



FACULTAD DE CIENCIAS
UNIVERSIDAD DE CANTABRIA

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

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

Written by
Cédric Prieëls

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

Santander, June 2020



FACULTAD DE CIENCIAS
UNIVERSIDAD DE CANTABRIA

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

Memoria para optar al
Grado de doctor

Escrita por
Cédric Prieëls

Bajo la supervisión de
Jónatan Piedra Gómez
Pablo Martínez Ruiz del Árbol

Santander, Junio 2020

Abstract

Resumen

Acknowledgments

First of all, I would like to thank the group of high energy physics of the Instituto de Física de Cantabria (IFCA) in general for the opportunity they gave me by offering 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 the personal and professional point 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. I cannot count the number of bugs and issues you helped me solve, so thank you for keeping your door open at all times anyway, and for spending so much time correcting the talks I gave over the years.

Pablo Martínez, you have been a great co-director, this work would probably not have been possible without your help and dedication to this analysis. You also became a friend and I will never be able to thank you enough for your help, especially in this last year which has been quite challenging for me.

I would also like to thank all the people who helped me at some point: Alicia, Rocío, Pupi, and Celso, mostly for all the administrative relative issues I faced. This work would not have been possible either without the help of many people involved in this particular analysis, such as Deborah, Dominic, . Thank you to all the persons involved in the development of the Latino framework, which quickly became a great tool to work with. UNIOVI

Thank you also to my PhD coworkers, Pedro, Nicolò, Celia, Andrea and Pablo for making the students office a better place to work in. I wish you all the best in the future.

Of course, I would especially like to thank to my father Philippe, my mother 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: you are the reason why I came to Spain and started this PhD in the first place and you can be sure I don't regret this decision. You brought me more than you can imagine and I will never forget all these great moments we spent during the last 7 years, and the many more still to come.

Acronyms used

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

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

Contents

1	Introduction	1
2	The experimental setup	5
2.1	The Large Hadron Collider (LHC) accelerator	5
2.1.1	The LHC in a nutshell	6
2.1.2	LHC key parameters	7
2.2	The Compact Muon Solenoid (CMS) detector	10
2.2.1	Tracker	12
2.2.2	Electromagnetic Calorimeter (ECAL)	15
2.2.3	Hadronic Calorimeter (HCAL)	16
2.2.4	Solenoid	18
2.2.5	Muon system	19
2.2.6	Trigger system	22
2.2.7	Data acquisition system	24
3	Objects reconstruction	25
3.1	Particle Flow (PF) algorithm	25
3.2	Primary Vertex (PV) definition	26
3.3	Leptons reconstruction	27
3.3.1	Muons	28
3.3.2	Electrons	30
3.4	Jets reconstruction	32
3.4.1	B-tagging	34
3.5	Missing Transverse Energy (MET)	35
3.6	Top reconstruction	36
3.6.1	Top reconstruction in practice	38
Appendices		39
A	Samples used	41
A.1	Data samples	41

A.2	Signal samples	41
A.3	Backgrounds samples	41
B	Neural network optimization	49
	Bibliography	53

Chapter 1

Introduction

The Standard Model (SM) of particle physics [1] is nowadays the most accepted mathematical model used to describe the elementary particles and three of the 4 fundamental forces of nature (electromagnetic, weak and strong interactions, while the gravitational interaction is out of reach of this model). This model is quite simple in concept, but has been able to describe most of the phenomena observed in nature so far with an incredible level of precision, and has made a lot of predictions that have now been proven to be true, such as the postulate of the Higgs mechanism [4, 5] followed by the discovery of the Higgs boson itself [6, 7] by the CMS [2] and A Toroidal LHC ApparatuS (ATLAS) [3] experiments analyzing the proton-proton collisions produced by the Large Hadron Collider (LHC) at a center of mass energy $\sqrt{s} = 13$ TeV, announced at the European Council for Nuclear Research (CERN) on the 4th of July 2012.

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

In this context, the first serious Dark Matter (DM) hypothesis was introduced in the 1970s because of gravitational anomalies observed by several astrophysicists, as a way to explain the apparent non-luminous missing mass in the Universe [8]. 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 [9], 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 [10]) and the anisotropies observed in the Cosmic Microwave Background (CMB) [11] are other evidences for the existence of DM, as explained in Section ??.

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) [12]. Understanding the nature and 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 Weakly Interactive Massive Particles (WIMPs), 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 the hint of 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 Beyond the Standard Model (BSM) physics. These searches will be summarized in Section ??.

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

In the context of this work, DM is searched for in association with one or two top quarks which play the role of the SM particle allowing us to trigger the event. The top quark, the 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 [13]. 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 dileptonic 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, as will be explained in Section ??.

The LHC has now been running for 10 years, and several similar searches have already been carried out and published in the past by the CMS and ATLAS collaborations, at different center of mass energies. First of all, at 8 TeV, several searches for a pair of top quarks were published by the CMS (in association with DM in the semileptonic [14] and dileptonic [15] final states) and ATLAS collaborations [16]. Then, at 13 TeV, the ATLAS collaboration published on one hand several studies, considering different final states and different luminosities (13.3 fb^{-1} and 36.1 fb^{-1}) [17, 18, 19]. On the other hand, the CMS collaboration published a few extremely important papers for this study [21, 22]. For the first time in 2019, the results obtained by the $t/\bar{t}+\text{DM}$ and $t\bar{t}+\text{DM}$ analyses have also been combined and published using the data collected during the year 2016 [23]. Our main objective is now to repeat and improve this analysis while considering the full Run II dataset, globally improving the analysis strategy and including the dileptonic final state for the first time in this combination.

After a general introduction about DM in Chapter ??, the experimental setup will be detailed in Chapter 2. 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 has been collected during the years 2016, 2017 and 2018 and corresponds to an integrated luminosity of $\sim 137 \text{ fb}^{-1}$, collected during the Run II of operation of the LHC and at a center of mass energy $\sqrt{s} = 13 \text{ TeV}$. A particular care will be given to the explanation of the Particle Flow (PF) algorithm, used to reconstruct the different objects of the analysis and that will be defined in Chapter 3, while the estimation of the different backgrounds and the selection of interesting events will be detailed throughout Chapters ?? and ??.

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

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

Chapter 2

The experimental setup

The data analyzed throughout this work is the result of the proton-proton collisions at a center of mass energy of 13 TeV provided by the Large Hadron Collider (LHC). It is therefore extremely important to describe this particle accelerator and to detail its main characteristics in order to understand better the origin of our data, as will be done in Section 2.1.

This kind of collisions at high energy typically produces hundreds of new particles that we then need to identify, and whose exact properties, such as their momentum or electric charge, need to be measured. This is done by using the Compact Muon Solenoid (CMS) detector, an incredibly sophisticated detector, result of the work of thousands of people and made of several different layers, which will be described in Section 2.2.

2.1 The LHC accelerator

The Large Hadron Collider (LHC) is a superconducting particle accelerator able to accelerate protons and lead ions up to velocities close to the speed of light (0.999999990 c). 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 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 [4, 5] but never observed in any experiment. The discovery of the Higgs boson, announced at CERN on the 4th of July 2012 [6, 7], was then an incredible 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 kind 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 the search for eventual DM production as the one performed in this work.

