\theta}$ that maximizes the value of the likelihood for a given μ, while $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are respectively the unconditional maximum-likelihood estimators of μ and $\boldsymbol{\theta}$.

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{8.5}$$

This equation is extremely important in the sense that it gives us a way to test the two hypotheses we have at our disposal against each other: the background-only hypothesis b (if $\mu = 0$) and the signal+background hypothesis $s + b$ (if $\mu = 1$). Additionally, this equation allows us to determine which values of signal strength μ can be discarded at a given confidence level, usually set to 95% within the **CMS!** collaboration. At this stage, it is important to note as well that if a significant excess is observed, additional tests are required to be able to claim for a discovery, as this method only allows to reject some signal models not significant enough.

In order to finally be able to establish upper limits on the value of the parameter μ, one final test-statistic q_μ is usually defined according to Equation **??**. Larger values for this parameter then represent higher incompatibilities between the data and the given signal strength. The reason why we set the estimator to 0 when $\hat{\mu} > \mu$ is simply to make sure to have a one-sided confidence interval, because eventual upwards fluctuations of the data able to pass this condition should not be considered as evidence for a μ-strong $s + b$ hypothesis.

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{8.6}$$

Additionally, p-values can be used in order to quantify the level of agreement between the data and the μ-strong signal under investigation, as shown in Equation **??** for an observed value $q_{\mu,\text{obs}}$, if $f(q_\mu|\mu)$ is the **PDF!** of the test-statistic q_μ assuming a certain value of μ.

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu) dq_\mu \tag{8.7}$$

The confidence level on the signal CL_s for a given signal strength is then defined in Equation **??**, where p_μ corresponds to the p-value obtained using the **PDF!** of the test-statistic q_μ assuming a value of μ, while p_b is the **PDF!** obtained for the same parameter but assuming $\mu = 0$. Most of these parameters introduced can be seen in Figure **??**.

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \tag{8.8}$$

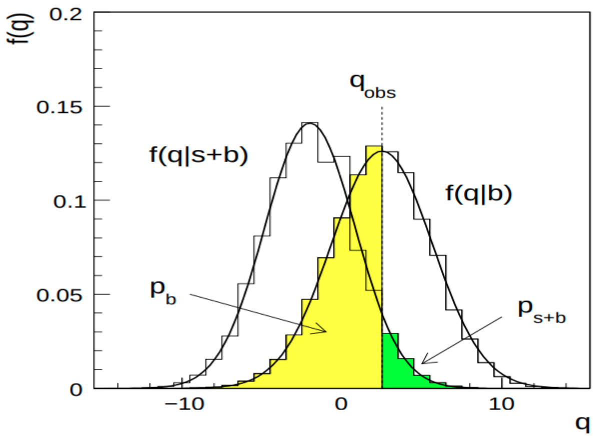


Figure 8.1: Schematic representation of the concept of **PDF!** and p-value used in the computation of our test-statistic q_μ, under the hypotheses $\mu = 0$ and $\mu = 1$.

This latest parameter CL_s can finally be tuned by modifying the value of μ until reaching a given threshold α, set to 0.05 in this thesis and in most of the publications of the **CMS!** collaboration.

Performing this complete calculation with observed data is giving us the so-called **observed upper limits**, but we also typically need to calculate the expected sensitivity of the analysis, defined as the mean value of μ expected for a background-only hypothesis, computed from N simulated datasets. Placing all the μ values obtained for a given threshold and a given dataset in a cumulative histogram then allows us to plot separate curves corresponding to both the observed and expected upper limits along with their corresponding error bars depending on the distribution of the μ values obtained for all the N datasets. This allows us to make sure that the sensitivity of the search is good enough to be able to effectively see a signal if it exists. Getting an expected value of $\mu < 1$ for a given mass point is then the sign that if the signal did in fact exist, the analysis should be sensitive enough to see it and, if we don't see any significant deviation between the expected and observed curves, we can then rule out the possibility of such model actually existing.

Finally, one last important concept can be discussed here. Two different kind of signals are searched for in this work at once, mainly the $t/\bar{t}$+DM and the $t\bar{t}$+DM processes. Even though we can search for both processes independently, they are both able to exclude dark matter production models and a combination of both searches is therefore expected to give better results than two separate searches. Some of the nuisance parameters are common between the channels, but not all of them and, if we take the same initial signal strength value for both channels and if the channels are statistically independent, which is a condition verified in this case, then the full likelihood function is simply given by the product of the individual likelihoods in each channel i, allowing us to define a common test-statistic parameter shown in Equation **??**.

$$\lambda(\mu) = \frac{\prod_i \mathcal{L}_i(\mu, \hat{\hat{\boldsymbol{\theta}}}_{\mu,i})}{\prod_i \mathcal{L}_i(\hat{\mu}, \hat{\boldsymbol{\theta}}_i)} = \prod_i \lambda_i(\mu) \tag{8.9}$$

Because of this, it is possible to determine the values of the profile likelihood ratio separately for each channel, which simplifies greatly the task of estimating the median significance that would result from the full combination. The same considerations can also apply when combining for example the different final states given by the different decays of the W bosons, or when combining the data collected during the different years of operation of the **LHC!** during the Run II.

8.2 Theoretical, statistical and systematic uncertainties

All the measurements we can make in high energy physics and in physics in general present some uncertainties, and the determination of their value is a critical point of every analysis, especially in this case since we try to detect a really low signal over a large background. Indeed, as we already saw, the nuisance parameters $\boldsymbol{\theta}$ are extremely important when defining the upper limits on the signal strength, so a good characterization of such source of uncertainties is crucial in most of the analyses searching for new physics such as this one. The uncertainties considered in this work belong to two major categories:

- On one hand, **statistical uncertainties** appear in any counting experiment [**?**] since every given measurement, made by definition of a finite set of observations, typically results in different observations when repeated, and the statistical uncertainty is then just a measure giving us an idea about the range of this kind of variations. Given the high rate of collisions happening at the **LHC!**, we usually assume that our counting experiment can be approximated with a Poisson distribution, for which we know that the error on the number of measurements is directly given by its square root.
- On the other hand, **systematic uncertainties** are just as important, but are usually harder to estimate and are different in nature [**?**] since they arise directly from the theory or from the detector itself. Some systematic uncertainties are treated as individual nuisance parameters when fitting the **MC!** observed to the data, nuisances which can either only affect the normalization of some processes, or affect the shape of the predictions across the distribution of the observables. These uncertainties share the need to be propagated through the full analysis chain all the way to the discriminating distributions.

Even though the signal enriched bins are to be found in a region of the **MVA!** discriminant with lower number of events, where the impact of statistical uncertainties is usually higher, systematic uncertainties still play an important role in most of the analyses performed at **CERN!**, including this one.

Statistical uncertainties

Shape uncertainties due to the limited size of the simulated signal and background samples are simply included by allowing each bin of the distributions included in the signal extraction to fluctuate independently according to the statistical uncertainty on the simulation.

Experimental uncertainties

Additional uncertainties related to the detector or to the experimental conditions are also taken into account when producing the upper limits on the signal strength.

- **Luminosity**. The actual integrated luminosity collected is not a fixed number and comes with an associated uncertainty of 2.5%, 2.3% and 2.5% for 2016, 2017 and 2018 respectively, based on van der Meer scans performed [**?**, **?**, **?**].
- **Pileup modeling**. Varying the inelastic cross-section used to calculate the pileup distribution in simulation results in a systematic of the order of 5% [**?**]. This systematic uncertainty is considered to be correlated across the years.

- **Lepton triggers**. An uncertainty associated to the lepton trigger efficiency is taken into account. It has an impact of the order of $\sim 2\%$, estimated from a general Tag and Probe method, by varying the Z window considered and the tag lepton p_T selection cut.
- **Lepton efficiency and energy scale**. Scale factors are applied to the **MC!** processes in order to mimic the measured lepton reconstruction and selection efficiencies in data, and such scale factors come with estimated p_T and η uncertainties of the order of $\sim 2\%$ for electrons and muons, which have been considered as well.
- **JES! (JES!)**. Single nuisance parameter applied to the p_T of each jet, as a function of the jet p_T and η values. These variations are then coherently propagated to important discriminating variables, such as the E_T^{miss} and M_{T2}^{ll}, according to a procedure described by the JET/MET **POG!** [**?**]. These uncertainties are expected to be below 3% across the phase space considered by most of the analyses ($p_T > 30$ GeV, $|\eta| < 5.0$) and even below 1% in the barrel region ($|\eta| < 1.3$) [**?**].
- **MET! mismodelling**. The uncertainties on jets, electrons and muons that make up the **MET!** are applied to the respective objects, while the **MET!** is recalculated with the same up and down variations alongside. Additionally, any eventual unclustered energy, mostly due to **ECAL!** or **HCAL!** deposits not assigned to any object is shifted up and down, and the impact of such variations has been estimated on the **MET!**.
- **b-tagging efficiency**. The b-tagging process is not perfect and its efficiency, typically different in data and **MC!** also needs to be taken into account by applying a p_T, η and flavour dependent scale factor on the **MC!**. The uncertainties on such scale factors are measured in an independent control sample and propagated to the analysis [**?**].
- **Top quark p_T reweighting**. As described in Section **??**, the measured p_T spectrum is expected to be softer than the one obtained using simulation, so a factor correcting this effect is usually applied to the **MC!**. A systematic assumed correlated through the years and covering for such differences observed between data and **MC!** is taken into account.
- **ECAL! prefiring**. An uncertainty accounting for the prefiring of the **ECAL!** in 2016 and 2017 is also taken into account, even though its impact is expected to be quite small overall.

Background related uncertainties

The systematic associated to the backgrounds normalization is typically one of the largest source of uncertainties in an analysis, and is computed by estimating the normalization of the backgrounds that are estimated on data control samples whenever possible.

- **Statistical uncertainties** are first assigned to all the **MC!** samples as well.
- **DY! (DY!) background**. To cover for the non-flatness of the $R_{\text{in-out}}$ transfer factor and possible mismodeling of this particular background, especially at high pf**MET!** values, a 20% systematic is associated to this process.
- **Minor backgrounds**. An uncertainty of 30% (for all the minor backgrounds, except for the ttV, for which a 50% uncertainty was actually used, obtained by looking at the data/**MC!** discrepancies in some regions enriched in this particular background) has been assigned to the normalization of all the minor backgrounds of the analysis.

Theoretical uncertainties

Finally, this category of systematic uncertainties is related directly to the theoretical models used in the analysis and is mostly related to the production of the **MC!** simulations according to the process described in Section **??**. Several different uncertainties belong to this particular category:

- **PDF and higher order corrections**. As accurate as Feynman diagrams can be, they do not represent the complete picture since we are computationally limited and therefore cannot consider high order perturbations, **MC!** samples being usually limited to the **LO!** (**LO!**) or **NLO!** (**NLO!**). This limitation introduce a small systematic uncertainty that usually needs to be taken into account. Additionally, uncertainties due to the choice of the **PDF!**s and α_s parameters themselves are usually estimated by considering a hundred NNPDF3.0 [**?**] replicas, according to the PDF4LHC recommendations for the **LHC!** Run II [**?**]. The uncertainty obtained in a given bin is then set as the standard deviation of the content of the bin obtained from all these computed replicas, and is typically of the order of 4% of the cross-section value.
- **Renormalization and factorization scales**. Uncertainties usually emerge due to the presence of the renormalization μ_R and factorization μ_F scales in the QCD calculations regarding the hard collisions of hadrons happening at the **LHC!**, which are updated by a factor 0.5 and 2 and then propagated to the distributions of the analysis. This uncertainty is considered to be uncorrelated among the different background processes considered, and takes values typically lower than 4% of the cross-section value as well.
- **Underlying event and parton shower modeling**. The choice of the parton shower generator and **UE!** tune is also to have an effect on the final results, but since we did not observe any dependency of **UE!** variations on the number of jets, a flat 1.5% **UE!** uncertainty is assigned to cover all the up and down variations.

8.3 Results

First of all, the impact plots, showing the impact that each systematic has on the final upper limits obtained, are shown in Appendix **??**. In such plots, we can observe that the normalization of the ttV process is the dominant systematic across all the years and for all the mediators considered. We can also observe that both the **MET!** and **JES!** systematics are strongly constrained by the fit performed by the combine tool. According to our investigations, such constraints come from the low region of the output shape of the **BDT!**, enriched in backgrounds (mainly $t\bar{t}$) and which plays the role of a control region, therefore able to constrain a lot these two systematics (as a cross-check, we tried removing the negative values from the shape of the **BDT!** and such constraints simply disappeared).

The post-fit plots showing the shape of the **BDT!** after the fit performed by combine can also be found in Appendix **??**. Such plots look quite healthy, as the post-fit shape looks globally better than the pre-fit, which indicates that the fit was performed well enough.

Expected and observed upper limits on the different signals production cross-sections at the 95% confidence level have been obtained and plotted against all the possible dark matter mediator masses considered in this actual analysis, considering both the scalar and pseudoscalar mediators, as shown in Figure **??**, for $m_\chi = 1$ GeV and when setting all the couplings of the different models to 1. In both cases, limits obtained separately for the $t/\bar{t}$+DM and $t\bar{t}$+DM have been obtained, along with a combination of the limits obtained considering these two signals of interest at once.

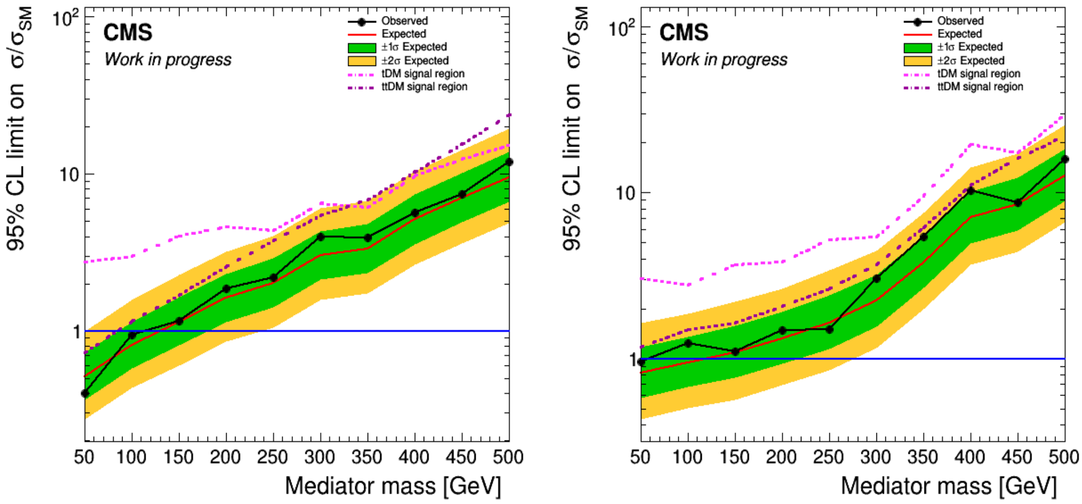
Finally, Figure **??** shows the $t/\bar{t}$+DM and $t\bar{t}$+DM combined expected and observed limits obtained for each data taking period in the same plot, along with a second combination of all the different exclusion limits obtained for each year, giving us the final exclusion limits for this analysis, for both the scalar and pseudoscalar mediators. These limits were obtained individually, considering both signal samples together in each signal region (pink and purple lines), and then combined together.

By looking at all these plots, we can draw several different observations and conclusions:

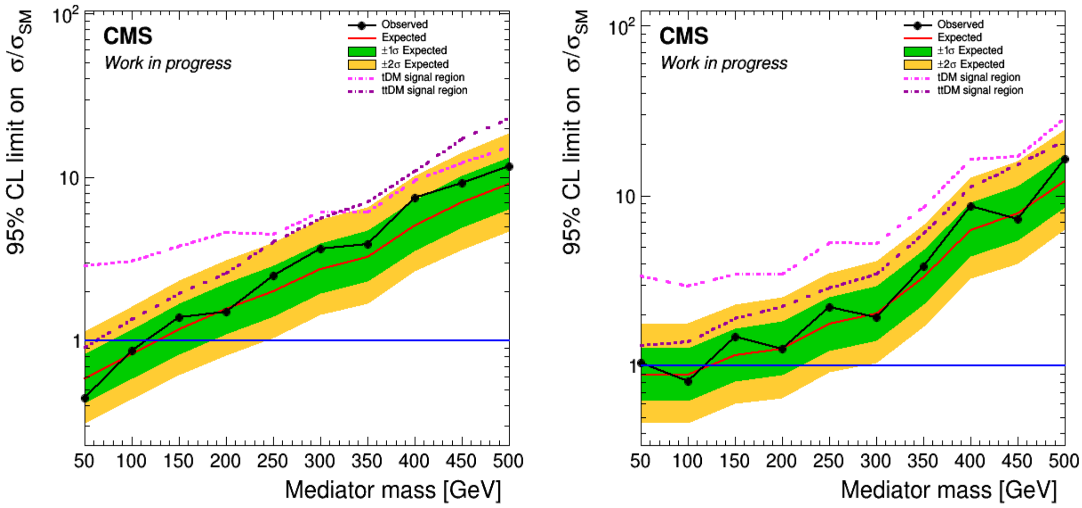
- First, these plots allow us to see that the **inclusion of the $t/\bar{t}$+DM signal is extremely important**, especially at high scalar mediator masses, where the limits obtained are comparable or even better than the ones obtained by considering the $t\bar{t}$+DM signal only.
- We can also observe that the limits obtained in 2018 are slightly better than the other two data taking periods. This is an expected effect as expected limits are roughly expected to decrease with the square root of the luminosity, slightly higher in 2018. The difference is not noticeable for all the mass points though, as the log scale can hide this small difference.
- Finally, we can observe that low masses mediators are easier to exclude than higher masses simply because, even though they feature a worse global signal/background discrimination, their cross-section is also much higher. In this sense, pseudoscalar exclusion limits are also typically slightly worse than the scalar ones, mainly because of the difference in cross-section between such processes.

Another way to look at this is to consider each signal individually in both signal regions, and then combine these results. The plots obtained in this case can be found in Figure **??**.

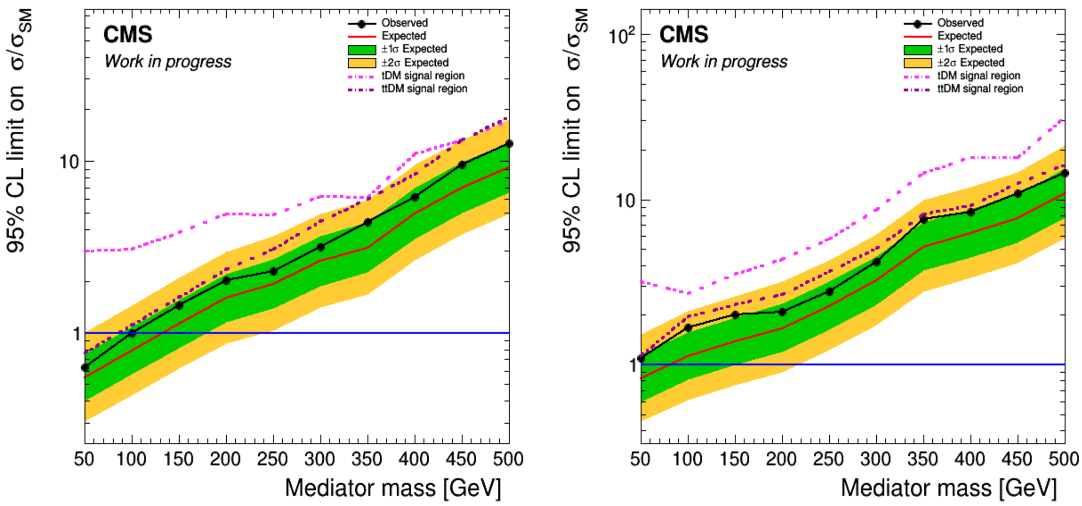
At the end of the day, this analysis allowed us to achieve an expected (observed) exclusion for both scalar and pseudoscalar mediators up to 215 (180) and 250 (220) GeV respectively. Even though similar results than the ones obtained in 2016 (as shown in Figure **??**) are found for the 2016 $t\bar{t}$+DM model alone, the inclusion of more statistics and of the new $t/\bar{t}$+DM production model allows us to obtain much better results this time around.