Let's now start with a global description of the general design of this particle accelerator in Section 2.1.1 before describing in more details the key parameters allowing us to evaluate its actual performances in Section 2.1.2.

2.1.1 The LHC in a nutshell

The LHC is an underground particle accelerator built in the same 27 kilometers tunnel where the Large Electron Positron collider (LEP) was previously used. This machine is accelerating two beams made out of 10^{10} protons or lead ions in each direction and is mostly made out of more than 4000 superconducting magnets, in majority dipoles and quadrupoles, allowing respectively to curve the beam to maintain a nominal circular trajectory and to focus it by compensating its natural dispersion due to the repulsion of the protons making up these beams. A dedicated small section of the accelerator is then made out of 16 radio-frequency cavities synchronized in such a way that these protons always face a negative electric charge, which is used as the driving process of the actual acceleration of such particles.

Once the nominal center of mass energy $\sqrt{s} = 13$ TeV is reached (this concept will be described in Subsection 2.1.2), the protons are then smashed together in four different places on the LHC, where the four detectors (ATLAS, CMS, A Large Ion Collider Experiment (ALICE) and LHCb) have been placed in order to study the collisions. Both ATLAS and CMS are general detectors able to study exotic processes such as the production of DM and to make precision measurements on SM physics as well (the decision to build two separate detectors was made in order to introduce some redundancy and to check the results). ALICE on the other hand is mostly dedicated to the study of heavy ions collisions that happen $\sim 10\%$ of the time in order to study in particular a specific state of matter, called quark-gluon plasma [86]. Finally, LHCb has been designed to study in particular the CP violation phenomena, which could be the sign of some new physics [87].

It is important to note at this point that the LHC is not a standalone accelerator in the sense that protons enter the LHC with a velocity already close to the speed of light. In order to reach such input energies (~ 450 GeV), previous smaller accelerators of CERN are still used today. A chain of accelerators is then formed: first of all, the protons, extracted from a bottle of ordinary hydrogen, are injected into the LINAC 2, a linear accelerator, before being transferred to the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS) and finally the LHC itself (all this chain of acceleration can be found in Figure 2.1).

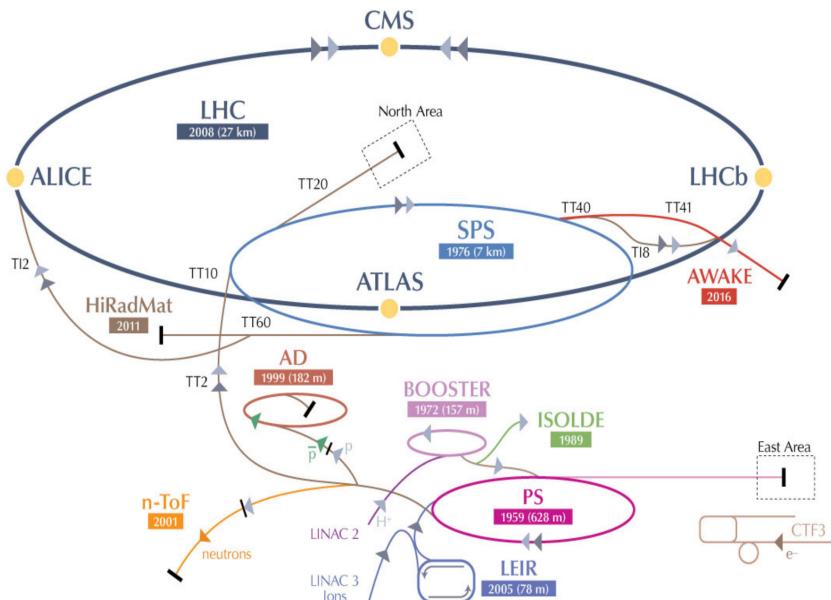


Figure 2.1: LHC injection chain and experiments performed at CERN [89].

Parameter	Run I	Run II	Run III	Design
Energy [TeV]	7 → 8	13	14	14
Bunch spacing [ns]	50	25	25	25
Intensity [10^{11} 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 → 25	30 → 25	55
Crossing angle [μrad]	-	300 → 260	300 → 260	285
Peak luminosity [$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$]	0.8	2.0	2.0	1.0
Peak PU	45	60	55	25

Table 2.1: Expected and observed main parameters of pp operation of the LHC across the different eras of operation [88].

During this phase of acceleration, the beam is separated in 2808 bunches of protons nominally separated by 25 ns (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 Pile up (PU), has to be taken into account as well and will be described in the next section.

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

The data analyzed in this work corresponds to the second phase of the Run II of operation of the LHC, from 2016 to 2018, while the Run III is now expected to start in the Spring of 2021. The summary of the main parameters of operation of the LHC across the different Runs of operations can be found in Table 2.1.

2.1.2 LHC key parameters

Center of mass energy

The center of mass energy is defined as a Lorentz invariant quantity under longitudinal boosts resulting of the collisions of two protons (defined as E_1, \vec{p}_1, m_1 and E_2, \vec{p}_2, m_2) with a θ angle, as developed in Equation 2.1. It is a key variable of the LHC since the phase space of particles that

can be explored directly depends on this value.

$$\sqrt{s} = \sqrt{(m_1)^2 + (m_2)^2 + 2(E_1 E_2 - 2|\vec{p}_1| |\vec{p}_2| \cos(\theta))} = 13 \text{ TeV} \quad (2.1)$$

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

Luminosity

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

Mathematically, we can first of all define in Equation 2.2 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 can be extracted easily the number of expected interactions N in a certain amount of time T as well.

$$\begin{cases} R = \mathcal{L} \cdot \sigma \\ N(T) = \sigma \int_0^T \mathcal{L}(t) dt = \sigma L \end{cases} \quad (2.2)$$

This instantaneous luminosity we just introduced \mathcal{L} can be defined using Equation 2.3, assuming that the beams have a Gaussian profile, while the integrated luminosity L can be defined by simply integrating the instantaneous luminosity over time $L = \int \mathcal{L} dt$ [90].

$$\mathcal{L} = \frac{\gamma f_{\text{rev}} k_B N_p^2}{4\pi \epsilon_n \beta^*} R, \text{ where } R = \frac{1}{\sqrt{1 + \frac{\theta_C \sigma_z}{2\sigma}}} \quad (2.3)$$

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

- 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 25 ns bunch spacing)
- N_p is the number of protons per bunch ($1.15 \cdot 10^{11}$ protons)
- ϵ_n is the transverse normalized emittance ($3.75 \mu\text{m}$)
- β^* is the betatron function at the interaction point (0.55 m)

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

As previously stated, increasing the luminosity of the LHC is always something interesting in order to produce processes having a low production cross-section, and we can then see that many different parameters can be tweaked in order to achieve the highest possible instantaneous luminosity, such as the number of protons per bunch, the number of bunches per beam or the beam crossing angle at the interaction point. New 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 ~ 10 (HL-LHC project [91]).