(a) 2016 exclusion limits

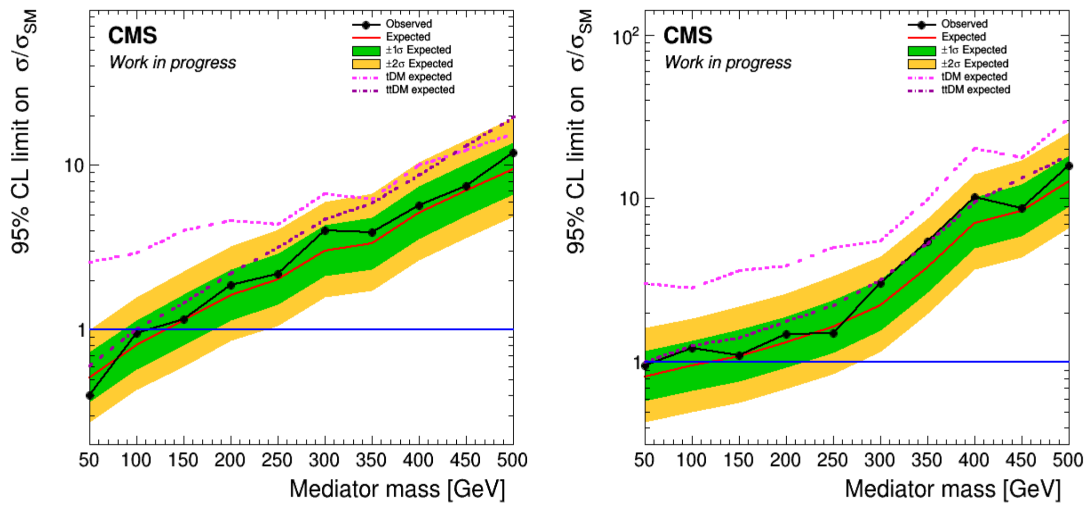


(b) 2017 exclusion limits

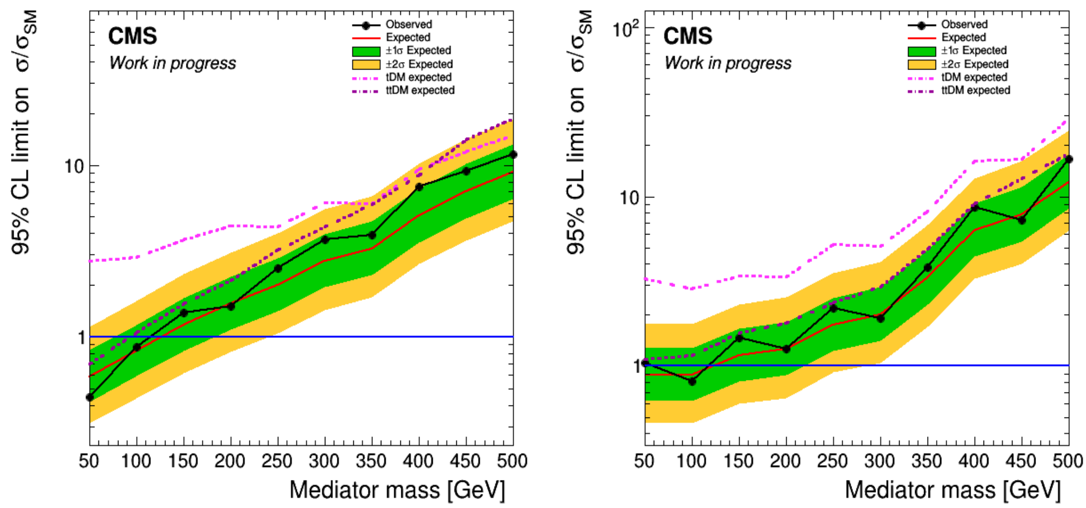


(c) 2018 exclusion limits

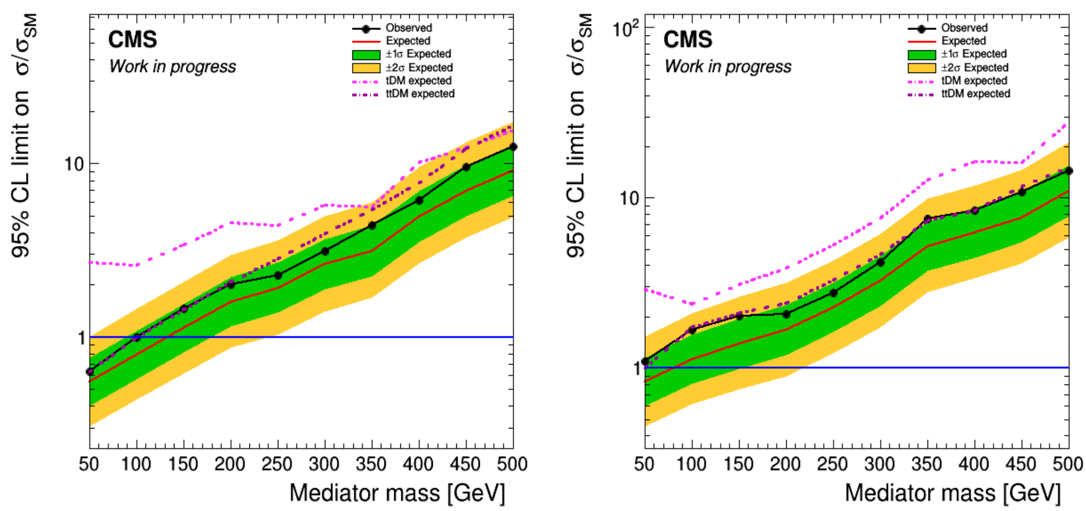
Figure 8.2: Expected and observed upper limits on the signal strength μ for scalar (on the left) and pseudoscalar (on the right) models, considering with $m_\chi = 1$ GeV and $g_\chi = g_q = 1$, obtained by considering both signals in each of the signal regions, and then combining these regions together.



(a) 2016 exclusion limits



(b) 2017 exclusion limits



(c) 2018 exclusion limits

Figure 8.3: Expected and observed upper limits on the signal strength μ for scalar (on the left) and pseudoscalar (on the right) models, considering with $m_\chi = 1$ GeV and $g_\chi = g_q = 1$, obtained by considering individually each signal in both signal regions and then combining the signals together.

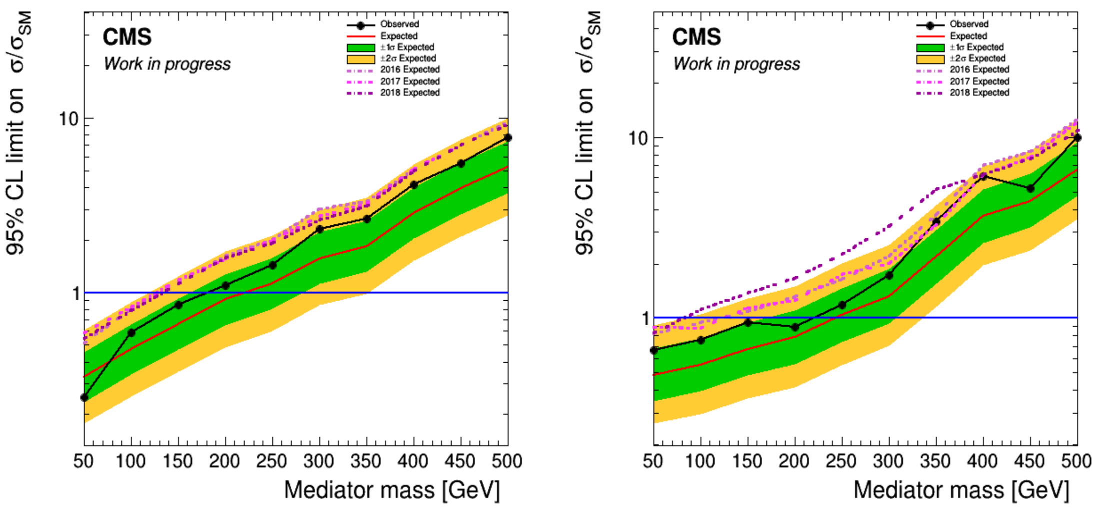


Figure 8.4: Expected and observed upper limits on the signal strength μ for scalar (on the left) and pseudoscalar (on the right) models, considering with $m_\chi = 1$ GeV and $g_\chi = g_q = 1$, after combining the different Run II data taking periods.

CHAPTER 9

Conclusions

In conclusion, a search for the production of dark matter in association with one or two top quarks has been performed in this work, by studying in particular the dilepton decay channel of both production modes. This analysis was done by considering the full (137.1 ± 2.0) fb^{-1} of proton-proton collisions data collected by the **CMS!** detector during the Run II of operation of the **LHC!**, at a center of mass energy of $\sqrt{s} = 13$ TeV. No evidence for the existence of dark matter has been found, but upper limits on the signal strength have been obtained by considering different production models and channels. Several steps where needed to reach this goal:

- First, the triggers were chosen in such a way to collect as much interesting data as possible. Different objects were also defined and chosen at this point in coordination with other groups, such as the working point used for leptons and for the b-tag. The choice regarding the **MET!** used in this particular analysis was a bit trickier, given the impact this choice might have on the final results, but we finally settled on the usual pf typeI corrected **MET!**, applying all the corrections recommended to this particular variable;
- Once this was done, a full top reconstruction method was developed in order to reconstruct in the best was possible the kinematic of the $t\bar{t} \rightarrow 2l$ process. This method allowed us for example to get the information related to each top quark and neutrino separately, and to define new discriminating variables, such as as estimation of the mediator p_T;
- All the signals samples were then produced, first privately and then centrally for the $t\bar{t}$+DM model, considering all the different models that might be of interest for this analysis;
- The background were then carefully studied. Even though most of the backgrounds are estimated directly from **MC!**, some of them did require a bit extra work, such as the **DY!** process, for which a semi data-driven method was used for its estimation. The predictions

made for most of the major backgrounds of the analysis by such simulations were then checked in dedicated control regions;

- Several discriminating variables available to us were then explored, and a MVA method was set up in order to combine the discriminating power of all these variables into a single output, used to perform a general shape analysis. The optimization of all the hyperparameters was thoroughly done, as we ended up choosing an **BDT!** to perform the actual analysis, with a simple set of 6 different input variables;
- Finally, exclusion limits on the signal strength at the 95% confidence level were set for the different models considered.

At the end of the day, this analysis allowed us to achieve an expected (observed) exclusion for both scalar and pseudoscalar mediators up to 215 (180) and 250 (220) GeV respectively, improving by a factor of more than 3 the previous results obtained in 2016 for the $t\bar{t}$+DM model alone.

9.1 Future prospects

This is the first time that such a search combining the $t/\bar{t}$ and $t\bar{t}$+DM models is performed in the dilepton final state. Even though large parts of the phase space have already been excluded, several additions could be considered to improve and complete the results obtained by this analysis:

- First of all, the combination of this work with the semileptonic and hadronic final states is expected to increase the exclusion limits by at least a factor of 2, given the large branching ratio and therefore sensitivity of such processes.
- The remaining models to be excluded correspond to high mediator masses, which feature a large signal/background discrimination on their own but which also have a low cross-section of production. The continuous operation of the **LHC!** is therefore expected to give more data to analyze, in turn decreasing the limits on the signal strength roughly as $\left(\frac{1}{\mathcal{L}}\right)^{1/2}$. This channel will therefore benefit from the larger luminosity collected during the full Run III of the LHC data taking period, expected to start soon in 2022;
- Given that the normalization of the ttV process is consistently the dominant systematic uncertainty in all the signal regions, it will be interesting in the future to define a dedicated control region for this background in order to constrain it a bit more and reduce its overall impact on the final results;
- A single set of couplings was studied in this work, but exploring other accessible regions of the space (by considering different m_χ values, even tough they are expected to have a minimal impact on the kinematics of the event, and mostly by changing the value of the couplings considered) would in this sense be an interesting addition to this particular analysis;
- One possible addition to this work would be to include a ttH to invisible reinterpretation of this search, given the relative similarities between this kind of model and the dark matter models considered here.

This analysis is in any case expected to gain momentum and provide us with even better exclusion limits over the course of the next years of operation of the **LHC!**.

Appendices

APPENDIX A

Resumen en español

El Modelo Estándar (ME) de la física de partículas [?] es hoy en día el modelo matématico que mejor describe hasta el momento las partículas fundamentales y tres de las cuatro interacciones fundamentales conocidas que existen entre ellas. Este modelo ha permitido explicar muchos de los fenómenos observados en la naturaleza con un grado de precisión excelente, y ha sido capaz a lo largo de los años de hacer predicciones sobre la existencia de nuevas partículas, como el postulado del mecanismo de Brout-Englert-Higgs [?, ?], seguido por el descubrimiento del bosón de Higgs [?, ?] por los experimentos **CMS!** [?] y **ATLAS!** [?] del **CERN!**, analizando colisiones entre protones producidas por el **LHC!** (**LHC!**) a las energías de colisión en el centro de masas de $\sqrt{s}$ = 7 TeV y 8 TeV.

A pesar de estas predicciones, sabemos hoy en día que el Modelo Estándar no es completo en el sentido de que no es capaz de explicar algunas observaciones cosmólogicas hechas a lo largo del siglo XX [?]. Eventuales partículas exóticas podrían por ejemplo ser el signo de la existencia de física más allá del Modelo Estándar. Este trabajo consiste por lo tanto en tratar de mejorar nuestro entendimiento del universo en el cual vivimos buscando partículas de materia oscura que podrían ser producidas con las colisiones entre protones del **LHC!** del **CERN!**.

A.1 Materia oscura

La primera hipótesis seria sobre la posible existencia de la materia oscura se desarolló en los años 70, después de que varios astrofísicos observaran anomalías gravitacionales, siendo una manera sencilla de explicar la aparente falta de materia luminosa en el universo [?]. En efecto, la masa visible dentro de las galaxias parecía ser muy baja para explicar algunos fenómenos observados, como las curvas de rotación de las galaxias [?], que parecen ser incompatibles con las leyes de la gravitación de Newton. Algunas medidas adicionales del efecto lente (en el Bullet Cluster, por ejemplo [?]) y las anisotropías observadas en el fondo cósmico de microondas [?] son otras evidencias de la posible existencia de materia oscura.

Estas observaciones cosmologícas han permitido determinar que la materia bariónica ordinaria solo constituye el 5% del universo visible, mientras la materia oscura cuenta para el 26% de la densidad de energía total del universo (la energía oscura, otro fenómeno completamente distinto constituye la parta faltante de esta densidad de energía). Estudiar la naturaleza y las propiedades de este nuevo tipo de partículas es por lo tanto muy importante para entender las leyes físicas que rigen nuestro universo, con muchos científicos y experimentos alrededor del mundo dedicados a este tipo de búsquedas.

La existencia de la materia oscura está muy bien motivada, pero no hemos sido capaces de observar este tipo de partículas hasta la fecha y la única evidencia sobre su posible existencia viene de sus efectos gravitaciones observados a larga distancia. Se desconocen la masa, espín y propiedades básicas de este tipo de materia, pero muchas de las teorías contemplan un tipo particular de candidatos para constituir la materia oscura: las **WIMP!**s (**WIMP!**s), ya que este tipo de partículas cumplen con propiedades generalmente asociadas a esta idea de materia oscura. En efecto, sabemos que el candidato ideal para formar esta materia exótica debe de ser entre otros negro (es decir, insensible a la radiación electromagnética), no-bariónico, frío, estable a largo plazo (por lo menos un tiempo superior al tiempo de vida del universo) y en un rango de masa entre 10 GeV y 1 TeV.

Diferentes métodos se pueden usar para buscar a este nuevo tipo de partículas, perteneciendo a tres categorías principales:

- Las **búsquedas directas**, buscando posibles interacciones entre materia oscura y partículas del ME, estudiando por ejemplo choques y por lo tanto intercambio de energía entre estos dos sectores. Estas búsquedas se suelen llevar en laboratorios muy aislados, y a menudo debajo de la tierra (como el experimento Super-Kamiokande en Japón), para reducir tanto como sea posible la contaminación debida a fenómenos radiactivos naturales, por ejemplo;
- Las **búsquedas indirectas**, tratando de encontrar nuevas partículas del Modelo Estándar (típicamente rayos γ, neutrinos o rayos cósmicos que podrían emerger de la interacción entre dos partículas de materia oscura;
- Y la **producción dentro de colisionadores de partículas**, el método de investigación principal de este trabajo, ya que se supone que chocar dos partículas del Modelo Estándar con una determinada energía podría llegar a producir partículas de materia oscura. Este tipo de búsquedas resulta ser muy interesante ya que permite estudiar candidatos a materia oscura de muy baja masa.

Chocar partículas fundamentales se puede hacer hoy en día en aceleradores de partículas como el **LHC!**, pero estudiar la posible producción de materia oscura resulta ser muy complicado por la naturaleza misma de este tipo de materia, ya que no interacciona con el detector dispuesto alrededor del punto de colisión. Esto significa que, incluso si logramos producirlas, se espera que estas nuevas partículas escapen del detector sin ser detectadas y este tipo de búsquedas tiene por lo tanto que buscar materia oscura como energía faltante producida en asociación con partículas del Modelo Estándar que se pueden detectar.

En este trabajo en particular, se busca materia oscura producida en asociación con uno o dos quarks top con el detector **CMS!** a una energía en el centro de masa $\sqrt{s} = 13$ TeV, considerando los ~ 137 fb^{-1} de datos recogidos durante el Run II de operación del **LHC!**, a lo largo de los años 2016, 2017 y 2018. Este canal de investigación resulta ser muy interesante ya que se espera que el acoplamiento entre los sectores bariónico y oscuro sea de tipo Yukawa, y que sea por lo tanto mayor para partículas que tengan una masa mayor, como es el caso del quark top. Estos quarks tienen sin embargo un inconveniente en el sentido de que se desintegran rápidamente, antes de llegar al detector, y solo podemos observar el resultado de su desintegración, siendo en general formado por unos leptones y unos jets, resultado del proceso de hadronización de los quarks producidos. El número de leptones que esperamos observar depende de la desintegración del bosón W que aparece en la cadena de desintegración del quark top. En este trabajo, nos concentramos en la desintegración dileptónica de este bosón, que presenta la ventaja de ser un canal de desintegración bastante limpio, con pocos fondos pero con el inconveniente de tener una fracción de desintegración bastante pequeña comparada con los canales semi-leptonics y hadronicos.

A.2 El dispositivo experimental

El LHC

El **LHC!** es un acelerador y colisionador de partículas ubicado a 100 metros bajo tierra en la frontera entre Francia y Suiza; es el resultado de la colaboración entre cientos de institutos, científicos, ingenieros e informáticos, entre otras profesiones, del mundo. Capaz de acelerar protones o iones de plomo a una velocidad muy cercana a la velocidad de la luz, este dispositivo, de una circunferencia de 27 kilómetros, empezó a funcionar en el año 2010 a una energía en el centro de masa de 7 TeV. A lo largo de los años, esta energía subió hasta llegar a los 13 TeV al empezar el Run II de funcionamiento en el 2016, después de una parada técnica de unos años. El **LHC!** es una máquina ideal para estudiar física más allá del Modelo Estándar por los niveles de energías alcanzables y por su alto número de colisiones producidos por segundo (su luminosidad instantánea), lo que permite estudiar procesos con baja tasa de producción.

El detector CMS

En sitios diversos del **LHC!**, 4 detectores han sido construidos alrededor del *beam pipe* para estudiar las partículas producidas en las colisiones entre los protones o iones pesados. Estos detectores, **CMS!**, **ATLAS!**, **ALICE!** y LHCb, han sido desarrollados usando equipos y métodos distintos, para lograr varios objetivos.

En particular, en este trabajo se estudian los datos tomados por el detector **CMS!**, un detector de forma cilíndrica, de propósito general y diseñado para ser lo más hermético y compacto posible. Este detector pesa unos 12 500 toneladas, mide unos 14 metros de diámetro y 22 metros de altura, y está compuesto de diferentes capas, cada una diseñada con un objetivo distinto:

- Primero, en el interior del detector y muy cerca del punto de colisión, un sistema perfeccionado de detección de trazas cargadas conocido como el *tracker* ha sido instalado, para identificar el punto de origen de la colisión y ayudar con la identificación y la medida del momento de las partículas creadas. Este detector, compuesto por unas capas de píxeles de silicio rodeada por un detector de tiras de silicio de 10 capas y por sus tapas correspondientes para que esta capa sea lo más hermética posible, tiene que ser muy resistente a la radiación y rápido para medir todas las colisiones, que se producen cada 25 nanosegundos;
- Luego vienen los dos calorímetros: el **ECAL!** y el **HCAL!**, que permiten medir la energía de las partículas interaccionando de manera electromagnética o hadrónica respectivamente. El **ECAL!** está hecho de una parte central compuesta por 61 200 cristales de tungsteno de

plomo $PbWO_4$ y completada por dos barriles de 7 324 cristales adicionales. El **HCAL!** por el otro lado está hecho por unas capas de centelladores y de material que permite desarrollar cascadas hadrónicas, lo que permite medir con precisión la energía de los eventuales hadrones producidos;

- La parte central del detector es un imán capaz de crear un campo magnético de 3.8 T, lo que permite determinar el momento de las partículas gracias a la relación de Lorentz;
- Fuera de esta parta central se encuentra el sistema de detección de muones, ya que este tipo de partículas interacciona poco con la materia y puede escapar del detector. Este último sistema de detección se basa en varias tecnologías: su zona central, donde el campo magnético es uniforme, viene por ejemplo formada por las cámaras de derivas (las **DT!**s (**DT!**s)). Estas cámaras están divididas en 4 capas y están formadas por más de 172 000 cables que permiten medir la posición exacta de los muones gracias al fenómeno de ionización. Las tapas de este sistema, en la zona donde el campo magnético no es uniforme y donde la tasa de muones y de fondos es más alta, las cámaras de tiras catódicas **CSC!**s (**CSC!**s) están instaladas y, por fin, el sistema está completado por las cámaras de tiras resistivas **RPC!**s (**RPC!**s), mucho más rápidas que las **CSC!**s.

Debido a limitaciones computacionales, es imposible guardar todas las colisiones que se producen en el **LHC!** y, de todas maneras, la gran mayoría de las colisiones son muy bien conocidas. Por lo tanto, un sistema de *trigger* de dos niveles ha sido desarrollado para guardar solamente unos 1000 eventos por segundo que parecen ser muy energéticos y interesantes desde el punto de vista de la física para un análisis ulterior.

A.3 Reconstrucción de objetos

Tal y como lo acabamos de ver, el detector **CMS!** está hecho por varias capas dedicadas a las medidas de diferentes propiedades de las partículas que se generan después de cada colisión entre dos protones en el *beam pipe*. Cada capa del detector atravesada por una de estas partículas genera señales eléctricas que se guardan. Un algoritmo conocido como el *Particle Flow* ha sido desarrollado para analizar las señales que provienen de cada parte del detector, para combinarlas y extraer información física sobre la colisión estudiada.

Para la búsqueda de materia oscura llevada a cabo en este trabajo, se seleccionan eventos que tienen dos leptones y dos jets y, por lo tanto, un sistema eficiente de detección, reconstrucción e identificación de este tipo de objetos es imprescindible. Los muones se reconstruyen primero con el **PF!**, ya que son las únicas partículas cargadas capaces de llegar a las cámaras de muones en el exterior del detector. Luego, se pueden reconstruir los electrones, identificados como una cáscada electromagnética en el **ECAL!** asociada con trazas compatibles en el tracker y, por fin, se identifican los hadrones cargados y neutros asociando los hits que quedan en el tracker con la información del **HCAL!**. Los quarks se hadronizan y se manifiestan como jets de partículas en el detector, más difíciles de reconstruir en general. Se desarolló para llevar a cabo esta tarea el algoritmo *anti-kT*, capaz de agrupar los hadrones, fotones y leptones que forman un solo jet. Este algoritmo es también capaz de distinguir un jet procedente de un quark bottom, ya que los hadrones formados a partir de estos quarks tienen un tiempo de vida un poco más alto y que pueden por lo tanto viajar unos poco milímetros antes de decaer. Este pequeño desplazamiento basta para distinguir estos jets, llamados b-jets, que tienen un papel importante en este análisis.

El *Particle Flow* es también capaz de reconstruir un objeto llamado **MET!** y que representa la energía transversal faltante de una colisión, siendo definido como la suma del momento transverso de todas las partículas creadas y medidas por el detector, menos el módulo de la suma vectorial. Como sabemos que estas colisiones se producen de manera frontal, el momento transverso total de una colisión es exactamente 0 al principio y se tiene que quedar así después del choque por el fenómeno de conservación del momento. Sin embargo, no siempre es el caso y a veces se puede observar un momento transverso total no nulo más que nada por imperfecciones en el detector, o debido a la producción de partículas capaces de escapar del detector sin dejar ninguna señal, tal y como los neutrinos del Modelo Estándar o eventuales partículas exóticas. Una reconstrucción completa y precisa de esta variable es por lo tanto muy importante en este análisis.

Finalmente, hemos desarrollado para esta análisis un método completo de reconstrucción *offline* de los sucesos en los que se producen pares de quarks top. Este algoritmo es muy importante ya que se espera que muchas de las variables que nos permiten distinguir entre la señal buscada y los fondos dependen del 4-momento de los quarks top y de los neutrinos, típicamente no disponible sin este proceso de reconstrucción.

A.4 Análisis de datos

El objetivo principal de este tipo de búsqueda de física nueva consiste principalmente en observar los datos recogidos por el detector **CMS!** en este caso, y compararlos con simulaciones de tipo Monte-Carlo que permiten simular la naturaleza aleatoria de los procesos cuánticos involucrados en la física de partículas y la interacción de las partículas producidas con el detector. Calquiera desviación que se podría ver entre los datos y simulaciones podrían ser el signo de la existencia de física nueva y, por lo tanto, el proceso de generación de muestras de Monte-Carlo tiene que ser muy preciso y suele ser costoso computacionalmente.

Para este análisis, se estudian los (137.1 ± 2.0) fb^{-1} de datos acumulados por el detector **CMS!** a lo largo de los años 2016, 2017 y 2018, con el objetivo de intentar descubrir algunos eventos en los cuáles se observa materia oscura producida en asociación con uno o dos quarks top. Dado lo poco que se conoce sobre la materia oscura hoy en día, diferentes modelos se consideran para este tipo de búsquedas, considerando diferentes canales de producción, masas (de 10 GeV a 1 TeV) y espines (0 o 1) de mediadores.

Dos modelos principales de producción de materia oscura se consideran en este análisis: con un solo quark top, llevando al análisis llamado $t/\bar{t}$+DM, o bien con dos quarks top, denominado $t\bar{t}$+DM. En ambos casos, en este trabajo se estudia solamente el estado final dileptónico, aunque se puedan observar en general 0, 1 o 2 leptones en el estado final, que vienen acompañados por un número de b-jets y una cantidad de energía faltante que puede variar. La concepción de estos modelos siguió las recomendaciones del ATLAS-CMS Dark Matter Forum [**?**].

Definición de objetos

Los objetos que se quieren usar en al análisis mismo se tienen que definir de antemano:

- Los *triggers*, que permiten reducir el flujo de datos que se guardan, seleccionando eventos que tengan leptones aislados (o no) en este caso, son de esta manera los primeros objetos que hemos definido para los periodos de toma de datos 2016, 2017 y 2018. Se tomaron precauciones a la hora de seleccionar estos *triggers*, más que nada para evitar calquier problema de *double counting* que podría producirse si un evento tiene la posibilidad de entrar en dos *triggers* distintos a la vez;

- Luego, otra parte fundamental del análisis consiste en definir los leptones que se quieren usar. De esta manera, definimos electrones basandónos en la definición usual de un electrón *medium* dada por el Grupo de física de electrones (EGamma **POG!** (**POG!**)) [**?**], con unos cortes adicionales en los parámetros de impacto d_{xy} y d_z para reducir la contaminación debida a los fakes, objetos identificados de manera incorrecta como leptones por el detector, en nuestras zonas de señal;
- Seguimos una estrategía similar a la hora de defnir nuestros muones, basandónos esta vez en el *medium working point* del muon **POG!** [**?**], con unos cortes addicionales en la isolación de los leptones (corte *tight*) y en los parámetros de impacto también;
- Luego se requiere que los jets tengan un momento transverso $p_T > 30$ GeV y una pseudorrapidez $|\eta| < 2.4$ (limite de la extensión del tracker a partir del cuál no se espera poder medir los jets con un alto grado de precisión) para formar parte del análisis. Además, el tight *working point* dado por el JET/MET **POG!** [**?**] se usa, porque ofrece una eficiencia muy alta de rechazo del ruido. Los jets de *pile-up*, definido como el cunjunto de colisiones que ocurren (y se recogen) de forma simultánea a la colisión que dispara el trigger, también se quitan aplicando además la definición de *tight* Jet **PU!** ID para jets de alto p_T (> 50 GeV);
- Por fin, dada la presencia de quark tops que tienen un tiempo de vida pequeño y que se desintegran rápidamente a quarks bottom, los b-jets son muy importantes y se definen a partir del *medium working point* del **POG!** de b-tagging y vertexing [**?**], para el cuál la tasa de identificación errónea de un jet ligero como un b-jet es del 1%.

Estimación de fondos

Por otro lado, la estimación de los distintos fondos del análisis también es muy importante, ya que sirve de referencia. La mayoría de los fondos están bastante bien entendidos y se pueden por lo tanto estimar usando simulaciones de Monte-Carlo directamente, mientras que zonas de control se suelen definir para comprobar el acuerdo datos/**MC!** de los fondos más importantes del análisis, tal y como el Drell-Yan (más que nada por su alta sección eficaz de producción, aunque reducir este fondo suele ser bastante fácil ya que su cinemática resulta ser muy distinta a la de nuestras señales de interés), el $t\bar{t}$ del Modelo Estándar y la producción de un simple top.

La mayoría de los fondos se estiman directamente a partir de simulaciones matématicas de tipo **MC!**, y se pueden controlar en zonas de control enriquecidas en fondos en particular. Aún así, la estimación del impacto que tiene el proceso conocido como el Drell-Yan tiene un tratamiento un poco peculiar. En efecto, en este se usan los datos mismos para calcular un factor de corrección que permite estimar el número de eventos de este proceso fuera del pico del Z a partir de los eventos observados dentro de este pico.

Extracción de señal

Les señales de interés en este análisis están muy cerca de algunos fondos del Modelo Estándar desde el punto de vista de la cinématica: se espera por ejemplo que sea complicado separar la señal de $t\bar{t}$+DM del proceso $t\bar{t}$ del Modelo Estándar. Se espera aún así que algunas variables presenten algo de discriminación entre estos tipos de procesos muy similares: por ejemplo, la **MET!** debería ser estadísticamente más alta cuando consideramos la señal ya que, aunque el proceso $t\bar{t}$ y la señal suelen presentar neutrinos con forma de energía transversa faltante en el detector, una contribución adicional a la **MET!** se espera en el caso de la señal, por la aparición de un par de partículas de materia oscura.

En este trabajo en concreto, un método multivariable avanzado ha sido desarollado con el objetivo de poder combinar el poder de discriminación de unas 10 variables que deberían presentar algo de discriminación entre las señales y los fondos del análisis, para poder definir zonas de señal enriquecidas en las señales de interés. Estas variables se han seleccionado tras realizar un estudio, en función de las cinemáticas de señal y fondos a partir de simulaciones de **MC!**. Con este objetivo de dicriminación en mente, se desarollaron en paralelo dos métodos de *machine learning* distintos: una **BDT!** (**BDT!**), y una **ANN!** (**ANN!**), ambas entrenadas con un conjunto etiquetado de simulaciones de Monte-Carlo para que sean capaces de asignar una probabilidad de pertenencia a cada una de las 2 categorías de interés: las dos señales ($t\bar{t}$+DM y $t/\bar{t}$+DM) o fondo. Un largo proceso de optimización de las redes tuvo por lo tanto lugar, con el objetivo de definir redes tan eficiente y pura como sea posible y, una vez el entrenamiento hecho y el proceso de validación terminado, pasamos los datos acumulados a traves de ambas redes para asignar una etiqueta a cada evento. De esta manera, definimos una única variable de salida que nos permite discriminar estos procesos y con la cuál hemos podido definir nuestras zonas de señal del análisis. Esta variable de salida se usa además para hacer un *shape analysis*, que consiste en aplicar métodos estadísticos avanzados para estudiar la forma de esta variable para los diferentes procesos del análisis, con el objetivo de tratar de descubrir una posible señal de materia oscura.

A.5 Resultados obtenidos

El resultado más importante de este trabajo consiste en obtener los limites de exclusión de la sección eficaz de producción de las señales de intéres, para las dos señales y para los 3 años de toma de datos por separado, y luego combinando todos estos parámetros. Esto se hace considerando diferentes fuentes de errores: estadísticos y sistemáticos. Los errores sistemáticos a su vez se pueden dividir en teóricos (tal y como el sistemático relacionado con la estimación de las **PDF!** de los diferentes procesos) y experimentales (tal y como el error asociado a la medida de la luminosidad colectada por el detector **CMS!**, y el error asociado al uso de *triggers* para recoger los datos).

En este sentido, hemos calculado los limites de exclusión observados y esperados de la sección eficaz de producción de las señales de intéres al nivel de confianza del 95%, considerando partículas de materia oscura que tengan $m_\chi = 1$ GeV y acomplamientos $g = 1$, para cada tipo de mediador y para los tres años por separado. Se combinaron después estos límites para obtener los límites de exclusión finales que se presentan en la figura **??**. En esta figura, vemos que se espera con este análisis excluir los modelos de producción de materia oscura hasta 215 GeV y 250 GeV (180 y 220 observado) considerando mediadores de tipo scalar y pseudoscalar respectivamente. Esto significa que logramos mejorar en un factor 3 los límites scalares obtenidos por **CMS!** en 2016, que consideraba solamente el modelo $t\bar{t}$+DM en su estado final dileptónico, mientras que logramos excluir modelos de tipo pseudoscalar por primera vez gracias a este análisis.

APPENDIX B

Samples used

B.1 Data samples

All the data samples considered for this analysis are listed in Tables **??**, **??** and **??** for 2016, 2017 and 2018 respectively. The luminosity of each dataset has been computed using the Brilcalc tool provided by **CMS!** [?], while the number of generated events has been obtained using the CERN official **DAS!** (**DAS!**).

B.2 Signal samples

The **MC!** signal samples have first been produced privately and then centrally (for the $t\bar{t}$+DM model only). They were generated considering different dark matter and mediator masses and different production channels, for both our signal models. All the mass points considered along with their respective cross sections can be found in Tables **??** ($t/\bar{t}$+DM signal) and **??** ($t\bar{t}$+DM signal). The mass points generated are the same for 2016, 2017 and 2018.

B.3 Backgrounds samples

All the background **MC!** samples considered for this analysis, along with their respective cross-sections, are listed in Tables **??**, **??** and **??** for 2016, 2017 and 2018 respectively.

Dataset	Events (size)	$\mathcal{L}$ [fb^{-1}]
Run 2016B		
/DoubleEG/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOD	143073268 (99.4 Gb)	5.8
/DoubleMuon/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOD	82535526 (53.2 Gb)	
/MuonEG/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOD	32727796 (26.8 Gb)	
/SingleElectron/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOD	246440440 (167.8 Gb)	
/SingleMuon/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOD	158145722 (96.4 Gb)	
Run 2016C		
/DoubleEG/Run2016C-Nano02Apr2020-v1/NANOAOD	47677856 (35.3 Gb)	2.6
/DoubleMuon/Run2016C-Nano02Apr2020-v1/NANOAOD	27934629 (19.7 Gb)	
/MuonEG/Run2016C-Nano02Apr2020-v1/NANOAOD	15405678 (12.8 Gb)	
/SingleElectron/Run2016C-Nano02Apr2020-v1/NANOAOD	97259854 (69.3 Gb)	
/SingleMuon/Run2016C-Nano02Apr2020-v1/NANOAOD	67441308 (42.4 Gb)	
Run 2016D		
/DoubleEG/Run2016D-Nano02Apr2020-v1/NANOAOD	53324960 (39.6 Gb)	4.2
/DoubleMuon/Run2016D-Nano02Apr2020-v1/NANOAOD	33861745 (24.1 Gb)	
/MuonEG/Run2016D-Nano02Apr2020-v1/NANOAOD	23482352 (19.4 Gb)	
/SingleElectron/Run2016D-Nano02Apr2020-v1/NANOAOD	148167727 (104.4 Gb)	
/SingleMuon/Run2016D-Nano02Apr2020-v1/NANOAOD	98017996 (61.3 Gb)	
Run 2016E		
/DoubleEG/Run2016E-Nano02Apr2020-v1/NANOAOD	49877710 (37.9 Gb)	4.0
/DoubleMuon/Run2016E-Nano02Apr2020-v1/NANOAOD	28246946 (20.8 Gb)	
/MuonEG/Run2016E-Nano02Apr2020-v2/NANOAOD	22519303 (19.0 Gb)	
/SingleElectron/Run2016E-Nano02Apr2020-v1/NANOAOD	117321545 (86.5 Gb)	
/SingleMuon/Run2016E-Nano02Apr2020-v1/NANOAOD	90984718 (58.7 Gb)	
Run 2016F		
/DoubleEG/Run2016F-Nano02Apr2020-v1/NANOAOD	34577629 (26.9 Gb)	3.1
/DoubleMuon/Run2016F-Nano02Apr2020-v1/NANOAOD	20329921 (15.3 Gb)	
/MuonEG/Run2016F-Nano02Apr2020-v1/NANOAOD	16002165 (13. 6Gb)	
/SingleElectron/Run2016F-Nano02Apr2020-v1/NANOAOD	70593532 (51.4 Gb)	
/SingleMuon/Run2016F-Nano02Apr2020-v1/NANOAOD	65489554 (42.4 Gb)	
Run 2016G		
/DoubleEG/Run2016G-Nano02Apr2020-v1/NANOAOD	78797031 (61.6 Gb)	7.6
/DoubleMuon/Run2016G-Nano02Apr2020-v1/NANOAOD	45235604 (34.2 Gb)	
/MuonEG/Run2016G-Nano02Apr2020-v1/NANOAOD	33854612 (29.0 Gb)	
/SingleElectron/Run2016G-Nano02Apr2020-v1/NANOAOD	153363109 (109.2 Gb)	
/SingleMuon/Run2016G-Nano02Apr2020-v1/NANOAOD	149912248 (94.6 Gb)	
Run 2016H		
/DoubleEG/Run2016H-Nano02Apr2020-v1/NANOAOD	85388734 (67.7 Gb)	8.6
/DoubleMuon/Run2016H-Nano02Apr2020-v1/NANOAOD	48912812 (37.3 Gb)	
/MuonEG/Run2016H-Nano02Apr2020-v1/NANOAOD	29236516 (26.0 Gb)	
/SingleElectron/Run2016H-Nano02Apr2020-v1/NANOAOD	128854598 (93.8 Gb)	
/SingleMuon/Run2016H-Nano02Apr2020-v1/NANOAOD	174035164 (110.2 Gb)	

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

Dataset	Events (size)	$\mathcal{L}$ [fb^{-1}]
Run 2017B		
/DoubleEG/Run2017B-Nano02Apr2020-v1/NANOAOD	58088760 (46.6 Gb)	
/DoubleMuon/Run2017B-Nano02Apr2020-v1/NANOAOD	14501767 (10.8 Gb)	
/SingleElectron/Run2017B-Nano02Apr2020-v1/NANOAOD	60537490 (42.2 Gb)	4.8
/SingleMuon/Run2017B-Nano02Apr2020-v1/NANOAOD	136300266 (86.2 Gb)	
/MuonEG/Run2017B-Nano02Apr2020-v1/NANOAOD	4453465 (4.1 Gb)	
Run 2017C		
/DoubleEG/Run2017C-Nano02Apr2020-v1/NANOAOD	65181125 (53.8 Gb)	
/DoubleMuon/Run2017C-Nano02Apr2020-v1/NANOAOD	49636525 (39.5 Gb)	
/SingleElectron/Run2017C-Nano02Apr2020-v1/NANOAOD	136637888 (102.5 Gb)	9.7
/SingleMuon/Run2017C-Nano02Apr2020-v1/NANOAOD	165652756 (109.5 Gb)	
/MuonEG/Run2017C-Nano02Apr2020-v1/NANOAOD	15595214 (15.0 Gb)	
Run 2017D		
/DoubleEG/Run2017D-Nano02Apr2020-v1/NANOAOD	25911432 (21.6 Gb)	
/DoubleMuon/Run2017D-Nano02Apr2020-v1/NANOAOD	23075733 (18.6 Gb)	
/SingleElectron/Run2017D-Nano02Apr2020-v1/NANOAOD	51526710 (38.5 Gb)	4.2
/SingleMuon/Run2017D-Nano02Apr2020-v1/NANOAOD	70361660 (47.2 Gb)	
/MuonEG/Run2017D-Nano02Apr2020-v1/NANOAOD	9164365 (8.9 Gb)	
Run 2017E		
/DoubleEG/Run2017E-Nano02Apr2020-v1/NANOAOD	56233597 (49.8 Gb)	
/DoubleMuon/Run2017E-Nano02Apr2020-v1/NANOAOD	51589091 (44.4 Gb)	
/SingleElectron/Run2017E-Nano02Apr2020-v1/NANOAOD	102121689 (81.3 Gb)	9.3
/SingleMuon/Run2017E-Nano02Apr2020-v1/NANOAOD	154630534 (111.0 Gb)	
/MuonEG/Run2017E-Nano02Apr2020-v1/NANOAOD	19043421 (19.2 Gb)	
Run 2017F		
/DoubleEG/Run2017F-Nano02Apr2020-v1/NANOAOD	74307066 (67.1 Gb)	
/DoubleMuon/Run2017F-Nano02Apr2020-v1/NANOAOD	79756560 (68.0 Gb)	
/SingleElectron/Run2017F-Nano02Apr2020-v1/NANOAOD	128467223 (105.2 Gb)	13.5
/SingleMuon/Run2017F-Nano02Apr2020-v1/NANOAOD	242135500 (178.3 Gb)	
/MuonEG/Run2017F-Nano02Apr2020-v1/NANOAOD	25776363 (26.3 Gb)	

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

Dataset	Events (size)	$\mathcal{L}$ [fb^{-1}]
Run 2018A		
/DoubleMuon/Run2018A-Nano02Apr2020-v1/NANOAOD	75499908 (62.6 Gb)	
/EGamma/Run2018A-Nano02Apr2020-v1/NANOAOD	327843843 (261.8 Gb)	13.5
/SingleMuon/Run2018A-Nano02Apr2020-v1/NANOAOD	241608232 (167.7 Gb)	
/MuonEG/Run2018A-Nano02Apr2020-v1/NANOAOD	32958503 (32.3 Gb)	
Run 2018B		
/DoubleMuon/Run2018B-Nano02Apr2020-v1/NANOAOD	35057758 (28.3 Gb)	
/EGamma/Run2018B-Nano02Apr2020-v1/NANOAOD	153822427 (123.1 Gb)	6.8
/SingleMuon/Run2018B-Nano02Apr2020-v1/NANOAOD	119918017 (82.3 Gb)	
/MuonEG/Run2018B-Nano02Apr2020-v1/NANOAOD	16211567 (15.8 Gb)	
Run 2018C		
/DoubleMuon/Run2018C-Nano02Apr2020-v1/NANOAOD	34565869 (27.6 Gb)	
/EGamma/Run2018C-Nano02Apr2020-v1/NANOAOD	147827904 (119.2 Gb)	6.6
/SingleMuon/Run2018C-Nano02Apr2020-v1/NANOAOD	110032072 (75.7 Gb)	
/MuonEG/Run2018C-Nano02Apr2020-v1/NANOAOD	15652198 (15.3 Gb)	
Run 2018D		
/DoubleMuon/Run2018D-Nano02Apr2020_ver2-v1/NANOAOD	168605834 (128.6 Gb)	
/EGamma/Run2018D-Nano02Apr2020-v1/NANOAOD	751348648 (583.6 Gb)	32.0
/SingleMuon/Run2018D-Nano02Apr2020-v1/NANOAOD	513867253 (344.5 Gb)	
/MuonEG/Run2018D-Nano02Apr2020_ver2-v1/NANOAOD	71961587 (68.6 Gb)	

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

Mass point	Cross-section [pb]
Scalar mediators	
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi10	$4.959 \cdot 10^{-2}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi20	$3.235 \cdot 10^{-2}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi50	$1.323 \cdot 10^{-2}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi100	$5.633 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi150	$3.397 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi200	$2.359 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi250	$1.720 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi300	$1.328 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi350	$1.018 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi400	$6.717 \cdot 10^{-4}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi450	$4.535 \cdot 10^{-4}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi500	$3.206 \cdot 10^{-4}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi1000	$3.045 \cdot 10^{-5}$
Pseudoscalar mediators	
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi10	$6.151 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi20	$5.869 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi50	$4.946 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi100	$3.658 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi150	$2.754 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi200	$2.097 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi250	$1.616 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi300	$1.253 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi350	$7.851 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi400	$4.371 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi450	$3.095 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi500	$2.321 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi1000	$2.791 \cdot 10^{-5}$

Table B.4: Signal samples mass points and **LO!** dilepton cross-sections considered for the $t/\bar{t}$+DM signal model used in this analysis.