The total integrated luminosity taken by the LHC during its different years of operation has been summarized in Figure 2.2. The final datasets available and analyzed in this work have an integrated luminosity of 35.9 fb^{-1} (2016), 41.5 fb^{-1} and 59.7 fb^{-1} (2018), resulting in a total dataset of 137 fb^{-1} recorded during the Run II of operation. This means that we roughly expect to produce around 137 events of any process whose cross section of production would be equal to 1 fb .

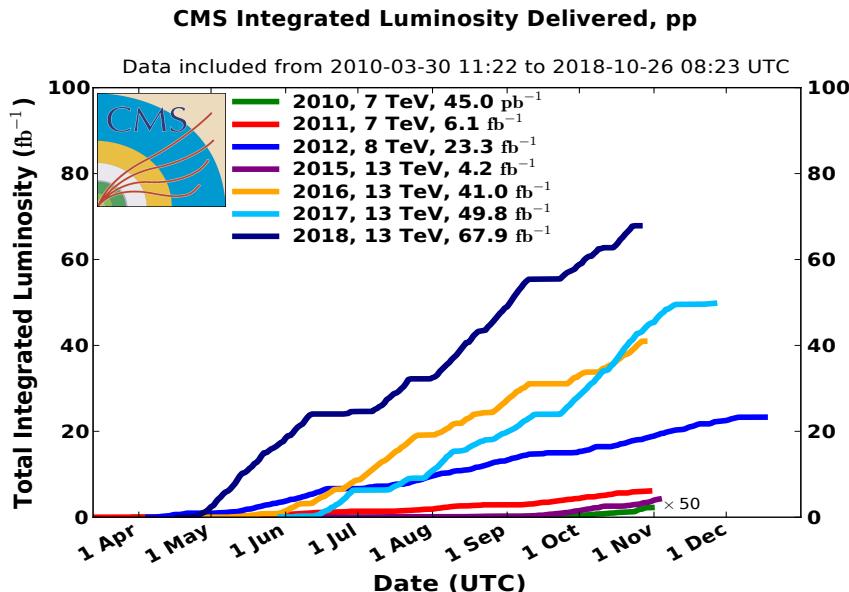


Figure 2.2: Integrated luminosity collected by CMS over its different years of operation so far.

Pile up (PU)

The last key parameter of the LHC discussed here is the PU. Usually, because of the high density of protons within the beams, a bunch crossing in an experiment produces around 30-35 collisions per proton-proton collision, as seen in Figure 2.3, defining the PV as the most interesting one, while the other vertices are usually referred to as the PU. The tracker of CMS, which will be introduced in Section 2.2.1, therefore needs to be able to reconstruct all these events in order to define the PV of the interaction.

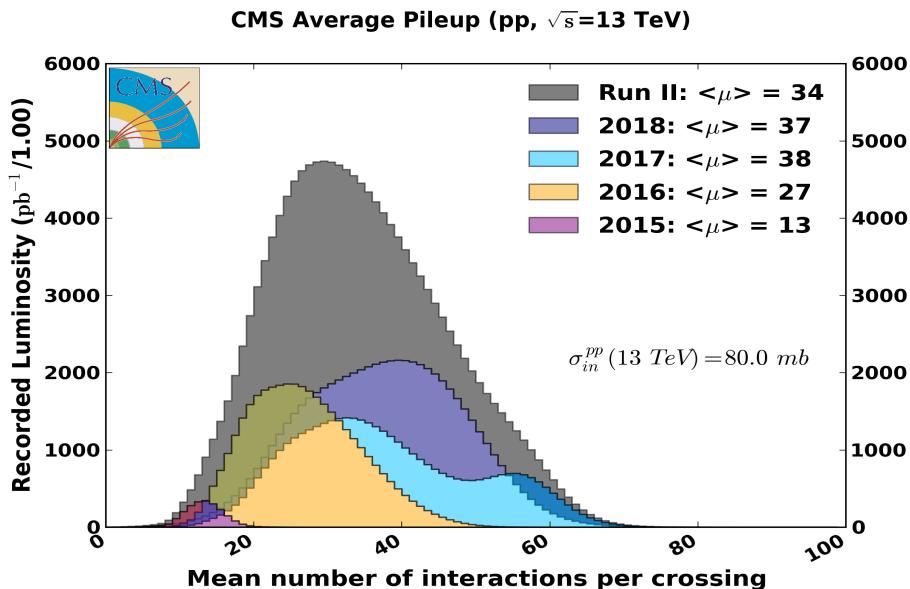


Figure 2.3: Mean PU distribution and luminosity recorded by CMS over the different years of operation of the LHC.

2.2 The CMS detector

The Compact Muon Solenoid (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 measurement of most of the SM processes and to try and discover BSM physics, such as the possible existence of DM.

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

Main goals

The CMS detector has been built in order to meet the goals of the LHC physics program, which have been summarized in [92] as:

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

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

CMS subdetectors

As shown in Figure 2.4, CMS is made out of different layers corresponding to different subdetectors, each able to provide different kinds of information about the particles created by each collision [92]. These subdetectors will be described in detail in the following sections, but they have all been designed in order to make the reconstruction of the different events as efficient, fast and precise as possible and need to be extremely resistant to the radiation produced.

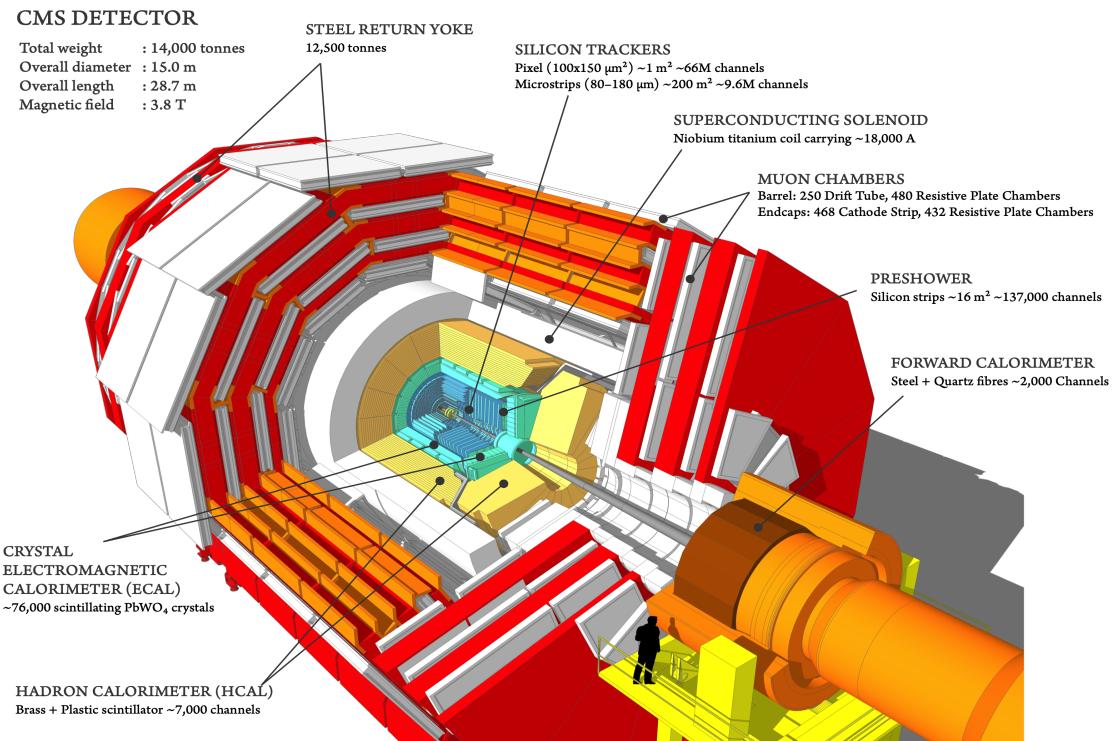


Figure 2.4: Schematic representation of the CMS detector, along with all its subdetectors and main characteristics.

The inner part of the CMS detector is the so-called tracker, a device made out of silicon pixels and strips, described in Section 2.2.1 and responsible for the precise reconstruction of all the charged particles coming from the different interaction vertices. A bit further outside can be found the Electromagnetic Calorimeter (ECAL), made out of thousands of crystals as described in Section 2.2.2 and used to precisely measure the energy of particles able to interact electromagnetically by producing an electromagnetic shower that can be detected. Then comes the Hadronic Calorimeter (HCAL), described in Section 2.2.3 and whose main purpose is to identify and measure the main properties of the hadrons produced in the collisions by producing hadronic showers.

The CMS name partly comes from its central piece, a huge superconducting solenoid described in Section 2.2.4 and able to produce a 3.8 T magnetic field in the detector, with a magnetic flux

density increased even more by the addition of the steel return yoke layers (red parts in Figure 2.4). This magnetic field is essential in the sense that it is able to deflect the charged particles which have been produced via the Lorentz interaction, therefore giving us a perfect way to measure their charge and energy from the measurement of the curvature of the induced binding.

Finally, on the outside of the detector can be found the complete muon system and particularly performing in CMS. This subdetector is currently made out of three main sub-systems (the Drift tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs)), as explained in Section 2.2.5, and is responsible for the identification and measurement of the momentum of the muons produced by the collisions.

Coordinates system

Before starting with the description of all these subdetectors, it is important to detail the coordinates system typically used within the CMS collaboration. As a convention, it has been decided to use a right-handed Cartesian coordinate system with the origin defined as the Primary Vertex (PV), and with an x-axis pointing towards the center of the ring, an y-axis pointing upwards and a z-axis pointing towards the Jura mountains (along the counterclockwise beam direction), as represented in Figure 2.5.

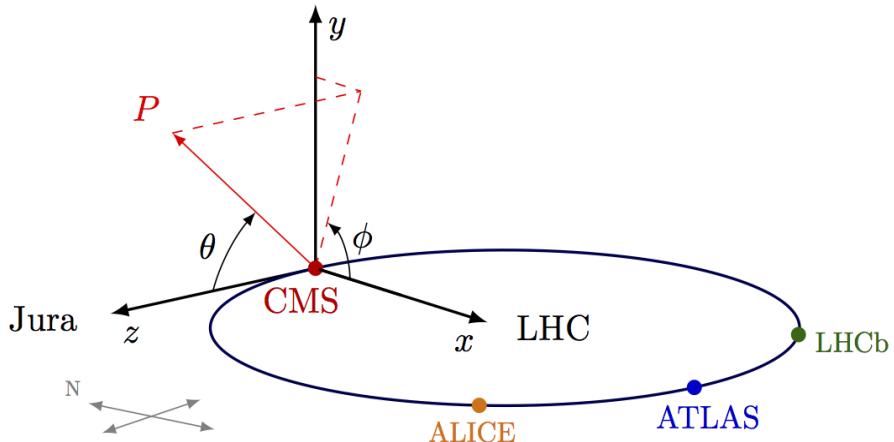


Figure 2.5: Schematic representation of the CMS coordinate system used by convention.

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 2.4, a Lorentz invariant quantity under longitudinal boosts quite often used in the different analyses since the multiplicity of high energy particles is roughly constant in η .

$$\eta = -\log \left(\tan \left(\frac{\theta}{2} \right) \right) \quad (2.4)$$

2.2.1 Tracker

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

them and measure their individual momentum.

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

This device is then made out of silicon pixels and strips mainly because of the granularity, reading velocity and radiation hardness offered by such material. It has been set up on several different layers disposed in such a way to make the detector as hermetic as possible, as shown in Figure 2.6. 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 reconstruction algorithms that will be detailed in Chapter 3.

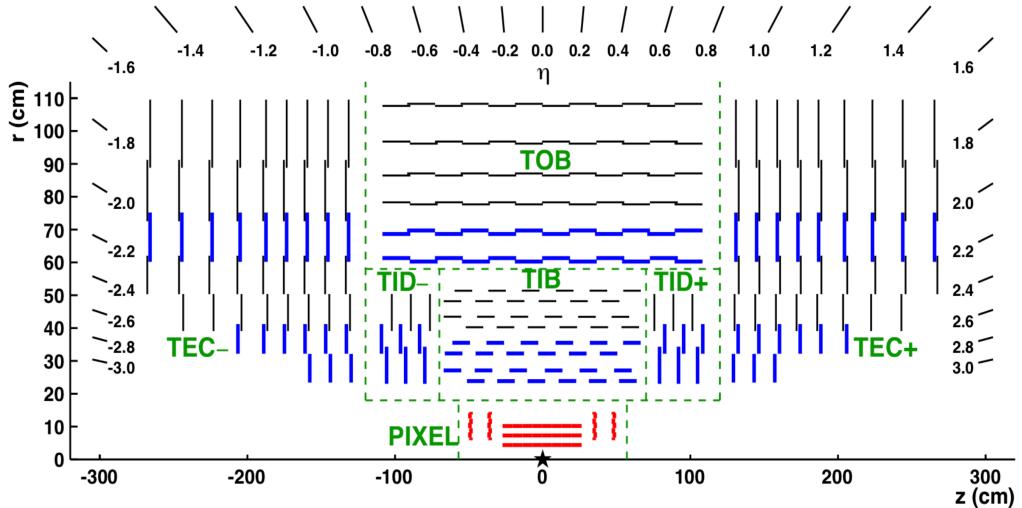


Figure 2.6: Schematic representation of the CMS tracker, for different pseudorapidity values and along the z -axis [93].

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, more than 60 millions pixels make up the 1440 modules of this detector, covering an area corresponding to ~ 1 m².

- The **silicon strip detector** on the other hand is composed of three different sub-systems, as seen in Figure 2.6 and covers a total area of $\sim 200 \text{ m}^2$. First of all, the Tracker Inner Barrel and Disks (TIB/TBD) add an additional 4 barrel layers and 3 endcap disks to the tracker system up to a distance $r = 55 \text{ cm}$ to the PV of the interaction, using silicon micro-strip sensors parallel in the barrel and perpendicular to the beam axis in both endcaps. Then, the Tracker Outer Barrel (TOB) surrounds this first layer; having an outer radius of 166 cm and going up to 118 cm along the z-axis, it adds 6 layers to the inner tracking system. Finally, the Tracker EndCaps (TECs) are made out of 9 disks and complete the measurement of particles emitted along the z-axis and having a high pseudorapidity η .