Mass point	Cross-section [pb]
Scalar mediators	
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_50	$3.405 \cdot 10^{-1}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_100	$8.027 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_150	$2.673 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_200	$1.158 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_250	$6.020 \cdot 10^{-3}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_300	$3.579 \cdot 10^{-3}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_350	$2.376 \cdot 10^{-3}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_400	$1.443 \cdot 10^{-3}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_450	$9.025 \cdot 10^{-4}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_500	$6.204 \cdot 10^{-4}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_20_mPhi_100	$7.993 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_30_mPhi_100	$8.052 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_40_mPhi_100	$8.147 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_45_mPhi_100	$8.319 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_49_mPhi_100	$8.304 \cdot 10^{-2}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_51_mPhi_100	$9.735 \cdot 10^{-4}$
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_55_mPhi_100	$4.835 \cdot 10^{-4}$
Pseudoscalar mediators	
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_50	$3.440 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_100	$2.164 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_150	$1.414 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_200	$9.773 \cdot 10^{-3}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_250	$6.753 \cdot 10^{-3}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_300	$4.808 \cdot 10^{-3}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_350	$2.742 \cdot 10^{-3}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_400	$1.409 \cdot 10^{-3}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_450	$9.302 \cdot 10^{-4}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_500	$6.618 \cdot 10^{-4}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_20_mPhi_100	$2.166 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_30_mPhi_100	$2.164 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_40_mPhi_100	$2.162 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_45_mPhi_100	$2.180 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_49_mPhi_100	$2.151 \cdot 10^{-2}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_51_mPhi_100	$1.993 \cdot 10^{-3}$
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_55_mPhi_100	$7.750 \cdot 10^{-4}$

Table B.5: Signal samples mass points and **LO!** dilepton cross-sections considered for the $t\bar{t}$+DM signal model used in this analysis.

Process	Sample	Cross section [pb]
Drell-Yan	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18610.0
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 ($H_T < 70$ GeV)	6077.22
	DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	169.9
	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	147.4
	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	40.99
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.678
	DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.367
	DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.6304
	DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.1514
	DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.003565
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	80.95
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	136.02
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85
TTToSemiLeptonic	TTToSemilepton_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	364.35
ttV	TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.2814
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
VZ	WWTo2L2Nu_13TeV-powheg	12.178
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.42965
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_powheg_pythia8	3.22
Others	WWW, WWZ, WZZ, ZZZ, WWG	//

Table B.6: Main 2016 **MC!** simulations for the different background processes considered for this analysis and their respective cross sections.

Process	Sample	Cross section [pb]
Drell-Yan	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 ($H_T < 70$ GeV)	6077.22
	DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8	169.9
	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	147.4
	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	40.99
	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.678
	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.367
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.6304
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1514
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003565
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	80.95
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
TTToSemiLeptonic	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	364.35
ttV	TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.2814
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
VZ	WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8	12.178
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.42965
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
Others	WWW, WWZ, WZZ, ZZZ, WWG	//

Table B.7: Main 2017 **MC!** simulations for the different background processes considered for this analysis and their respective cross sections.

Process	Sample	Cross section [pb]
Drell-Yan	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 ($H_T < 70$ GeV)	6077.22
	DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	169.9
	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	147.4
	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	40.99
	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.678
	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.367
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.6304
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1514
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003565
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	80.95
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
TTToSemiLeptonic	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	364.35
ttV	TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.2814
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
VZ	WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8	12.178
	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	4.42965
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
	ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
Others	WWW, WWZ, WZZ, ZZZ, WWG	//

Table B.8: Main 2018 **MC!** simulations for the different background processes considered for this analysis and their respective cross sections.

APPENDIX C

Non prompt leptons contamination

Even though not extremely important in the sense that its kinematics allows us to remove most of the contribution from this particular background anyway in the signal regions, this background is interesting in the sense that it can be estimated using a data-driven method instead of being taken directly from **MC!**.

A few definitions are first of all needed to explain this 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 actual **PV!** of a *pp* collision;
- The **PR! (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 **??**, and real leptons coming from eventual heavy flavour decays;
- The **FR! (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.

This background is particularly important at low p_T, where the misidentification rate is higher, and is usually not expected to be modeled correctly by **MC!** because of its complexity: a general **tight-to-loose data-driven 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

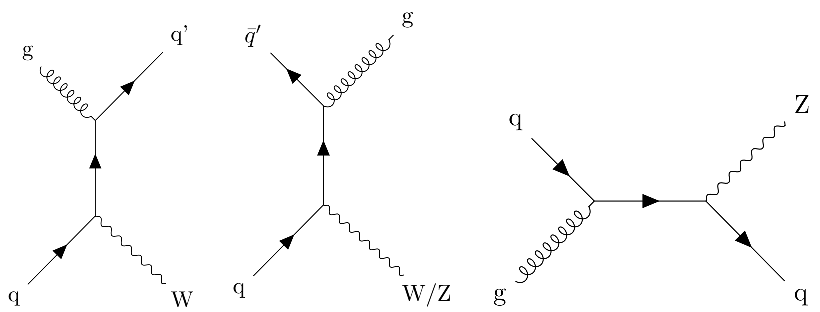


Figure C.1: Possible Feynman diagrams for the production of a W/Z boson with a jet.

the signal region of the analysis and the definition of a same sign control region enriched in fakes in order to perform a closure test of the yields and kinematics of this background.

FR! (FR!) computation

Because of its definition, the **FR!** is computed in a prompt lepton-free region, typically in a l1loose **QCD!** enriched region, defined with the following cuts:

- Exactly 1 lepton
- $p_T > 13$ (10) GeV for e (μ)
- $|\eta| < 2.5$ (2.4) for e (μ)
- mtw1 < 20 GeV
- pf**MET!** < 20 GeV
- PassJets

All the previous cuts have been designed to define a l1loose **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, ranging 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 **??**.

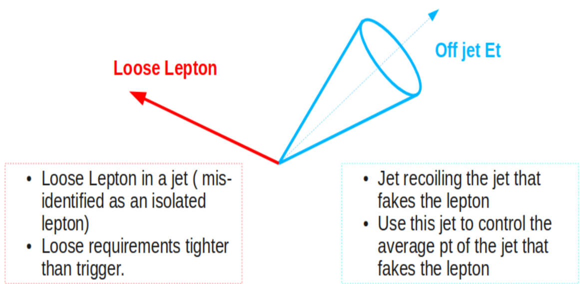


Figure C.2: Schematic representation of the two jets used for the systematics and for the jet faking a lepton in the tight-to-loose data-driven method.

Events passing the following pre-scaled triggers are then selected in this region:

$$\text{Muon triggers} = \begin{cases} \text{HLT\_Mu8\_TrkIsoVVL (if } p_T < 20 \text{ GeV)} \\ \text{HLT\_Mu17\_TrkIsoVVL (if } p_T \geq 20 \text{ GeV)} \end{cases} \tag{C.1a}$$

$$\text{Electron triggers} = \begin{cases} \text{HLT\_Ele8\_CaloIdL\_TrackIdL\_IsoVL\_PFJet30 (if } p_T < 25 \text{ GeV)} \\ \\ \text{HLT\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\_PFJet30 (if } p_T \geq 25 \text{ GeV)} \end{cases} \tag{C.1b}$$

The remaining of the **EWK!** (**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 chosen for the leptons of the analysis, they are computed separately depending on the flavour of the lepton and 2D histograms (accounting for the p_T and η of the lepton) 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 **??**, as they are usually easier to interpret.

PR! (PR!) computation

The **PR!**, taking into account the real lepton contamination in the region defined, is also important to calculate, even though the objects working points are usually chosen is such a way that this ratio is quite close to 1 and can therefore typically 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 $Z \to ll$ 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 **??**.

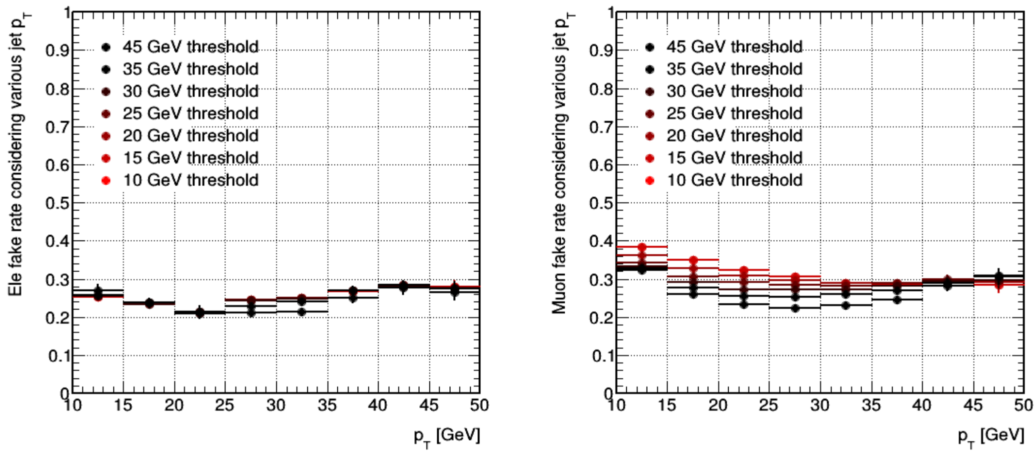
Fake weight calculation

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 signal regions of the analysis (usually, a l2loose region). For this, a simple set of equations can be used.

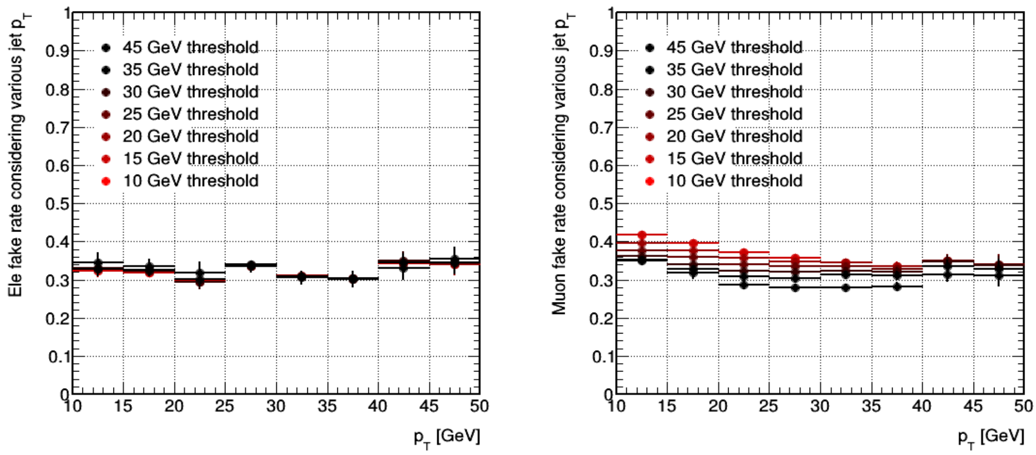
We start by defining the following quantities:

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

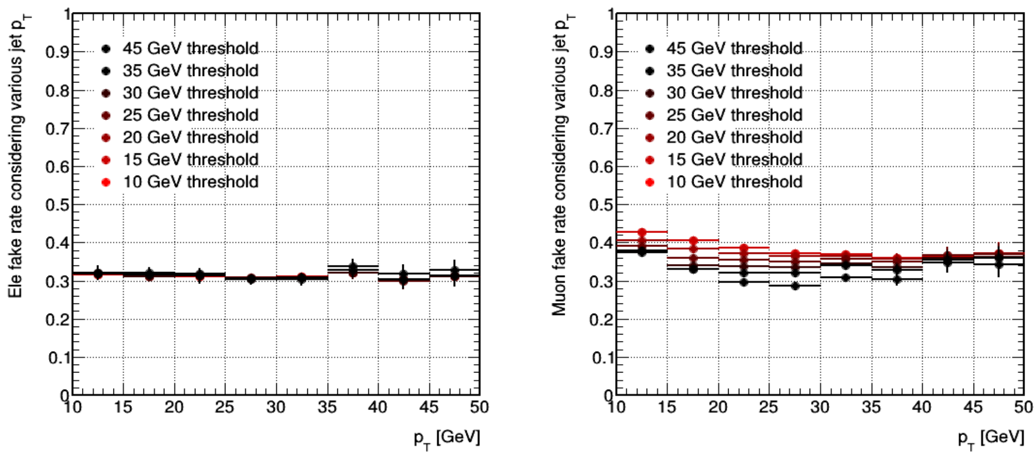
It is then possible to see in Equation **??** that, if p is the **PR!** and f the **FR!** previously calculated,



(a) 2016 electron and muon fake rates

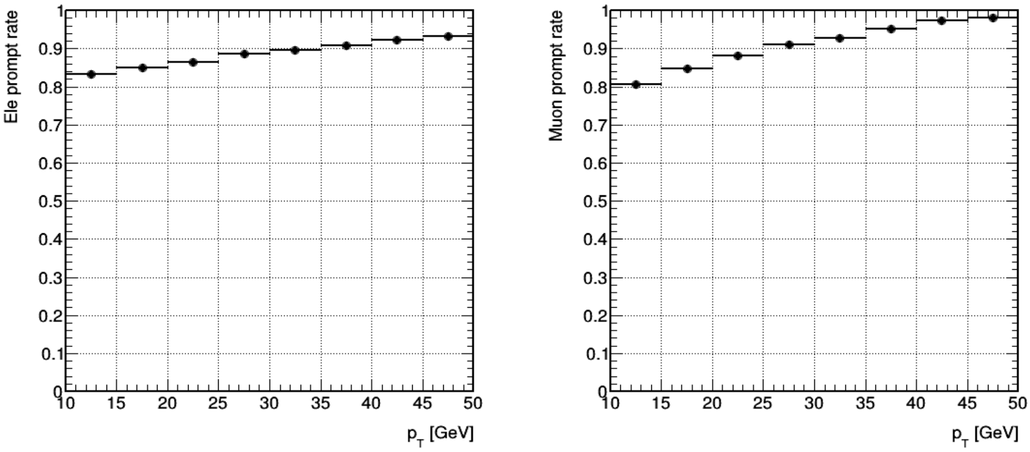


(b) 2017 electron and muon fake rates

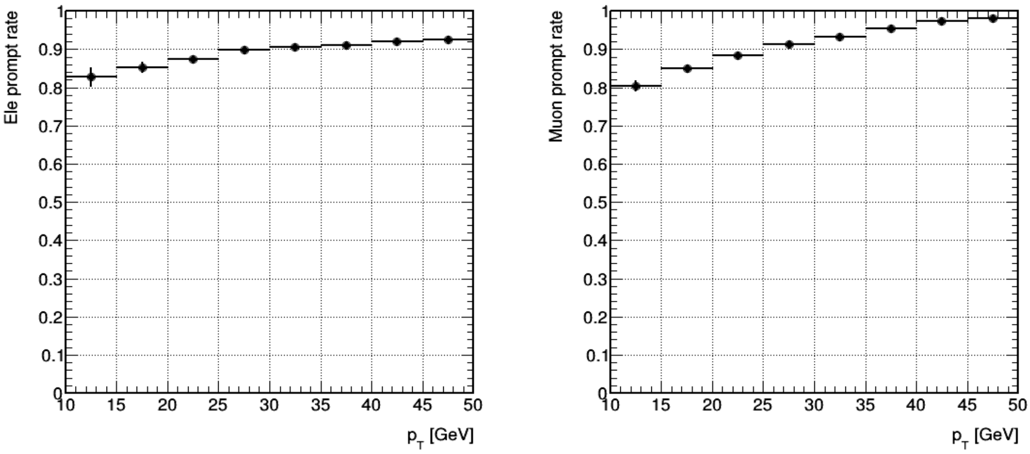


(c) 2018 electron and muon fake rates

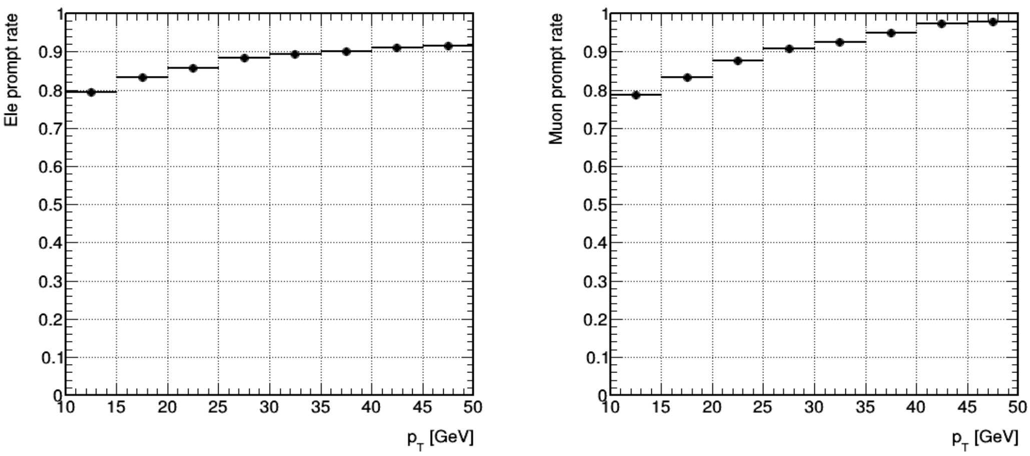
Figure C.3: Electron (on the left) and muon (on the right) **FR!** obtained in a QCD enriched region for different jet E_T thresholds for 2016, 2017 and 2018 with respect to the p_T of the lepton.



(a) 2016 electron and muon prompt rates



(b) 2017 electron and muon prompt rates



(c) 2018 electron and muon prompt rates

Figure C.4: Electron (on the left) and muon (on the right) **PR!** obtained in a Z+jets enriched region by a tag and probe method for 2016, 2017 and 2018 with respect to the p_T of the lepton.

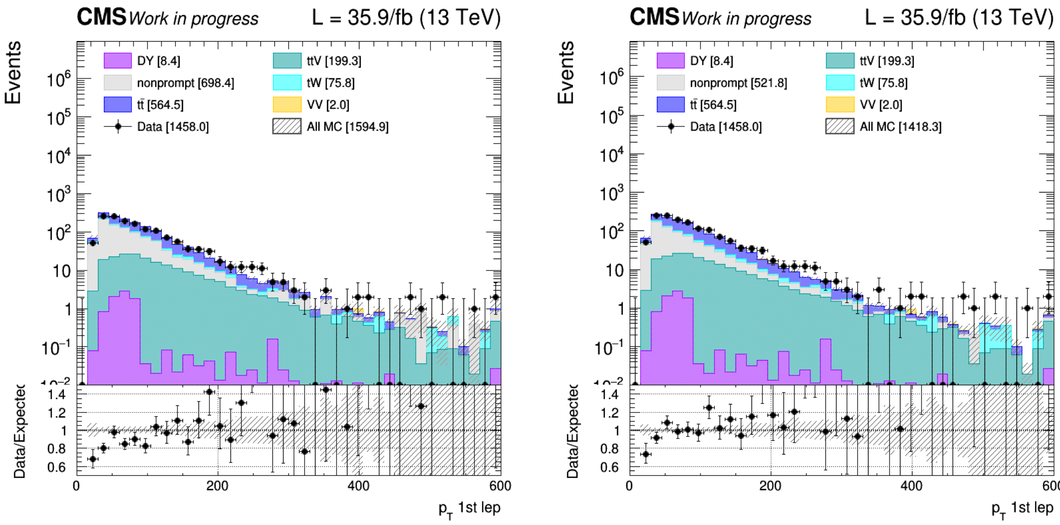
we can find an expression relating all these quantities.

$$\begin{cases} N_l = N_{pp} + N_{fp} + N_{ff} = N_{t2} + N_{t1} + N_{t0} \\ N_{t0} = (1-p)^2 N_{pp} + (1-p)(1-f) N_{fp} + (1-f)^2 N_{ff} \\ N_{t1} = 2p(1-p) N_{pp} + \big(f(1-p) + p(1-f)\big) N_{fp} + 2f(1-f)^2 N_{ff} \\ N_{t2} = p^2 N_{pp} + pf N_{fp} + f^2 N_{ff} \end{cases} \tag{C.2}$$

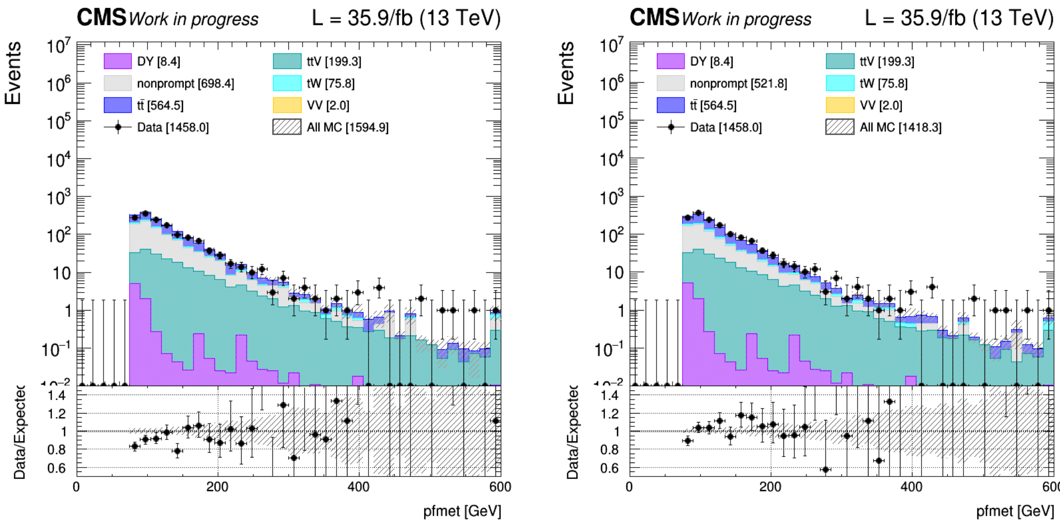
These equations can be inverted in order to represent the unknowns with respect to the known variables, as shown in Equation **??**, giving us a way to apply the weights previously calculated to this particular l2loose region.