The CMS tracker is extremely performing: one can see first of all in Figure 2.7 that for high momentum tracks of 100 GeV, the resolution is 1-2% up to $|\eta| < 1.6$.

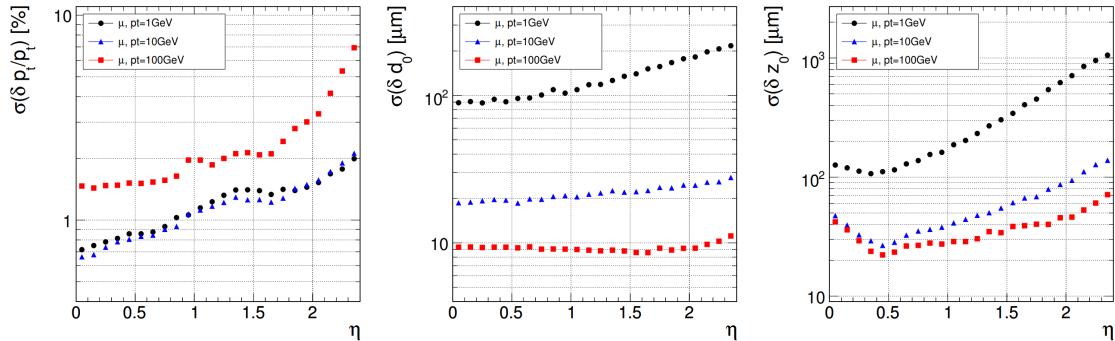


Figure 2.7: Expected resolution of muons transverse momentum (left), transverse impact parameter (middle) and longitudinal impact parameter (right), as a function of pseudorapidity and muon momentum (1, 10 and 100 GeV) [92].

Additionally, Figure 2.8 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.

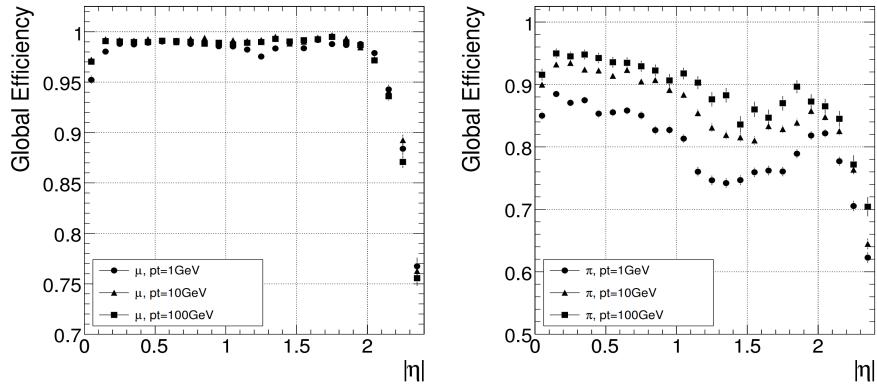


Figure 2.8: Tracker reconstruction efficiency of muons (on the left) and pions (on the right) calculated in simulation for different pseudorapidities and particle momenta (1, 10 and 100 GeV) [92].

2.2.2 Electromagnetic Calorimeter (ECAL)

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

The ECAL can also be divided into several sections: first of all, at pseudorapidities $|\eta| < 1.479$ is found the barrel part of the ECAL (EB), made out of 61 200 lead tungstate (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 2.9, and the preshower completes the ECAL.

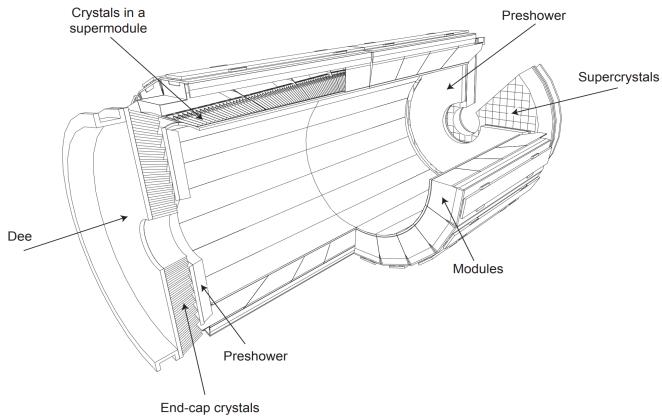


Figure 2.9: Schematic representation of the sub-systems of the CMS ECAL [92].

Its principle of action is simple and based on electromagnetic showers: when a particle such as an electron or a photon enters the ECAL, it will start to interact in different ways, depending on its nature. Photons will mainly produce pairs of electrons and anti-electrons, while the electrons themselves tend to emit additional photons by bremsstrahlung effect. This results in a chain reaction during which the incident particle will give most of its energy to the detector itself, energy measurable using photodetectors and photomultipliers. This effect is known as electromagnetic shower, is represented in Figure 2.10 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.

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 $2.2 \times 2.2 \times 23$ cm in the barrel and $3 \times 3 \times 22$ cm in the endcaps, cover a solid angle equal to $(\Delta\eta, \Delta\phi) = (0.0174, 0.0174)$, and have therefore a length corresponding to around 26 radiation lengths, more than enough to stop even the most energetic particles. Since the light output of such crystals is quite low (only around 4.5 photoelectrons per MeV of energy), they have to be connected to both avalanche photodiodes and vacuum phototriodes in order to multiply the signal measured. Finally, they have been mounted using a specific installation, slightly tilted with respect to both ϕ and η in order to

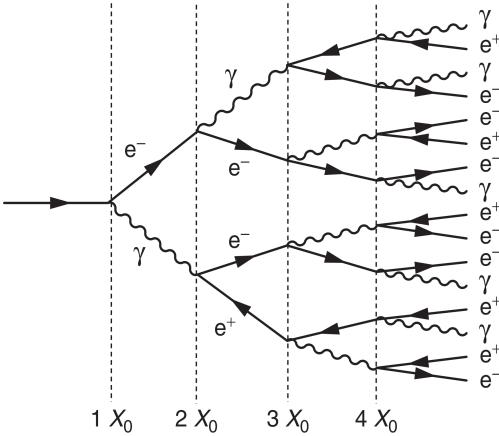


Figure 2.10: Schematic representation of a typical electromagnetic shower and the radiation length X_0 concept [92].

remove any possible gap between two adjacent crystals.

The typical energy resolution of the ECAL installed at CMS is given by Equation 2.5 [92], 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.

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12\%}{E}\right)^2 + (0.30\%)^2} \sim \frac{3 - 10\%}{\sqrt{E/\text{GeV}}} \quad (2.5)$$

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

2.2.3 Hadronic Calorimeter (HCAL)

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

In this case as well, showers of particles due to a chain reaction are expected since the primary hadronic interaction will produce several additional hadrons, themselves interacting even more with the medium while losing energy. This kind of hadronic showers is characterized by the 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 [90]. In this case, the energy resolution can therefore be expressed using Equation 2.6.

$$\frac{\sigma_E}{E} > \frac{50\%}{\sqrt{E/\text{GeV}}} \quad (2.6)$$

In CMS, the HCAL, represented in Figure 2.11, is also divided into a barrel (HB), radially constrained between radii values of 1.77 meters (outer radius of the ECAL) and 2.95 meters (inner radius of the solenoid), and two symmetrical forward regions (HF) extending the 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 Hadron Outer (HO), which has to be put outside of the solenoid in order to increase the amount of shower absorber material of the HCAL and therefore the effective nuclear radiation length λ .