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

The non-prompt contamination of this particular analysis however is expected to be dominated by the semileptonic decay of the **SM!** $t\bar{t}$ process, when one of the bottom quarks produced leads to a fake lepton. This particular process, contrary to the W+jets and Z+jets processes, is expected to be modeled quite well by the **MC!** simulations (this assumption can be checked in Figure **??**, obtained in a 2018 same sign control region, in which we do not observe major differences between the two ways we have to estimate the impact of this process). We therefore decided to rely on the **MC!** simulations only for this analysis mainly because they typically come with a smaller systematic uncertainty than the data-driven estimation for this process.



(a) Leading lepton p_T



(b) **MET!** distribution

Figure C.5: Some distributions obtained in a 2016 same sign control region considering the data-driven production of fakes (on the left) and the **MC!** simulation for this process (on the right).

APPENDIX D

MVA! optimization

The study and optimization of the different hyperparameters defining either the **ANN!** and the **BDT!** considered in this work was an important part of the analysis, to make sure to get the best possible discrimination between the backgrounds and the two signals of interest, the $t/\bar{t}$+DM and the $t\bar{t}$+DM. In general, such optimization is a complex task to perform mainly because of the high number of hyperparameters that can have an impact on the quality of the final results. As a starting point, the optimal parameters found by DESY (for the **BDT!**) and IFCA (for the **ANN!**) during the 2016 analysis were used.

The global strategy used for the actual optimization process is quite straightforward: even though some of the hyperparameters considered are actually correlated with each other, we decided for the sake of simplicity to treat them as independent and to optimize the model by optimizing each of these parameters individually. Among all the candidate values for a given parameter, the one leading to the best possible ROC curve while keeping the time needed to train the network reasonably low was then simply chosen for the analysis. Finally, the training optimization has been performed using the data collected in 2018 and considering the 100 GeV scalar mediators for both signal processes, an intermediate mass point that should feature some of the lowest discrimination between the signals and the backgrounds, given their kinematic similarities in this case.

Boosted Decisions Trees

The first obvious hyperparameter among all the parameters typical of a **BDT!**, described in Section **??**, that we decided to optimize was the number of trees, whose impact on the final discrimination obtained between the processes can be seen in the **ROC!** curves shown in Figure **??**.

Another important parameter of the **BDT!** which has been optimized is the maximal depth of the

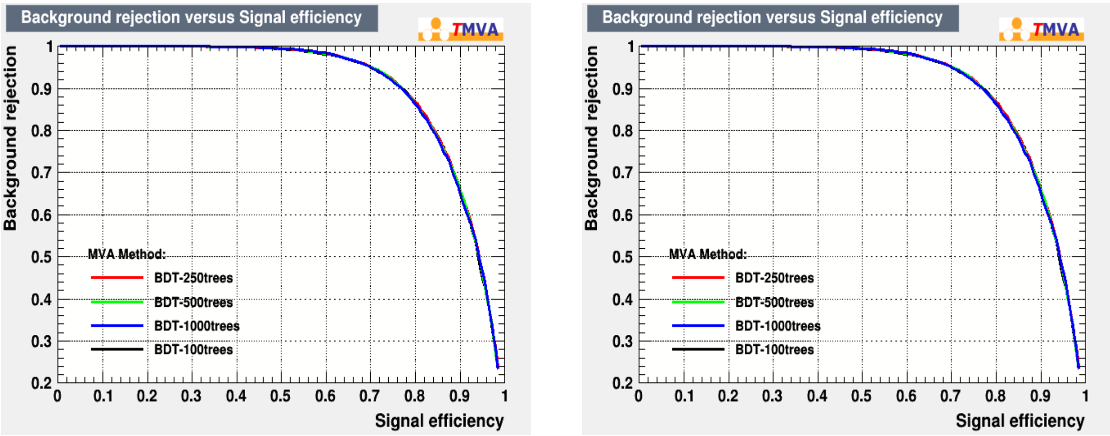


Figure D.1: Background rejection achieved for the **BDT!** by trying out different number of trees, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

trees that we define, as shown in Figure **??**. Given the relative small number of input variables used in this analysis, this parameter does not seem to be highly relevant since the results obtained are similar in all the cases.

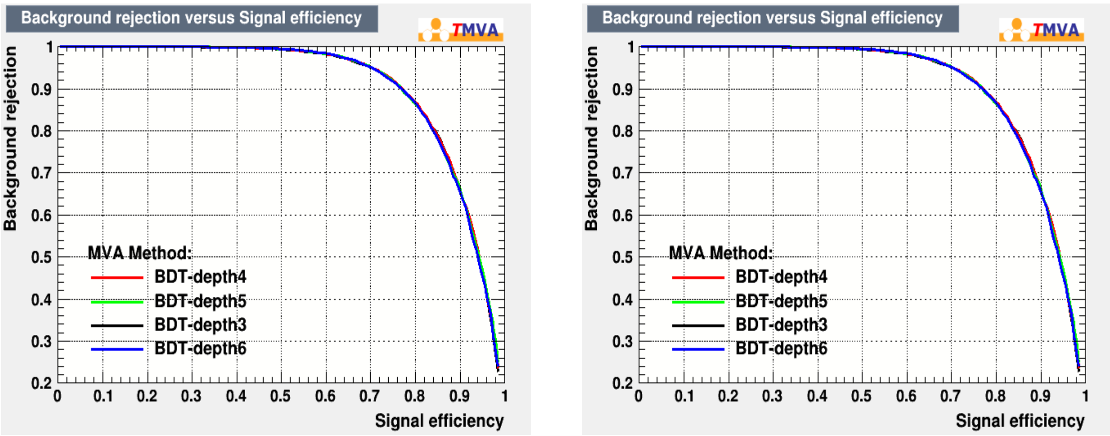


Figure D.2: Background rejection achieved for the **BDT!** by trying out different maximal depths, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

The n_{cut} parameter, corresponding to the granularity used when scanning over the variable range to find the optimal splitting criterion has also been studied, as shown in Figure **??**. Finally, the impact that the shrinkage, or learning rate for the GradBoost algorithm, has on the final discrimination results has also been studied and is shown in Figure **??**.

For most of the hyperparameters described so far, we usually expect that increasing (or decreasing) their values will result in better discrimination results but unfortunately, this usually also increases the time needed to train the **BDT!** so a compromise had to be found for each one of the hyperparameters described here, settling at the end for the set of parameters described in Table **??**.

Analysis Neural Network

The same process has then been repeated by considering this time the **ANN!** of the analysis. In this case, we decided to first of all focus the optimization of the neural network on its actual architecture, trying out different numbers of hidden layers and neurons in each layer, while trying to keep the validation loss obtained to a minimum. The results obtained for the optimization of this first parameter can be found in Figure **??**. We can see in these plots that with the 70%/30%

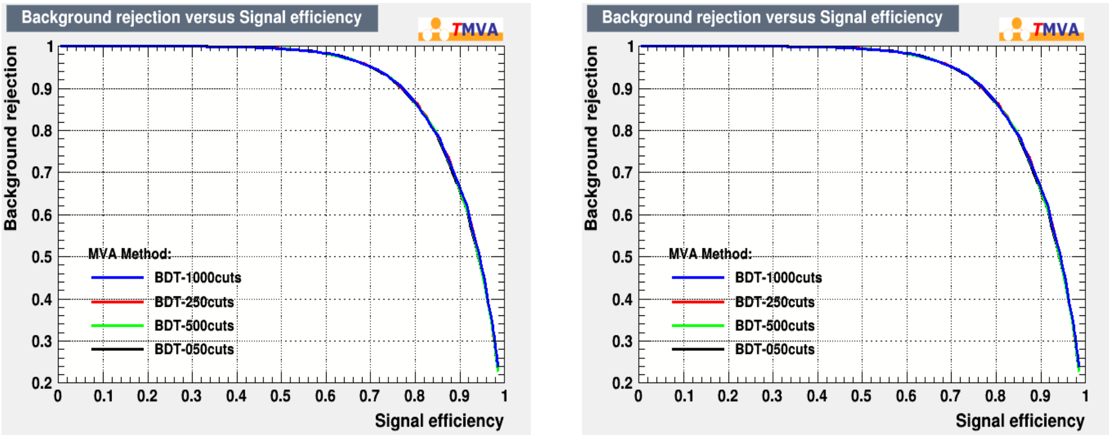


Figure D.3: Background rejection achieved for the **BDT!** by trying out different n_{cut} values, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

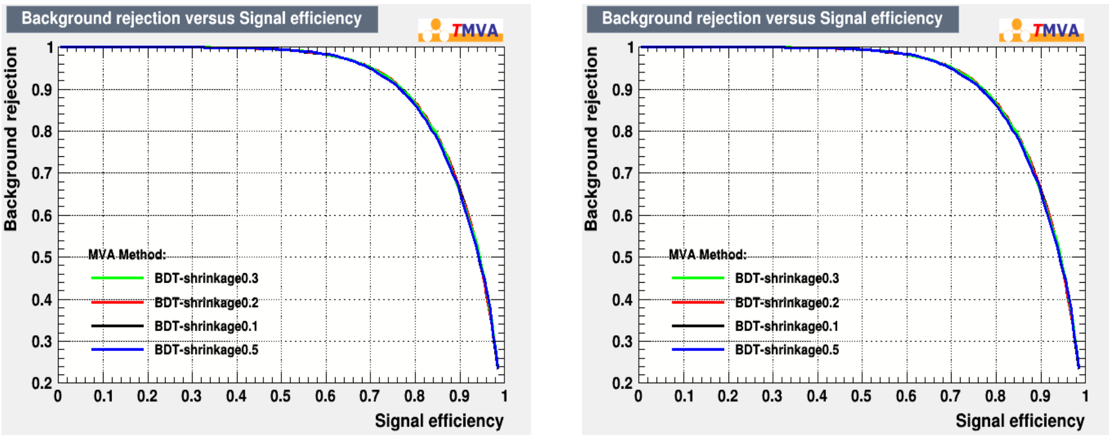


Figure D.4: Background rejection achieved for the **BDT!** by trying out different shrinkage values, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

training/test splitting used, which gave us around 50.000 total events for the training, all the neural networks defined seem to give similar results. At the end of the day though, slightly better results were obtained for the architecture with 4 hidden layers made out of 20/15/15/10 neurons, at least when looking at the combination between the background rejection and signal efficiency obtained.

After this first step of optimization, we kept the architecture which seemed to give the best results and started a subsequent optimization process, evaluating this time the learning rate of the Adam algorithm used, as shown in Figure **??**. Usually, a larger learning rate allow us to avoid any eventual local minimum in the curve corresponding to the error function defined and to reach the minimum global faster, but also increases the risk of not being able to converge to a solution at all, while a lower value is typically a bit more precise but also increases the training time needed in order to reach an optimal solution. Again, this trade-off implies that a comprise needs to be found when choosing the value of this hyperparameter as well.

The last parameter that has been optimized in this case is the batch size. Usually, the weights of the neural network have to be updated at each training epoch and this is done by looking at the whole dataset, which in reality can prove to be quite inefficient. This hyperparameter allows us to gain some time by only looking at a small number of events in the dataset before updating the weights of the neural network, therefore accelerating the training process. The results obtained considering different batch sizes are shown in Figure **??**. In this case, time was an important factor when choosing the parameter to use for the analysis, given the fact that it takes around two to three more times to train a **ANN!** with a batch size 20 than a higher value of 100.

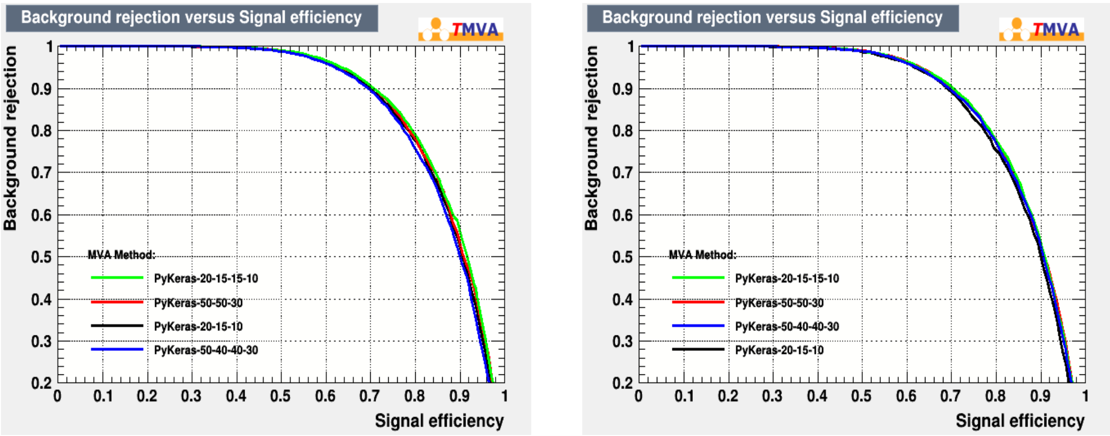


Figure D.5: Background rejection achieved for the **ANN!** by trying out different architectures, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

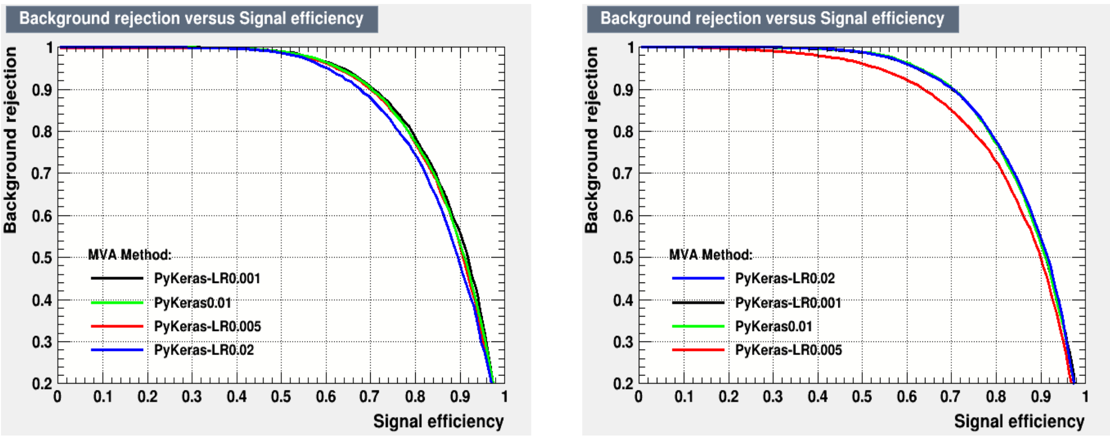


Figure D.6: Background rejection achieved for the **ANN!** by trying out different learning rates, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

For all the previous tests, we took care of defining a number of training iterations high enough by using a simple trial and error process, making sure to run at least 200 epochs, while implementing a system of early stop in case the validation loss did not decrease for several epochs in a row, in order to avoid overfitting the training data.

Input variables chosen

Finally, the input variables given to the **MVA!** is also an extremely important parameter to study. In order to study this parameter, we decided to train several times the **BDT!** and the **ANN!** considering each time a different set of input variables, all defined in Section **??**, as shown in Table **??**.

The **ROC!** curves for each set of input variables have been obtained and plotted on top of each other in order to simplify their comparison, as shown in Figure **??**. At the end of the day, the set of input variables number 5 was chosen over the others as it seems to be giving the best possible signal over background discrimination.

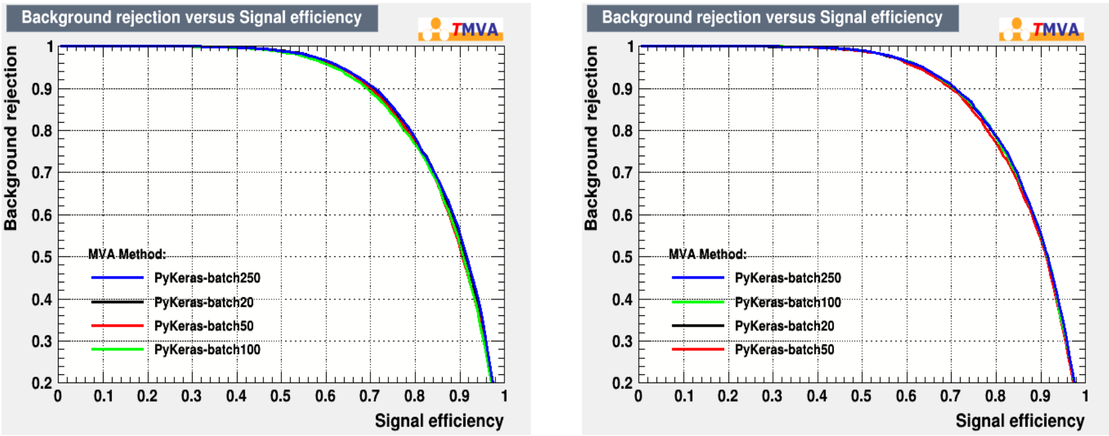
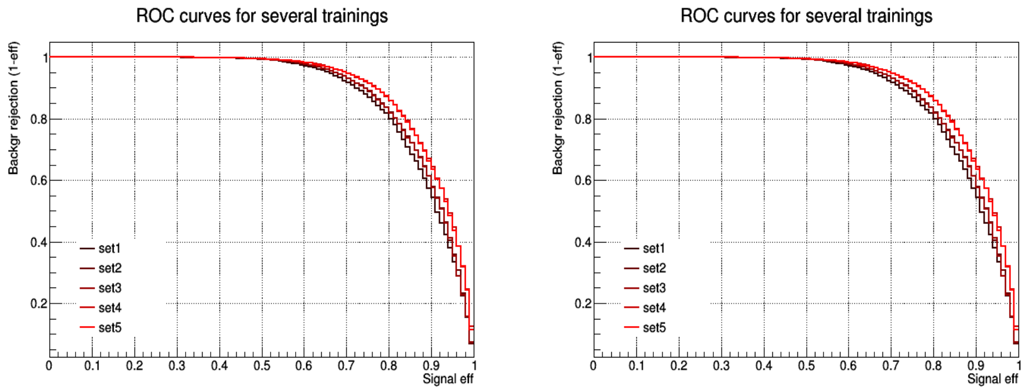


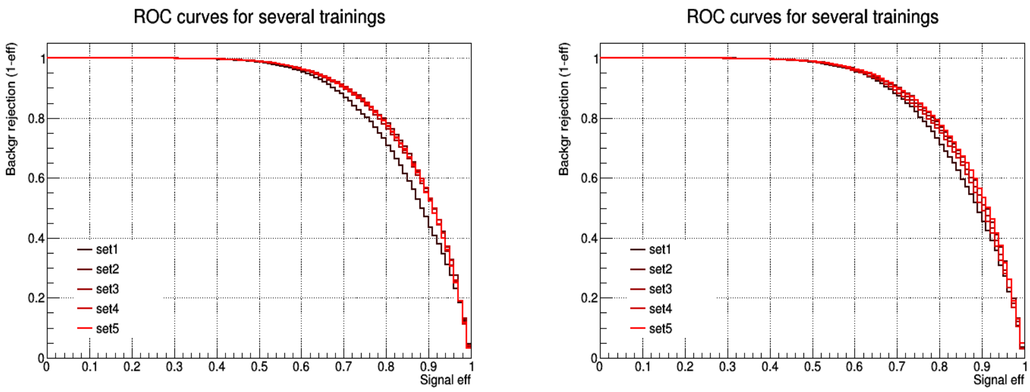
Figure D.7: Background rejection achieved for the **ANN!** by trying out different batch sizes, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) regions.

	Variables
Set 1	Number of b-jets
	m^t_{bl}
	pf**MET!**
	$\Delta\Phi(E_T^{\text{miss}}, ll)$
	M^{ll}_{T2}
Set 2	**Set 1 and**
	M^{bl}_{T2}
	massT
Set 3	**Set 2 and**
	Spin correlated variables
Set 4	**Set 3 and**
	$r2l$
	$r2l4j$
Set 5	**Set 4 and**
	reco weight w
	Dark p_T
	Overlapping factor R

Table D.1: Different sets of input variables considered for the optimization of both the **MVA!** methods considered.



(a) **BDT!** ROC curves



(b) **ANN!** ROC curves

Figure D.8: Background rejection achieved for the **BDT!** (at the top) and the **ANN!** (at the bottom) by trying out different sets of input variables, for the training done in the $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.

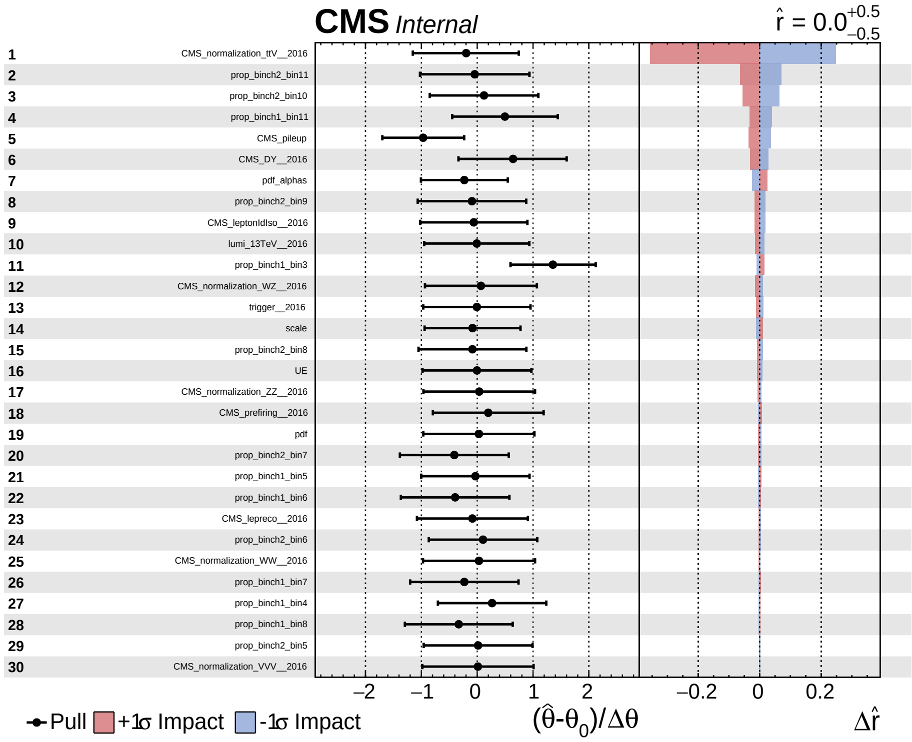
APPENDIX E

Post-fits, pulls and impacts plots

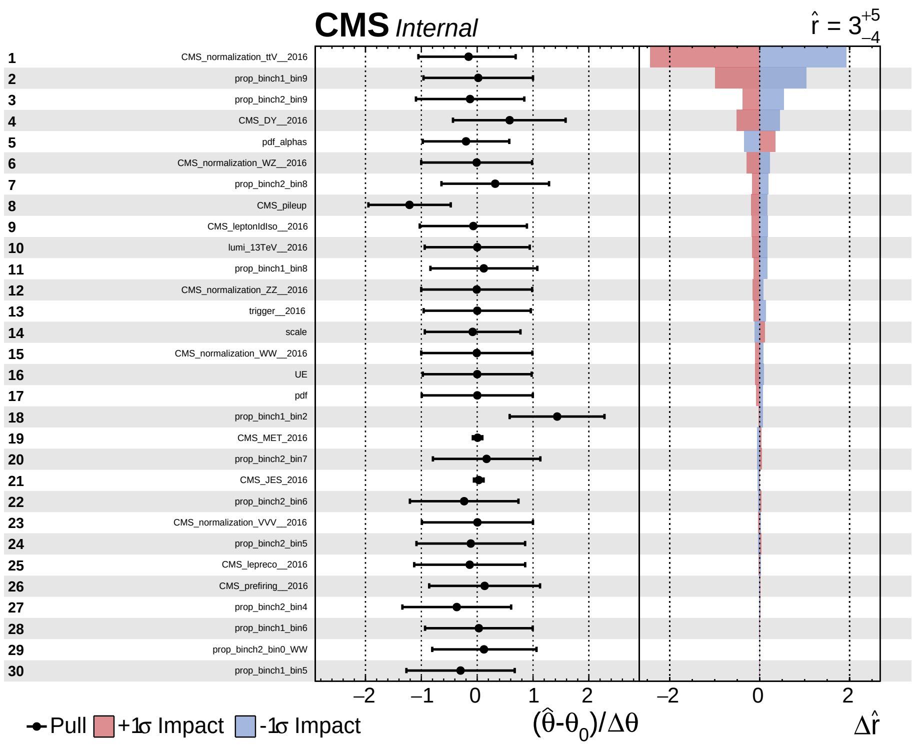
Given the large number of systematic uncertainties considered in this analysis, all described in Section **??**, it is first of all useful to try and understand the impact each one of them has on the final upper limits on the signal strengths obtained. This is where the so-called impact plots become useful, since they allow us to study the quality of the systematic estimation, expressed as the difference between the maximum likelihood estimator and our own estimation, along with the importance of every systematic of the analysis, according to the effect each systematic has on the signal strength μ. This is estimated by rerunning the fit performed with each nuisance parameter fixed at their $\pm 1\sigma$ values, to understand the impact they might have. Mathematically, the impact of a nuisance parameter θ on a parameter of interest μ (in this, the signal strength) is defined as the shift $\Delta\mu$ this is induced by bringing θ to its $+1\sigma$ and -1σ values.

The impact plots obtained for different mediators are shown in Figures **??** to **??**, for several data taking periods, mediator categories and mass points. In such plots, we can observe that the dominant uncertainty for this analysis is coming from the normalization of the ttV process.

Additionally, plots after performing the fit using the combine tool, the so-called post-fit plots, were also obtained, allowing us to compare the **BDT!** output shape before and after performing this fit. Some of the distributions obtained in both signal regions can be found in Figures **??** to **??**. In general, we can conclude from these results that the fit is doing a good job and looks healthy, as the post-fits look globally better than the pre-fits.

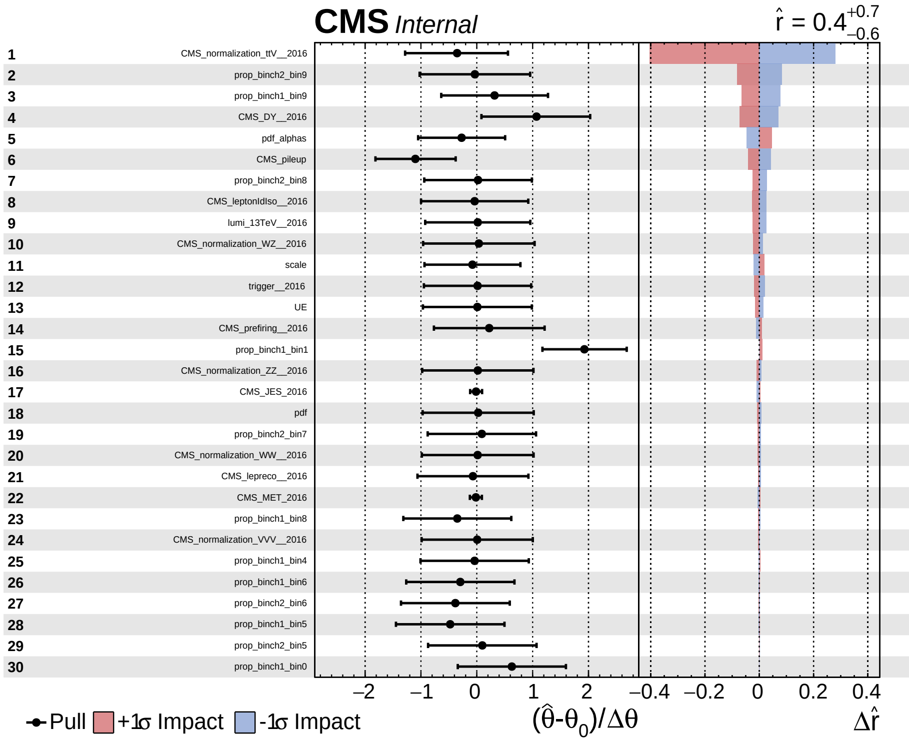


(a) 100 GeV

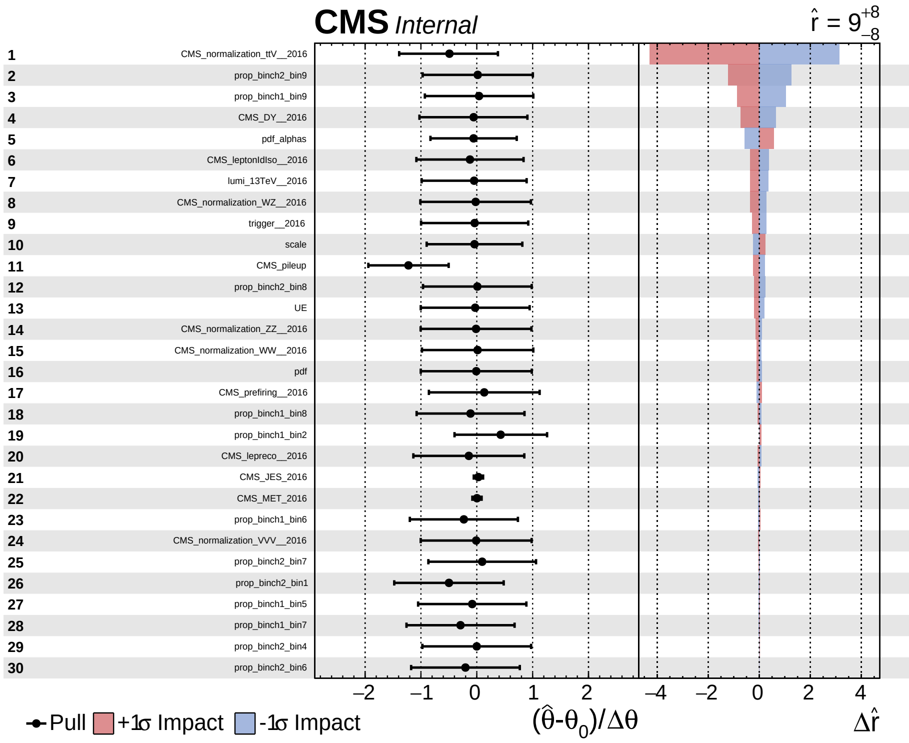


(b) 500 GeV

Figure E.1: 2016 impacts and pulls obtained for the most important systematics of the analysis, considering the 100 and 500 GeV scalar mediators.

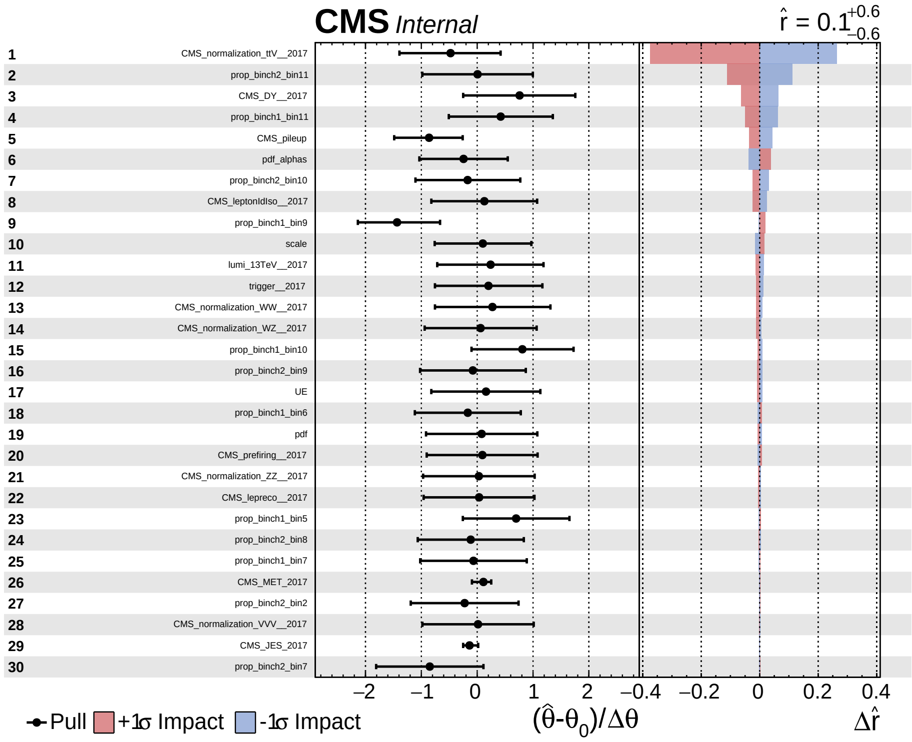


(a) 100 GeV

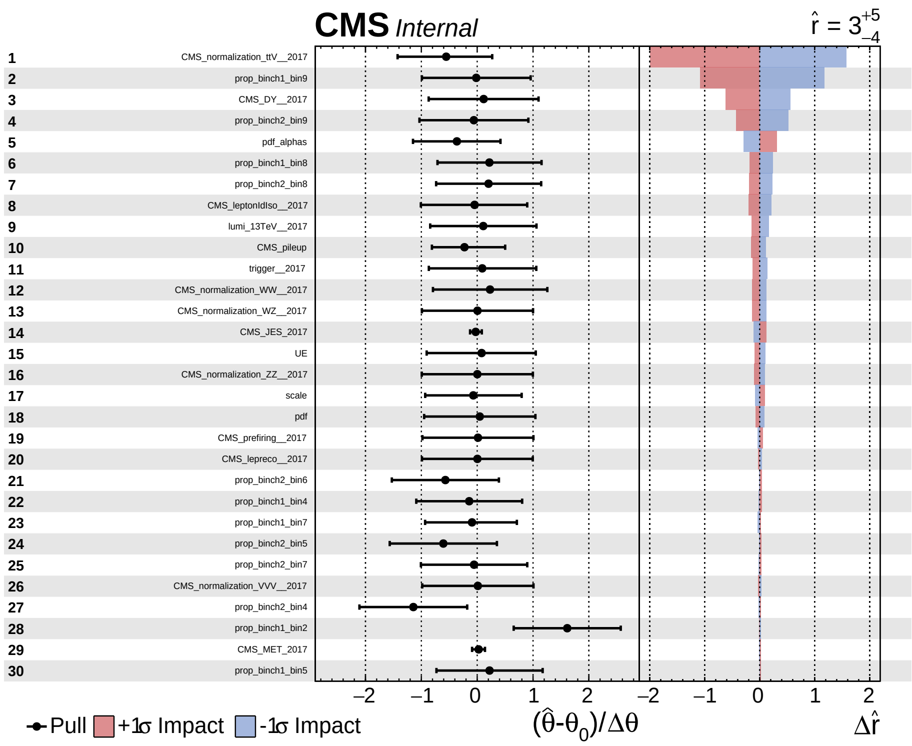


(b) 500 GeV

Figure E.2: 2016 impacts and pulls obtained for the most important systematics of the analysis, considering the 100 and 500 GeV pseudoscalar mediators.

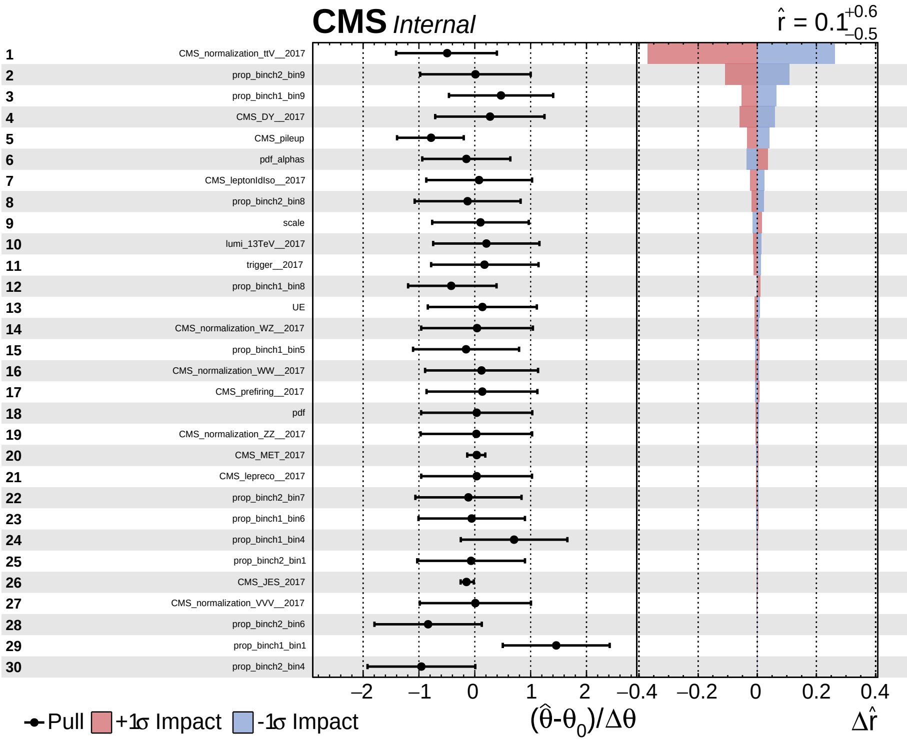


(a) 100 GeV

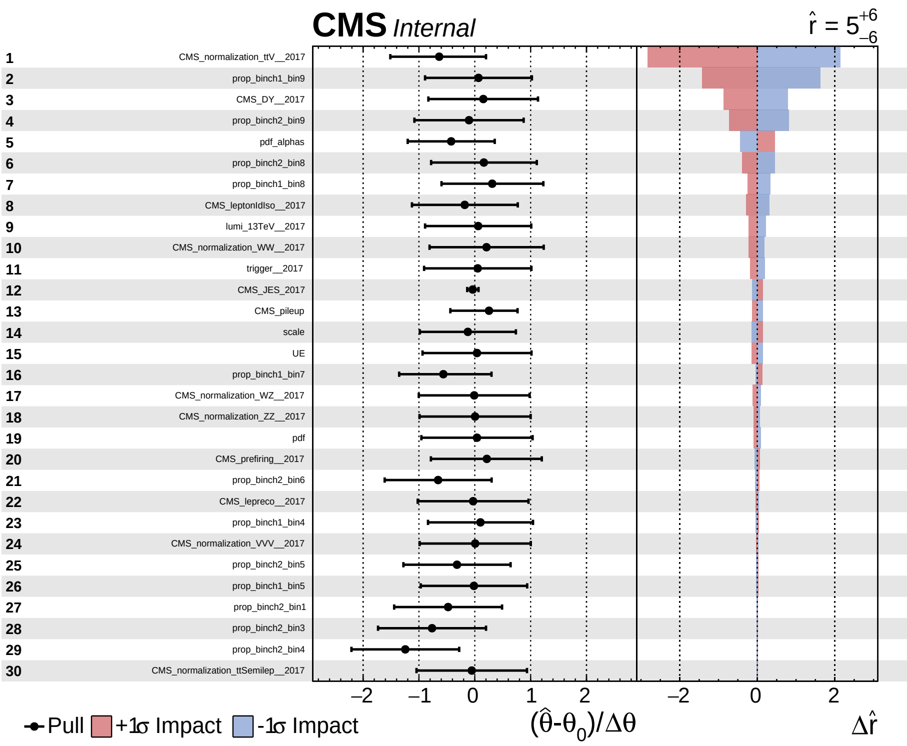


(b) 500 GeV

Figure E.3: 2017 impacts and pulls obtained for the most important systematics of the analysis, considering the 100 and 500 GeV scalar mediators.

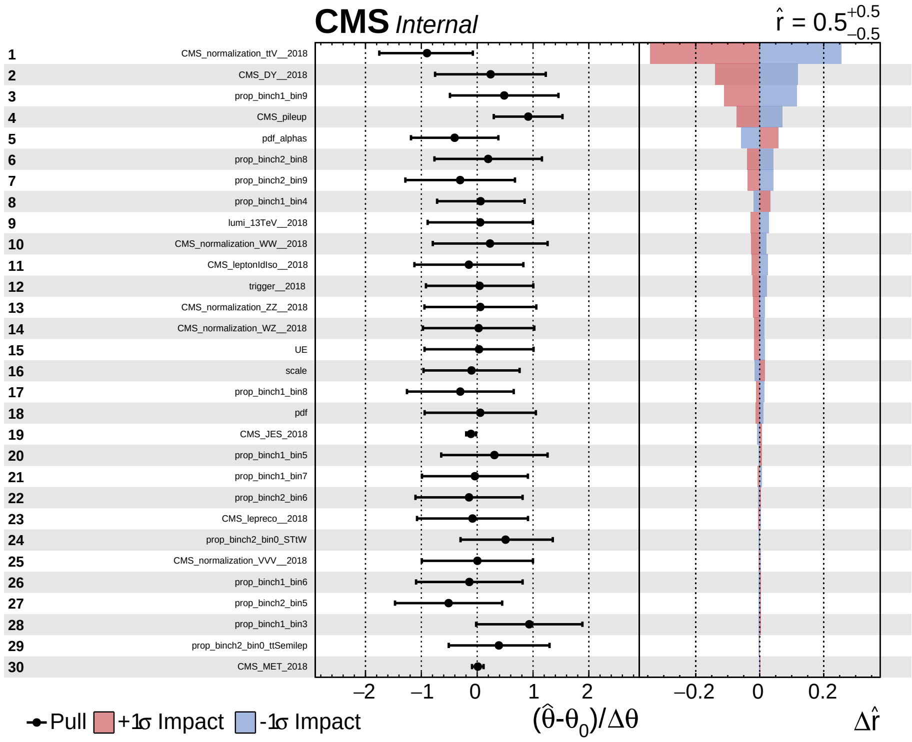


(a) 100 GeV

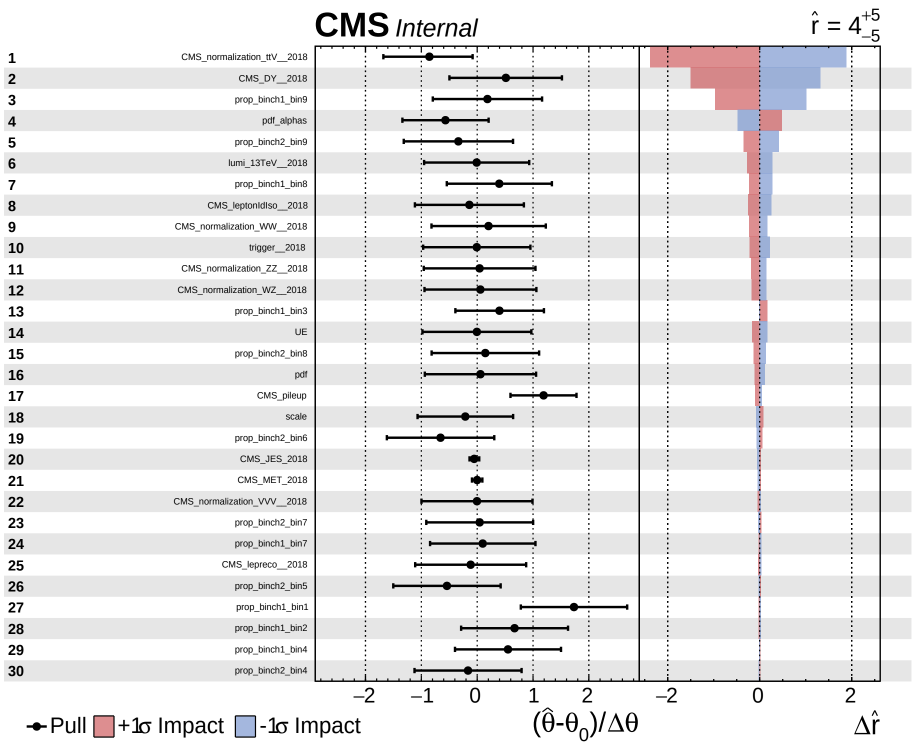


(b) 500 GeV

Figure E.4: 2017 impacts and pulls obtained for the most important systematics of the analysis, considering the 100 and 500 GeV pseudoscalar mediators.