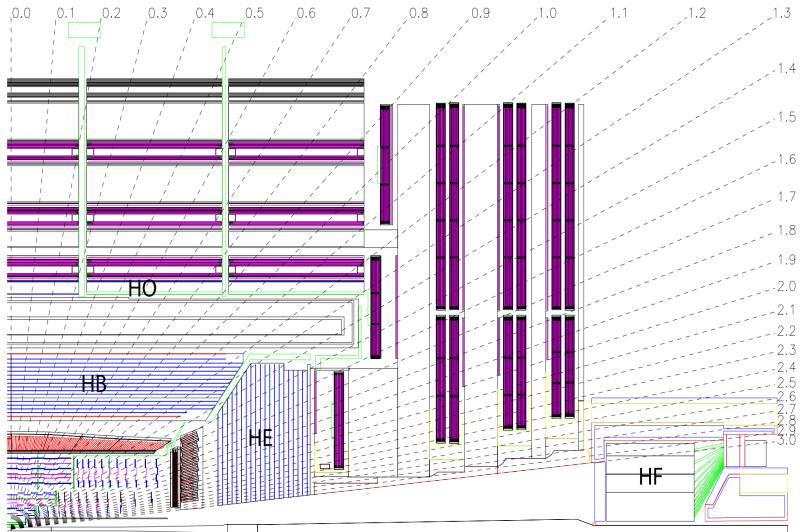


Figure 2.11: Schematic representation of the HCAL sub-system in CMS [92].

The HCAL barrel uses 36 identical azimuthal wedges as absorber, placed along the beam axis in such a way that the eventual cracks between them is smaller than 2 mm. The total absorber 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 2.12.

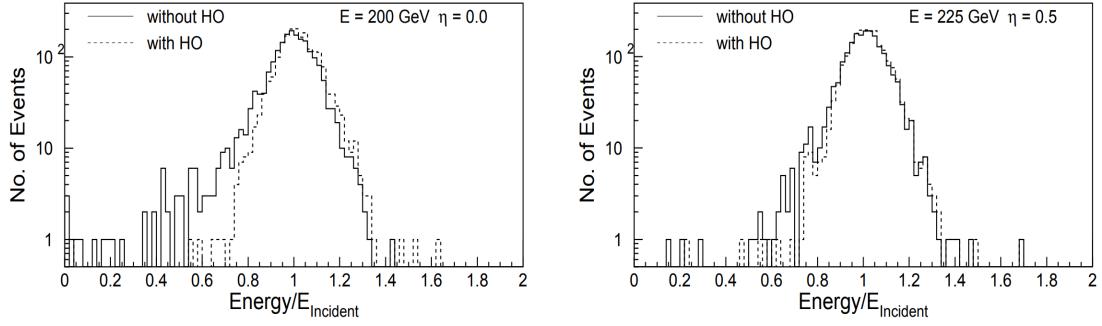


Figure 2.12: Distribution of the measured energy scaled to the incident energy for pions with incident energies of 200 GeV at $\eta = 0$ (on the left) and 225 GeV at $\eta = 0.5$ (on the right), with and without the inclusion of the HO in the HCAL system [92].

2.2.4 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 energy of 2.6 GJ. It is the largest magnet of its type ever constructed, therefore allowing the tracker, ECAL and HCAL calorimeters to be placed inside the coil, resulting in a detector that is, overall, quite compact compared to detectors of similar weight.

The magnetic field produced by this coil is extremely useful since it allows to measure quite precisely the charge and the momentum of the different charged particles interacting with the detector just by measuring the curvature of their track, according to the Lorentz equation 2.7. This solenoid has been designed to reach a momentum resolution $\Delta p/p \sim 10\%$ at $p = 1$ TeV.

$$\vec{F} = \frac{m \vec{v}^2}{R} = q \vec{E} + q \vec{v} \times \vec{B} = q \vec{v} \times \vec{B} \quad (2.7)$$

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 aluminium alloy.

The coil of the magnet is then completed with a huge steel yoke return system, as seen in Figure 2.13, made out of 6 endcap disks and 5 barrel wheels, weighting in total more than 12 000 tons and therefore accounting for most of the weight of CMS. This system is composed of many steel blocks up to 620 mm thick combined and actually also serves as the absorber plates of the HO and muon detection system that will be described in the next section.

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



Figure 2.13: Picture of the solenoid system of CMS being setup in the assembly hall.

extensively tested on the surface during the summer of 2006, before lowering down the complete solenoid in the experimental cavern where it now stands.

2.2.5 Muon system

The muon detection 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 of more than 25 000 m².

Let's now describe in details each category of muon chambers used in the different pseudorapidity areas, as shown in Figure 2.14, in order to form a muon system as hermetic as possible.

Drift tubes (DTs)

First of all, in the barrel region, where the flux of muons is low and where the magnetic field is mostly uniform and low as well, the **Drift tubes (DTs)** have been installed. This system covers the $|\eta| < 1.2$ area and has been divided into 4 different layers, each containing a number of stations optimized in order to provide a full coverage of the 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 DTs is represented in Figure 2.15.

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

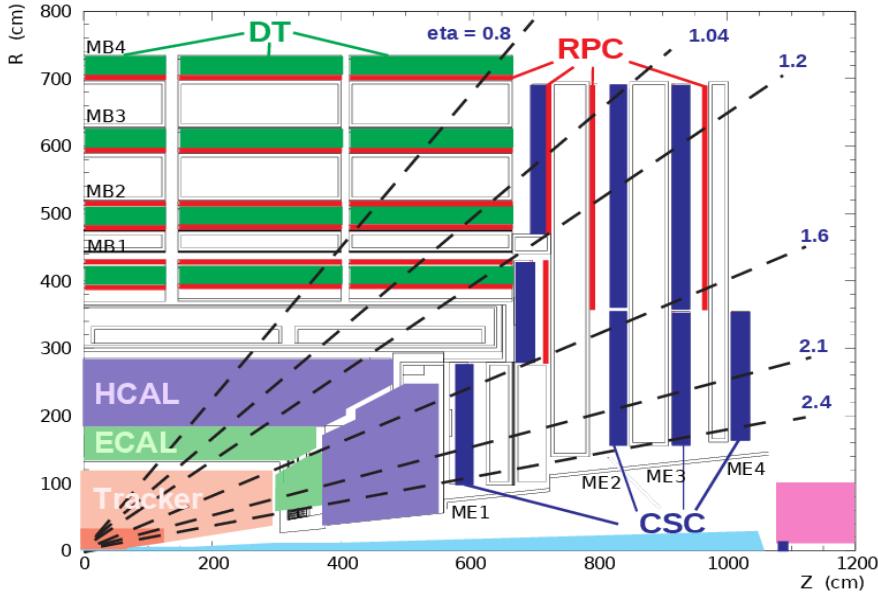


Figure 2.14: Geometrical repartition along the z-axis of the different muons chambers in CMS [94].