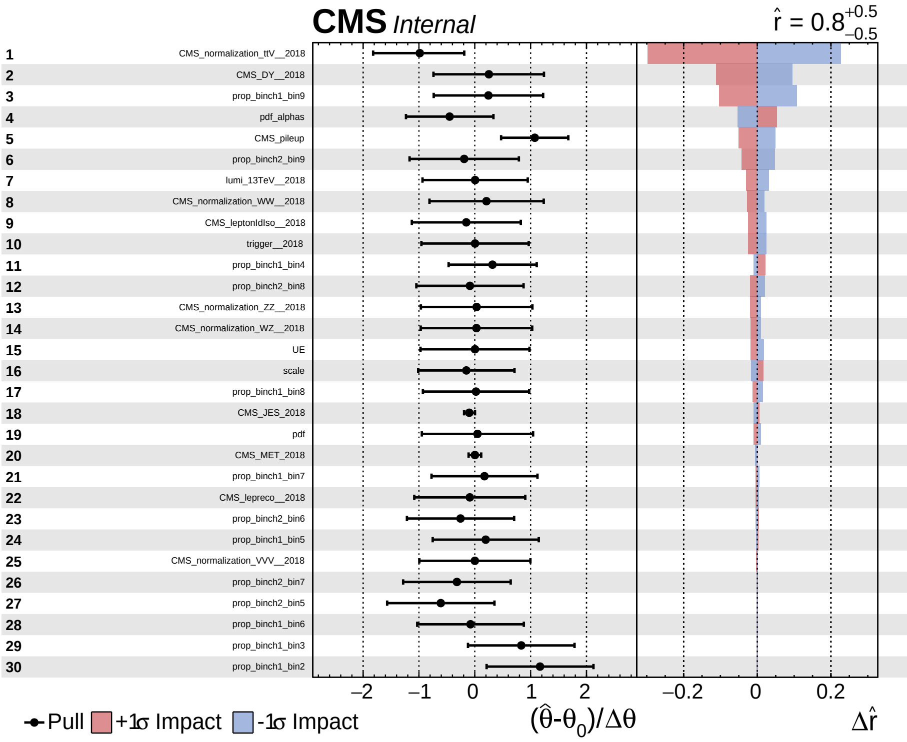


(a) 100 GeV

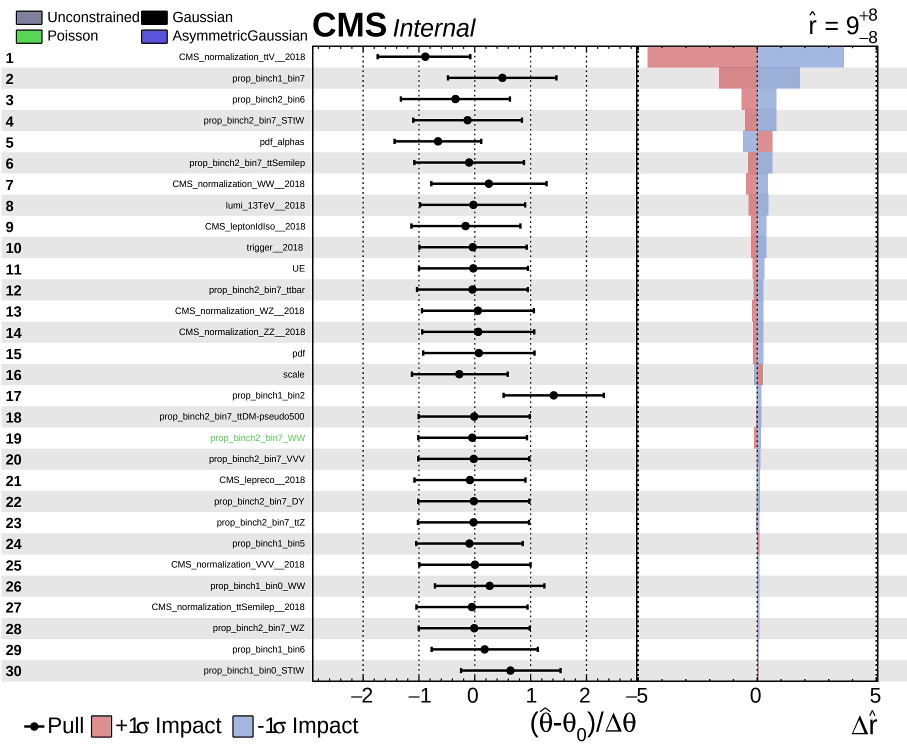


(b) 500 GeV

Figure E.5: 2018 impacts and pulls obtained for the most important systematics of the analysis, considering the 100 and 500 GeV scalar mediators.

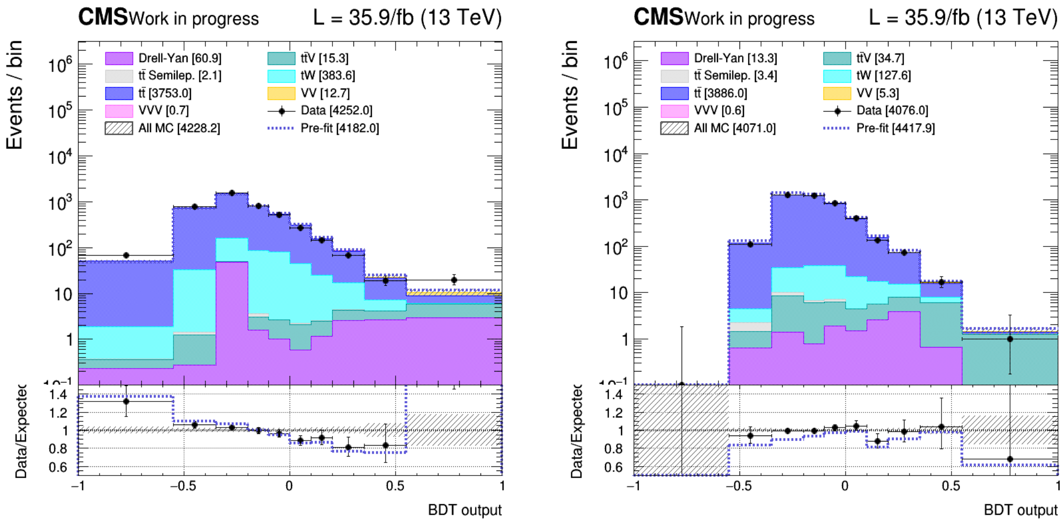


(a) 100 GeV

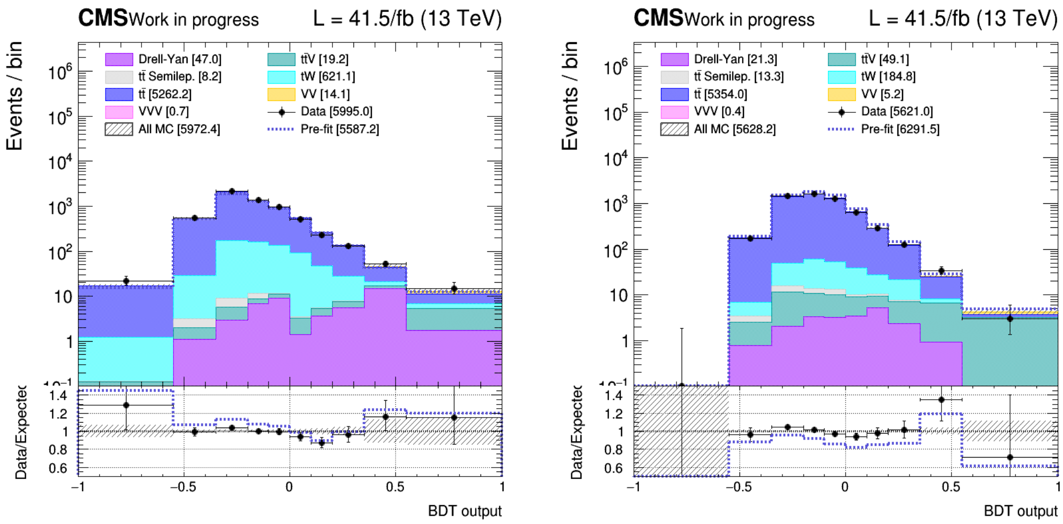


(b) 500 GeV

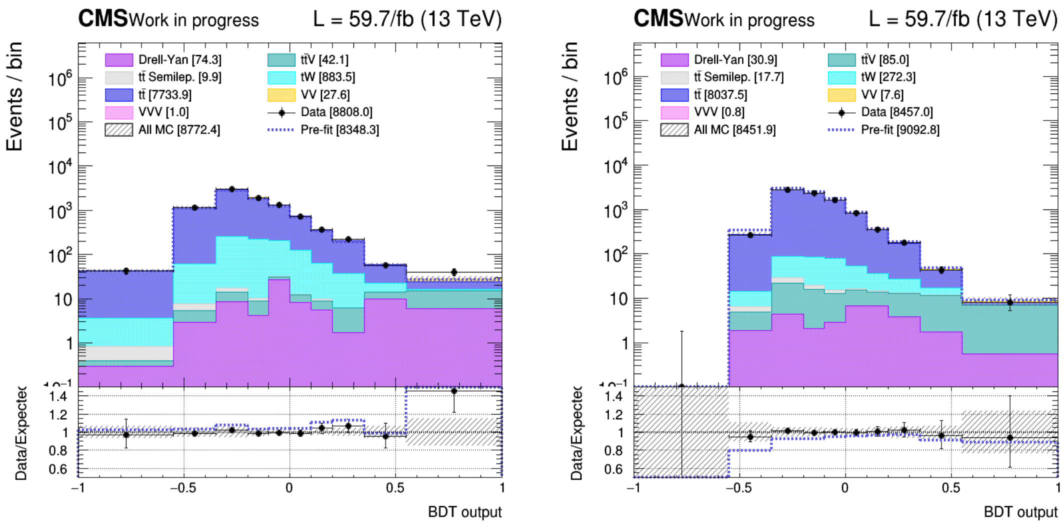
Figure E.6: 2018 impacts and pulls obtained for the most important systematics of the analysis, considering the 100 and 500 GeV pseudoscalar mediators.



(a) 2016 distributions

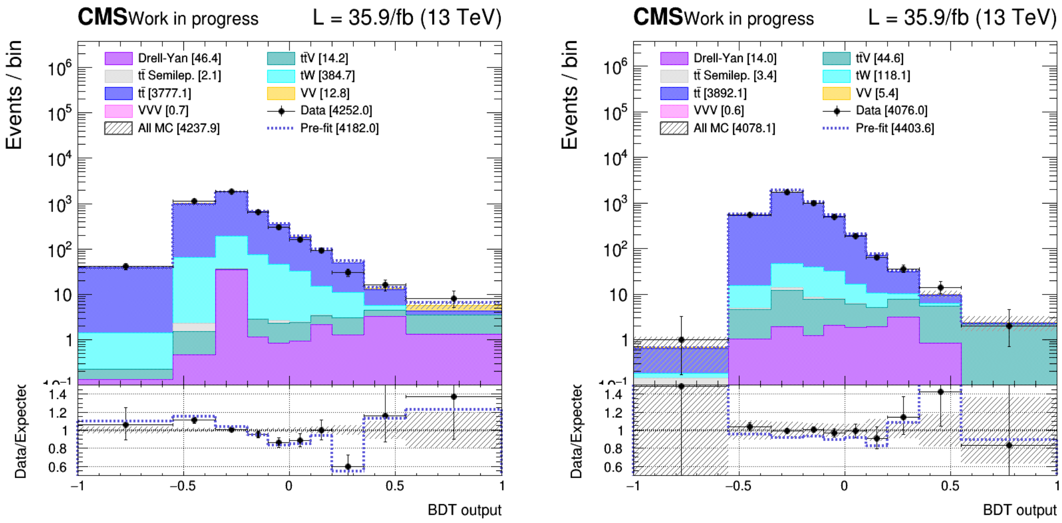


(b) 2017 distributions

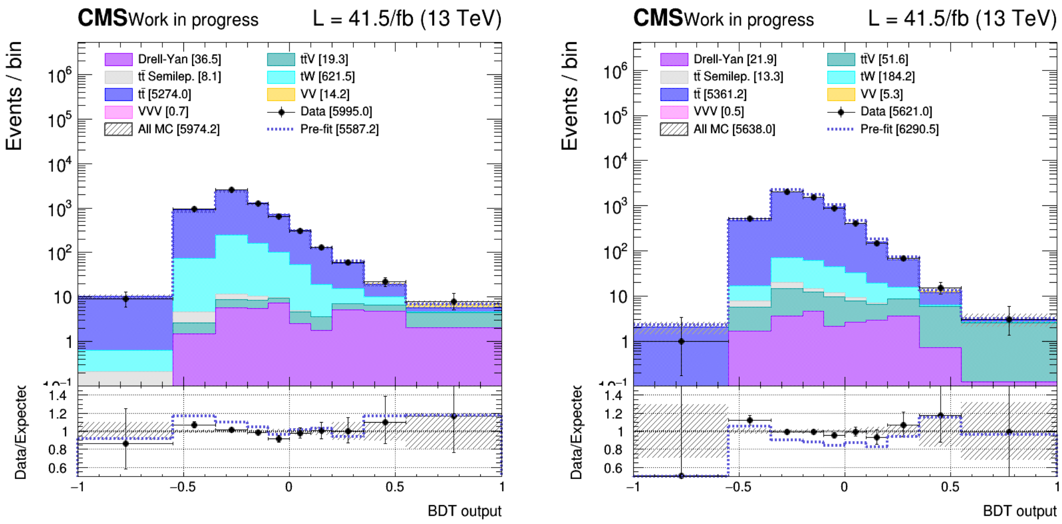


(c) 2018 distributions

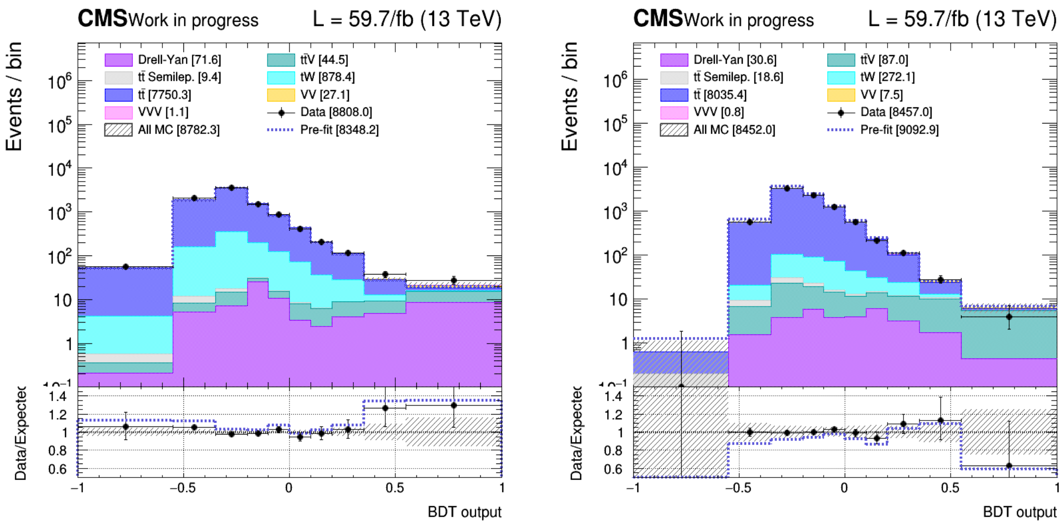
Figure E.7: Post-fits distributions obtained in the scalar 100 GeV $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.



(a) 2016 distributions

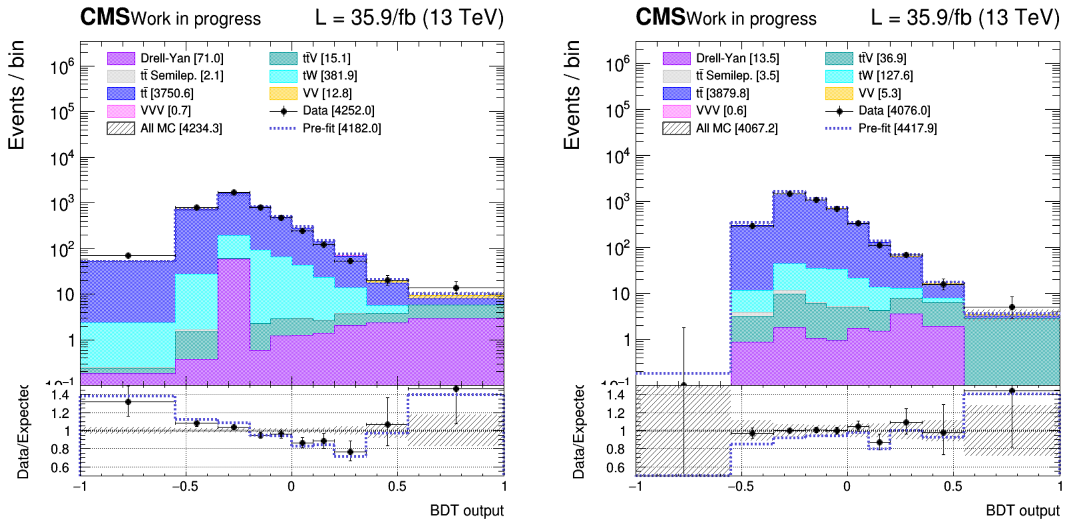


(b) 2017 distributions

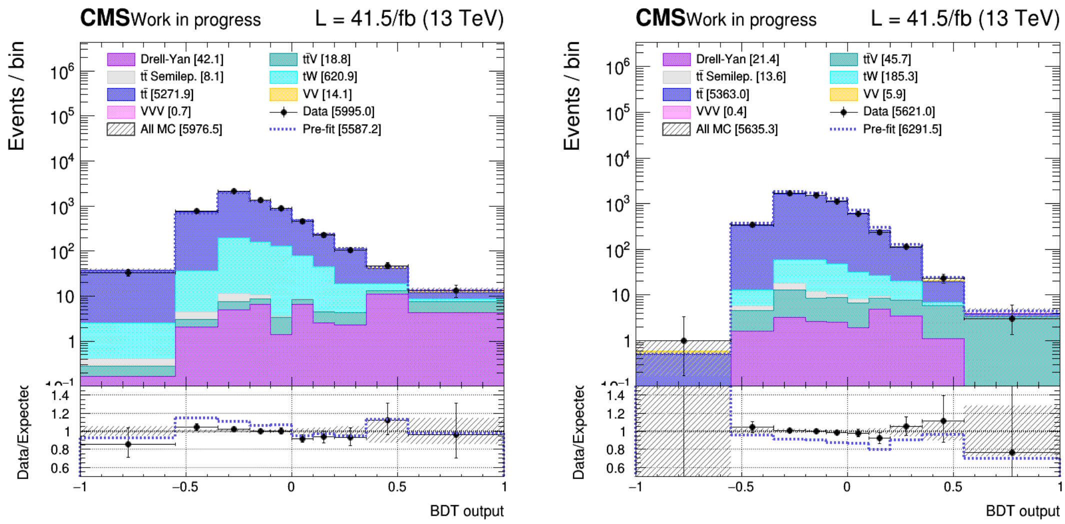


(c) 2018 distributions

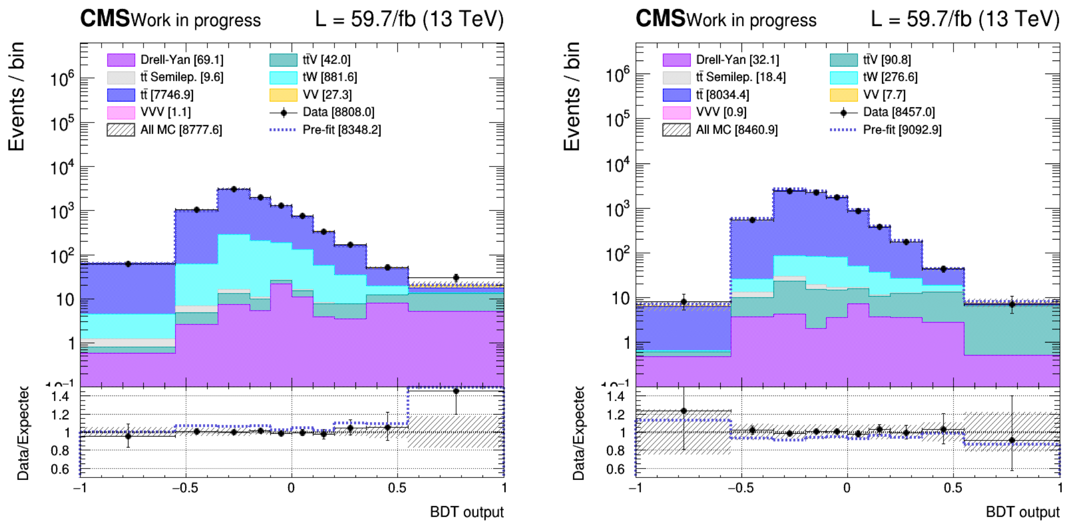
Figure E.8: Post-fits distributions obtained in the scalar 500 GeV $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.