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 CO₂, a value small enough to produce negligible occupancy in the different wires and to avoid the need of multi-hits electronics. Redundancy of the DTs provided by the installation of multiple layers is extremely important, mainly to reduce the backgrounds coming from eventual neutrons or photons, whose rate is actually much larger than the one obtained from prompt muons.

Cathode Strip Chambers (CSCs)

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

This sub-system is made out of 540 different chambers in total, all perpendicular to the beam pipe. The sensitive plates of this sub-system are made out of 2 million wires, cover about 5000 m² and the total gas volume included in such chambers is equal to about 50 m³.

Resistive Plate Chambers (RPCs)

Finally, some **Resistive Plate Chambers (RPCs)** have been added to the barrel and to the endcap regions in order to cope with the uncertainty associated with the eventual background rates and with the (in)ability of the previous muon system to identify unequivocally the correct bunch crossing when the LHC is running at full luminosity. Indeed, the time resolution of the DTs of 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 25 ns, making it an ideal candidate to trigger the event.

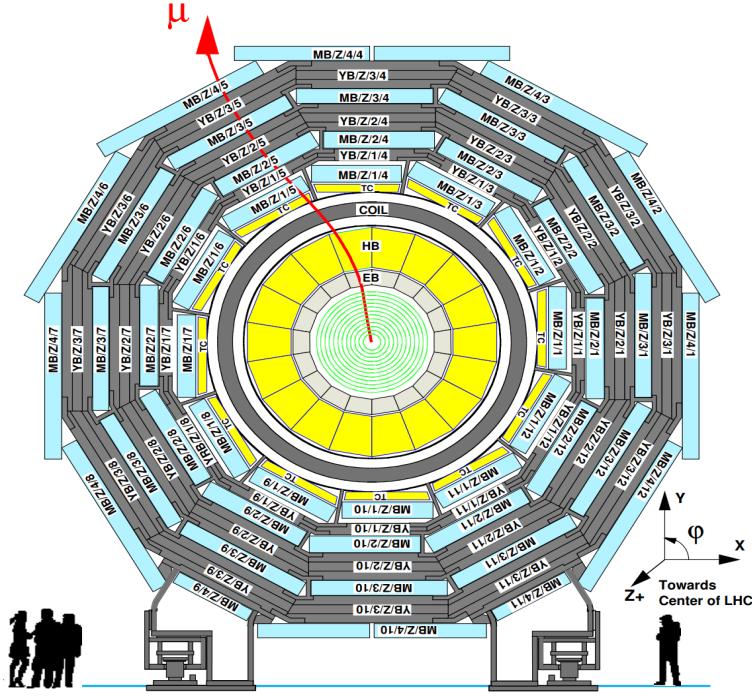


Figure 2.15: Lateral geometrical division of the different DT chambers in one of the 5 wheels of CMS [92].

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

Finally, the different features of the three muon sub-systems used by the CMS detector are summarized in Table 2.2.

Muon sub-system	DT	CSC	RPC
$ \eta $ coverage	0.0-1.2	0.9-2.4	0.0-1.9
Stations	4	4	4
Chambers	250	540	480 (barrel) 576 (endcaps)
Readout channels	172 000	266 112 (strips) 210 816 (anode channels)	68 136 (barrel) 55 296 (endcaps)
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 2.2: Comparison of the three main sub-systems currently used by CMS in order to identify muons [95].

Another advantage of the muon system such as the one built in CMS is that it can also directly be used by the trigger system, which will be described in Section 2.2.6, independently of the rest

of the detector and in addition of being able to detect, identify and measure several properties of muons crossing it.

The muon reconstruction efficiency obtained by the muon system strongly depends on the pseudorapidity value of the muon considered, along with its transverse momentum, as shown in Figure 2.16. 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. This distinction will be detailed when discussing about the muons reconstruction in Section 3.3.1).

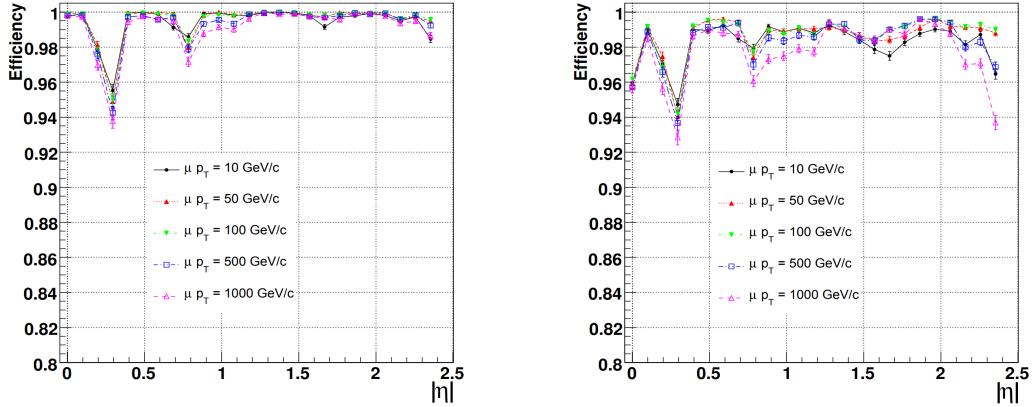


Figure 2.16: Muon reconstruction efficiency with different p_T and η values, considering only the muon system (on the left), and the combined information from the muon system and the tracker (on the right) [92].

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

The first 144 chambers of the GEM sub-system, filled with a mixture of Ar and CO₂ 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 2.17), while the rest will be set up during the next LS expected in 2024, before the phase II of operation of the LHC.

2.2.6 Trigger system

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

A system, called the trigger system, has therefore been put in place in order to select extremely quickly 1 kHz of interesting events out the 40 MHz. This system is based on two different levels: first of all, the Level-1 Trigger (L1), a hardware set of electronics selecting around 100 kHz of data,

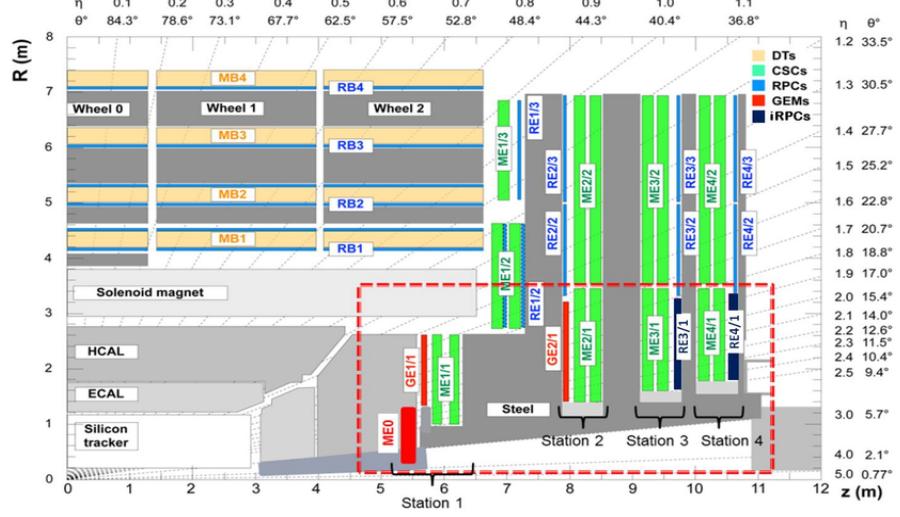


Figure 2.17: Location of the new GEM muon subsystem currently being installed in the very-forward region of CMS [92].

followed by the High-Level Trigger (HLT), a software layer improving the selection even more. Let's now give a bit more details about these two levels of trigger.