(a) 2016 distributions

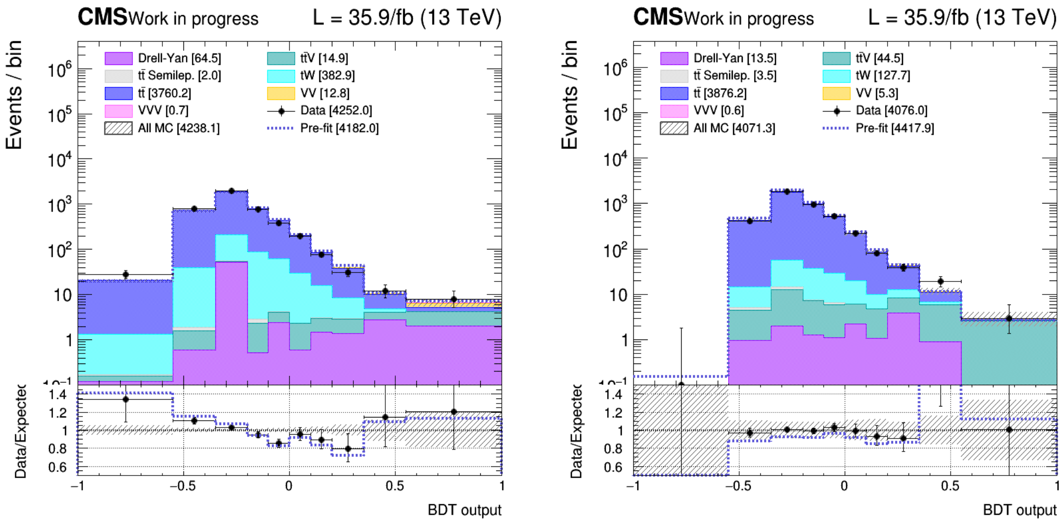


(b) 2017 distributions

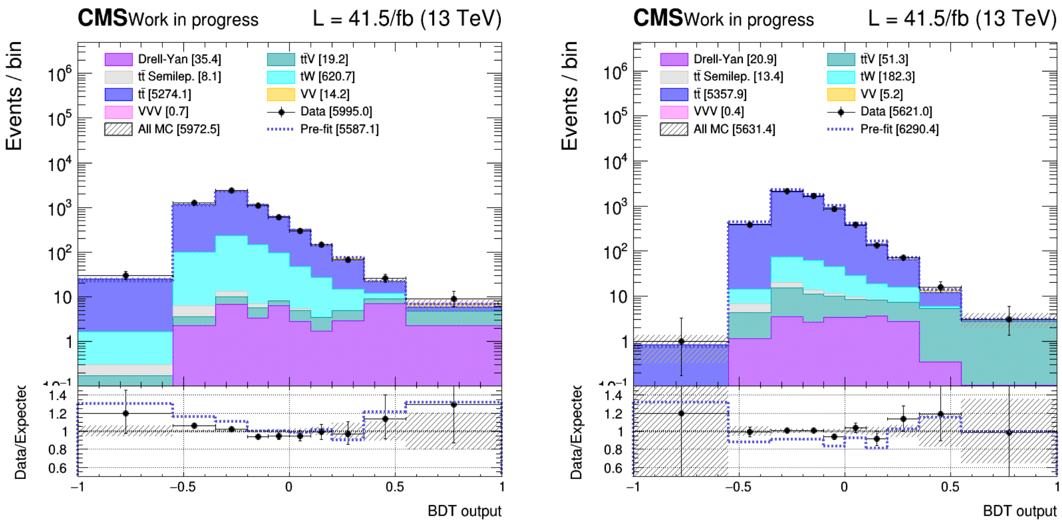


(c) 2018 distributions

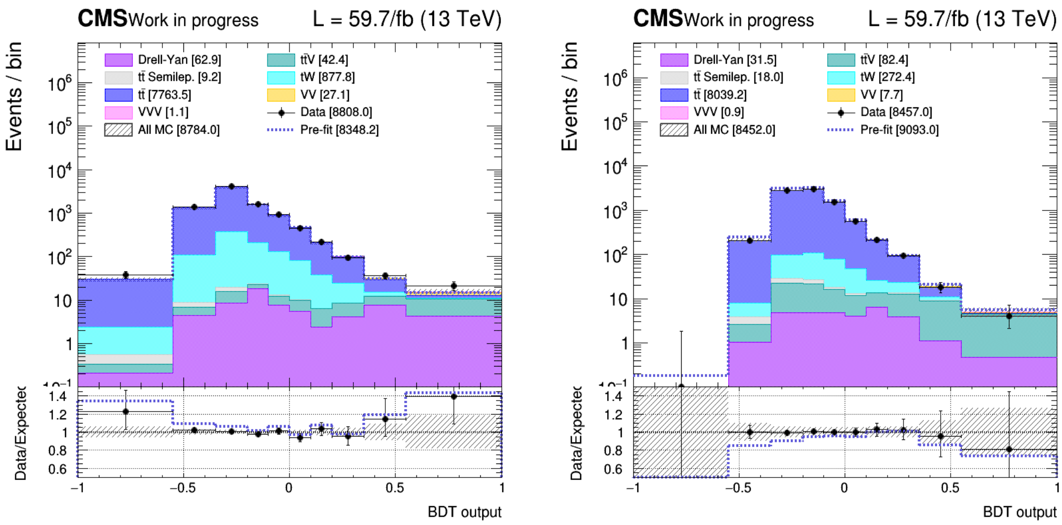
Figure E.9: Post-fits distributions obtained in the pseudoscalar 100 GeV $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.



(a) 2016 distributions



(b) 2017 distributions



(c) 2018 distributions

Figure E.10: Post-fits distributions obtained in the pseudoscalar 500 GeV $t/\bar{t}$+DM (on the left) and $t\bar{t}$+DM (on the right) signal regions.

List of figures

List of tables

Bibliography

[1] G. Altarelli, ”The Standard Model of Particle Physics”, CERN-PH-TH/2005-206, 2005

[2] F. Englert and R. Brout, ”Broken symmetry and the mass of gauge vector mesons”, Phys. Rev. Lett. 13, pp. 321-323, 1964

[3] P. W. Higgs, ”Broken symmetries and the masses of gauge bosons”, Phys. Rev. Lett. 13, pp. 508-509, 1964

[4] 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]

[5] 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]

[6] CMS Collaboration, ”The CMS Experiment at the CERN LHC”, JINST 3 S08004, 2008

[7] ATLAS Collaboration, ”The ATLAS Experiment at the CERN Large Hadron Collider”, JINST 3 S08003, 2008

[8] 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

[9] 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

[10] 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]

[11] J.B. Muñoz, C. Dvorkin and A. Loeb, ”21-cm Fluctuations from Charged Dark Matter”, Phys. Rev. Lett. 121, 121301 (2018) [arXiv: 1804.01092]

[12] A. Natarajan, ”A closer look at CMB constraints on WIMP dark matter”, Phys. Rev. D85, 2012 [arXiv:1201.3939]

[13] 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]

[14] 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]

[15] ATLAS Collaboration, ”Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, ATLAS-CONF-2020-046, July 2020

[16] 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]

[17] 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]

[18] S. Manzoni, ”The Standard Model and the Higgs Boson”, Physics with Photons Using the ATLAS Run 2 Data, Springer Theses, 2019

[19] 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]

[20] 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

[21] Y. Shadmi, ”Introduction to Supersymmetry”, CERN Yellow Report CERN 2016-003, pp. 95-123

[22] H. Poincare, ”The Milky Way and the Theory of Gases”, Popular Astronomy, vol. 14, pp.475-488, 1906

[23] F. Zwicky, ”Die Rotverschiebung von extragalaktischen Nebeln”, Helvetica Physica Acta , vol. 6, pp. 110-127, 1933

[24] S. Van den Bergh,Phys Rev D ”The early history of dark matter”, Dominion Astrophysical Observatory, 1999

[25] V.C. Rubin, W.K. Ford, ”Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions”, Astrophysical Journal 159, p. 379, 1970

[26] A. A. Penzias, R.W. Wilson, ”A Measurement of Excess Antenna Temperature at 4080 Mc/s”, Astrophysical Journal 142, pp. 419-421, 1965

[27] D.J. Fixsen, ”The temperature of the cosmic microwave background”, Astrophysical Journal, 2009

[28] Planck Collaboration, ”Planck 2018 results. I. Overview and the cosmological legacy of Planck”, 2018 [arXiv: 1807.06205]

[29] R. Tojeiro, ”Understanding the Cosmic Microwave Background Temperature Power Spectrum”, 2006

[30] Planck Collaboration, ”Planck 2018 results. VI. Cosmological parameters”, 2018 [arXiv: 1807.06209]

[31] "Astrophysical Constants and Parameters", 2019

[32] D. Clowe et all., "A Direct Empirical Proof of the Existence of Dark Matter", Astrophysical Journal Letters 648, 2006

[33] L. Heurtier, H. Partouche, "Spontaneous Freeze Out of Dark Matter From an Early Thermal Phase Transition", CPHT-RR065.112019 [arXiv: 1912.02828]

[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", PhD Royal Holloway, U. of London, 2017

[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. Morgante, "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] S.K. Vempati, "Introduction to MSSM", 2012 [arXiv: 1201.0334]

[54] B. Penning, "The Pursuit of Dark Matter at Colliders - An Overview", 2017 [arXiv: 1712.01391]

[55] M. Schumann, "Direct Detection of WIMP Dark Matter: Concepts and Status", J. Phys. G46 (2019) no.10, 103003 [arXiv: 1903.03026]

[56] S.C. Martin et all., "The RAVE survey: constraining the local Galactic escape speed", Mon.Not.Roy.Astron.Soc.379:755-772, 2007

[57] K. Freese, M. Lisanti, C. Savage, "Annual Modulation of Dark Matter: A Review", [arXiv: 1209.3339]

[58] T.M. Undagoitia and L. Rauch, "Dark matter direct-detection experiments", J. Phys. G43 (2016) no.1, 013001 [arXiv: 1509.08767]

[59] 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]

[60] J.M. Gaskins, "A review of indirect searches for particle dark matter", Contemporary Physics, 2016 [arXiv: 1604.00014]

[61] F.S. Queiroz, "Dark Matter Overview: Collider, Direct and Indirect Detection Searches", Max-Planck Institute of Physics

[62] LAT collaboration, "Constraints on Dark Matter Annihilation in Clusters of Galaxies with the Fermi Large Area Telescope", JCAP 05(2010)025 [arXiv: 1002.2239]

[63] A.A. Moiseev et all., "Dark Matter Search Perspectives with GAMMA-400", 2013 [arXiv: 1307.2345]

[64] L. Covi et all., "Neutrino Signals from Dark Matter Decay", JCAP 1004:017, 2010 [arXiv: 0912.3521]

[65] 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]

[66] J. Abdallah et all., "Simplified Models for Dark Matter Searches at the LHC", Phys. Dark Univ. 9-10 (2015) 8-23 [arXiv: 1506.03116]

[67] 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]

[68] D. Abercrombie et all, "Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum", Phys. Dark Univ. 26 (2019) 100371 [arXiv: 1507.00966]

[69] 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]

[70] 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]

[71] 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]

[72] 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]

[73] 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]

[74] 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]

[75] Atlas Collaboration, "Search for new phenomena in dijet events using 37 fb$^{?1}$ of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector", Phys. Rev. D 96, 052004 (2017) [arXiv: 1703.09127]

[76] CMS Collaboration, "Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles", JHEP 08 (2018) 130 [arXiv: 1806.00843]

[77] C. Munoz, "Models of Supersymmetry for Dark Matter", FTUAM 17/2, IFT-UAM/CSIC-17-005, 2017 [arXiv: 1701.05259]

[78] CMS Collaboration, "Searches for invisible decays of the Higgs boson in pp collisions at $sqrts =$ 7, 8, and 13 TeV", JHEP 02 (2017) 135 [arXiv: 1610.09218]

[79] J. Alimena et all., "Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider", 2019 [arXiv: 1903.04497]

[80] CMS Twiki, "CMS Exotica Summary plots for 13 TeV data", as seen in June 2020

[81] 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]

[82] M. Tanabashi et al., Particle Data Group, Phys. Rev. D98, 030001 (2018)

[83] G. Giacomelli and R. Giacomelli, "The LEP legacy", 2005 [arXiv:hep-ex/0503050]

[84] R. Schicker, "The ALICE detector at LHC", 2005 [arXiv:hep-ex/0509259]

[85] LHCb Collaboration, "LHCb Detector Performance", Int. J. Mod. Phys. A 30, 1530022 (2015) [arXiv: 1412.6352]

[86] J.T. Boyd, "LHC Run-2 and Future Prospects", 2020

[87] E. Gschwendtner, "AWAKE, A Particle-driven Plasma Wakefield Acceleration Experiment", CERN Yellow Report CERN 2016-001, pp.271-288 [arXiv: 1705.10573]

[88] M. Thomson, "Modern Particle Physics", Cambridge University Press, 2013

[89] G. Apollinari et all., "High Luminosity Large Hadron Collider HL-LHC", CERN Yellow Report CERN-2015-005, pp.1-19 [arXiv: 1705.08830]

[90] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004

[91] CMS Collaboration, "Precision measurement of the structure of the CMS inner tracking system using nuclear interactions", JINST 13 (2018) P10034 [arXiv: 1807.03289]

[92] M.S. Kim, "CMS reconstruction improvement for the muon tracking by the RPC chambers", 2013 JINST 8 T03001 [arXiv: 1209.2646]

[93] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", JINST 13 (2018) P06015 [arXiv: 1804.04528]

[94] CMS Collaboration, "CMS TriDAS project : Technical Design Report, Volume 1: The Trigger Systems", CERN-LHCC-2000-038, 2000

[95] CMS Collaboration, "CMS The TriDAS Project : Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger", CERN-LHCC-2002-026, 2002

[96] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS-PAS-PFT-09-001, 2009

[97] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", JINST 9 (2014) P10009 [arXiv: 1405.6569]

[98] V. Knunz, "Measurement of Quarkonium Polarization to Probe QCD at the LHC", Springer theses, 2015

[99] 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]

[100] J. Rembser, "CMS Electron and Photon Performance at 13 TeV", J. Phys. Conf. Ser. 1162 012008, 2019

[101] P.L.S. Connor, "Review of jet reconstruction algorithms", Ryan Atkin J. Phys. Conf. Ser. 645 012008, 2015

[102] Jet Energy Resolutions and Corrections Twiki, "Introduction to Jet Energy Corrections at CMS", `https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC`, 2016

[103] F. Beaudette, "The CMS Particle Flow Algorithm", 2014 [arXiv: 1401.8155]

[104] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", JINST 13 (2018) P05011 [arXiv: 1712.07158]

[105] 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]

[106] D. Bertolini et all., "Pileup Per Particle Identification", JHEP 1410 (2014) 59 [arXiv: 1407.6013]

[107] CMS Collaboration, "MET Corrections and Uncertainties for Run-II", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections`, 2020

[108] 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

[109] L. Sonnenschein, "Analytical solution of ttbar dilepton equations", Phys.Rev.D73:054015, 2016

[110] B.A. Betchart, R. Demina and A. Harel, "Analytic solutions for neutrino momenta in decay of top quarks", 2013 [arXiv: 1305.1878]

[111] A. Lantero Barreda, "Improvement of the signal-to-background discrimination in search for Dark Matter produced in association with a pair of top-antitop quarks", UNICAN repository, 2018

[112] CDF Collaboration, "Measurement of the Top Quark Mass using Template Methods on Dilepton Events in Proton-Antiproton Collisions at $\sqrt{s}$ = 1.96 TeV", Phys.Rev.D73:112006, 2006

[113] CMS Collaboration, "Search for dark matter production in association with dileptonically decaying top quarks at 13 TeV", AN2016-240

[114] M.H. Seymour and M. Marx, "Monte Carlo Event Generators", MCnet-13-05, 2013 [arXiv:1304.6677]

[115] B. Cabouat, J.R. Gaunt and K. Ostrolenk, "A Monte-Carlo Simulation of Double Parton Scattering", JHEP11(2019)061 [arXiv: 1906.04669]

[116] R. Placakyte, "Parton Distribution Functions", 2011 [arXiv:1111.5452]

[117] J. Alwall et all., "MadGraph 5 : Going Beyond", 2011 [arXiv:1106.0522]

[118] C. Oleari, "The POWHEG-BOX", Nucl.Phys.Proc.Suppl.205-206:36-41 [arXiv:1007.3893]

[119] S. Frixione et all., "The MC@NLO 4.0 Event Generator", CERN-TH/2010-216 [arXiv: 1010.0819]

[120] B. Webber, "Parton shower Monte Carlo event generators", Scholarpedia

[121] M. Bahr et all., "Herwig++ Physics and Manual", Eur.Phys.J.C58:639-707, 2008 [arXiv: 0803.0883]

[122] T. Sjostrand, "A Brief Introduction to PYTHIA 8.1" Comput.Phys.Commun.178:852-867, 2008 [arXiv: 0710.3820]

[123] A. Karneyeu et all., "MCPLOTS: a particle physics resource based on volunteer computing", European Physical Journal C 74 (2014) [arXiv: 1306.3436]

[124] V. Lefebure and S. Banerjee, "CMS Simulation Software Using Geant4", CMS-NOTE-1999-072, 1999

[125] S. Banerjee, "Validation of Geant4 Physics Models Using Collision Data from the LHC", J. Phys.: Conf. Ser. 898 042005

[126] 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

[127] U. Aisch and G. Polesello, "Searching for production of dark matter in association with top quarks at the LHC", High Energy Physics - Phenomenology, 2018 [arXiv: 1812.00694]

[128] 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

[129] C.G. Lester and B. Nachman, "Bisection-based asymmetric MT2 computation: a higher precision calculator than existing symmetric methods", JHEP03(2015)100, 1999 [arXiv: 1411.4312]

[130] K. Bloom, "CMS software and computing for LHC Run 2", ICHEP 2016 [arXiv: 1611.03215]

[131] W. Tanenbaum, "A ROOT/IO Based Software Framework for CMS", ECONFC0303241:TUKT010, 2003

[132] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS-PAS-LUM-17-001, 2017

[133] CMS Collaboration, "CMS Luminosity Measurements for the 2017 Data Taking Period", CMS-PAS-LUM-17-001, 2018

[134] CMS Collaboration, "CMS Luminosity Measurements for the 2018 Data Taking Period", CMS-PAS-LUM-17-001, 2019

[135] NNPDF Collaboration, "Parton distributions for the LHC Run II", 10.1007/JHEP04(2015)040 [arXiv: 1410.8849]

[136] **CMS!** Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", Eur. Phys. J. C 80 (2020) 4 [arXiv: 1903.12179]

[137] **CMS!** Twiki, "Reweighting recipe to emulate Level 1 ECAL prefiring", `https://twiki.cern.ch/twiki/bin/view/CMS/L1ECALPrefiringWeightRecipe`, as seen in May 2020

[138] **CMS!** Twiki, "Official Prescription for calculating corrections and uncertainties on Missing Transverse Energy (MET)", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETUncertaintyPrescription#Instructions_for_2017_data_with`, as seen in May 2021

[139] J. Hirschauer, "CMS status report", LHCC Open Session, September 2018

[140] 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]

[141] **CMS!** Twiki, "pt(top-quark) based reweighting of ttbar MC", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting`, as seen in May 2020

[142] **CMS!** Twiki, "The modeling of the top quark p_T", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting#Case_3_1_Analyses_with_SM_tt_as`, as seen in February 2021

[143] CMS Collaboration, "Measurement of the trigger efficiencies for a dilepton selection for a ttbar analysis with the full run 2 dataset", AN2019-140

[144] **CMS!** Twiki, "Electron and Photon Physics Object Offline Guide", `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEgamma`, as seen in May 2020

[145] **CMS!** Twiki, "Cut Based Electron ID for Run 2", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2`, as seen in May 2021

[146] **CMS!** Twiki, "Baseline muon selections for Run-II", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2`, as seen in May 2021

[147] **CMS!** Twiki, "Jet Identification", `https://twiki.cern.ch/twiki/bin/view/CMS/JetID`, as seen in May 2020

[148] **CMS!** Twiki, "Jet Identification", `https://twiki.cern.ch/twiki/bin/view/CMS/JetID#Recommendations_for_13_TeV_data`, as seen in May 2021

[149] **CMS!** Twiki, "Recommended Jet Energy Corrections and Uncertainties For Data and MC", `https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC`, as seen in May 2021

[150] **CMS!** Twiki, "Jet Energy Resolution", `https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Scaling_factors_and_Uncertai`, as seen in May 2021

[151] **CMS!** Twiki, "Heavy flavor identification at CMS with deep neural networks", `https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeVDPDeepCSV`, as seen in May 2020

[152] **CMS!** Twiki, "MET significance", `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMETSignificance`, as seen in August 2020

[153] K. Woodruff, "Introduction to boosted decision trees", Machine Learning Group Meeting, September 2017

[154] K. Albertsson et all, "TMVA 4: Toolkit for Multivariate Data Analysis with ROOT, users guide", CERN-OPEN-2007-007

[155] M. Jachowski, "Multivariate Analysis, TMVA, and Artificial Neural Networks", Michigan REU Final Presentations, August 2006

[156] **CMS!** Twiki, "LHC Higgs Combination Group (LHC-HCG)", `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HiggsCombination`, as seen in August 2020

[157] G. Cowan et all, "Asymptotic formulae for likelihood-based tests of new physics", Eur.Phys.J.C71:1554, 2011 [arXiv: 1007.1727]

[158] A.L. Read, "Presentation of search results: the CLs technique", J. Phys. G: Nucl. Part. Phys. 28 (2002) 2693-2704

[159] A.L. Read, "1st Workshop on Confidence Limits", CERN-2000-005, pp81-101"

[160] T. Junk, "Statistical Methods for Experimental Particle Physics", TRIUMF Summer Institute, July 2009

[161] P.K. Sinervo, "Definition and Treatment of Systematic Uncertainties in High Energy Physics and Astrophysics", PHYSTAT2003, Stanford Linear Accelerator Center, September 2003

[162] J. Buttwerworth et all., "PDF4LHC recommendations for LHC Run II", J. Phys. G: Nucl. Part. Phys. 43 023001, 2016 [arXiv: 1510.03865]

[163] CMS collaboration., "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", JHEP 07 (2018) 161 [arXiv: 1802.02613]

[164] CMS collaboration., "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", JINST 12 (2017) P02014 [arXiv: 1607.03663]

[165] CMS collaboration., "Identification of heavy-flavour jets with the CMS detector in pp collisions $\sqrt{s} = 13$ TeV", JINST 13 (2018) P05011 [arXiv: 1712.07158]

[166] CERN Twiki, "ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program", `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO`, as seen in August 2020

[167] CERN Twiki, "ATLAS-CMS recommended predictions for single-top cross sections using the Hathor v2.1 program", `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec`, as seen in August 2020

[168] CMS Collaboration, "BRIL Work Suite", `https://cms-service-lumi.web.cern.ch/cms-service-lumi/brilwsdoc.html`, as seen in June 2020