Level-1 Trigger (L1)

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

This trigger, whose architecture is represented in Figure 2.18 gets new data each 25 ns and today's electronics is not fast enough in order to deal with such a massive input of data, so an ingenious systems of buffers had to be put in place in order to put in line several events before analyzing all of them at once, using only basic segmented data provided by the detector.

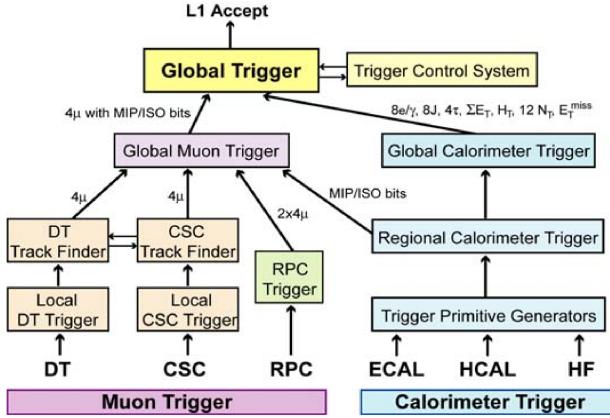


Figure 2.18: Architecture of Level-1 Trigger (L1) of CMS [92].

High-Level Trigger (HLT)

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

The HLT trigger is divided in several so-called **trigger paths**, especially developed in order to select a target object or physics process, usually with some p_T requirement in order to reduce the statistics collected by any given trigger, to avoid passing the bandwidth limit of 1kHz. These triggers paths can then for example require the presence of one or two leptons, a jet whose p_T is higher to a certain threshold or the presence of Missing Transverse Energy (MET) in the event. Some of these trigger paths need to record events with a reduced content in order to avoid saturating the data taking bandwidth. Such triggers, said to be **prescaled**, typically include the paths with low p_T requirements, as the ones used for the fake rate calculation explained in Section ??.

2.2.7 Data acquisition system

Finally, the CMS Data Acquisition System (DAQ), whose architecture is represented in Figure 2.19, has been designed to collect and analyze the data information at the nominal collision rate of 40MHz and is feeded directly from the Level-1 Trigger (L1). This means that it has to be able to read a flux of data of the order of 100kHz ($\sim 100\text{Gb/s}$) coming from approximately 650 different sources at once, while providing enough computing power for the HLT to be able to reduce this rate by a factor ~ 100 , while keeping some resources available for other tasks.

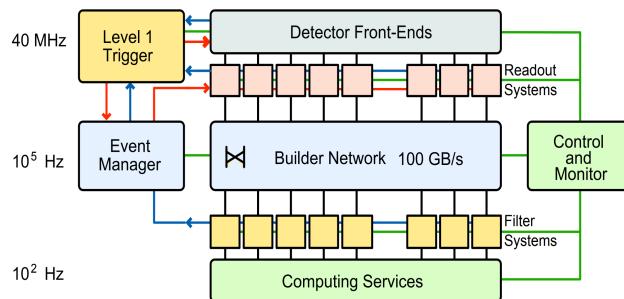


Figure 2.19: Architecture of CMS Data Acquisition System (DAQ) system [92].

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

Chapter 3

Objects reconstruction

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

The algorithm able to combine this raw data and to produce useful objects and variables is the so-called Particle Flow (PF) algorithm [96], which will be first of all described in Section 3.1. Then, a particular focus will be given to the definition and reconstruction of different objects of our particular analysis, such as the electrons and muons (Section 3.3), the jets (Section 3.4), the MET (Section 3.5) and the top reconstruction (Section 3.6) of the different pp collisions recorded.

3.1 Particle Flow (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, photons and hadrons. Then, more complex calculations are performed to identify eventual unstable particles, jets from the hadronization of quarks and to compute complex variables such as the leptons isolation and the MET.

The most basic elements used by this algorithm for the reconstruction of high-level physics objects are the tracks of charged particles, the clusters of energy left in the two calorimeters and the hits recorded in the muon chambers. For this algorithm to be as efficient as possible, the detector has been carefully designed, as described in Chapter 2: a magnetic field as large as possible and a small calorimeter granularity are indeed crucial in order to separate efficiently charged and neutral particles, and the tracker was designed to be as efficient and small as possible to have the smallest material budget possible in front of the calorimeters. The 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 3.1. Basically, the different kind of particles produced are going to interact with different parts of the detector and combining the information given by the tracker and the rest of the subsystems then allows to unequivocally identify each particle. This is usually done in a specific order in order to be as efficient as possible:

1. First of all, the most energetic **Primary Vertex (PV)** is identified by taking into account the PU and by assigning the different tracks to the different pp collisions happening during a single bunch-crossing.

All the particles originating from less energetic PV or from a secondary vertex of interaction are typically ignored and the corresponding hits in the tracker left by such particles can therefore be removed, leaving less hits available for the clustering algorithm later on, allowing for a more efficient reconstruction of the following objects.

2. Then, **muons** are the easiest particles to identify since they are at first order the only particles leaving many hits in the muon chambers placed on the outside of the detector. Each muon identified is associated to its track in the tracker, where all the hits matching a muon can therefore be removed to simplify the following reconstruction steps.
3. **Electrons** do have a charge, so they are visible by the tracker and by the ECAL, where they are going to produce an electromagnetic shower. Identifying electrons is a bit more challenging than muons because of their associated bremsstrahlung emission of photons that need to be attached to the original electrons to avoid any double counting. All the tracker hits corresponding to electrons are also removed after identification.
4. **Charged hadrons** also leave hits in the tracker and some energy deposits in the ECAL, but mostly in the HCAL, so they are easy to identify as well as a 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 then appear as some energy deposits in the ECAL for which no corresponding tracker track can be associated.
6. Finally, **neutral hadrons** can be identified as particles leaving some energy mostly in the HCAL for which no corresponding tracker track has been found as well.

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

3.2 Primary Vertex (PV) definition

Different kinds of vertices originating from a single pp collision can usually be defined, as shown in Figure 3.2: the **PU vertices**, corresponding to the different simultaneous collisions of a single bunch-crossing of the LHC, the **PV**, usually assumed to be the most energetic PU vertex and the only vertex considered in most of the physics analyses, and the **secondary vertices**, due to the eventual presence of long lived particles, decay chains or jets.

The first task of the PF algorithm is to identify all these vertices. 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

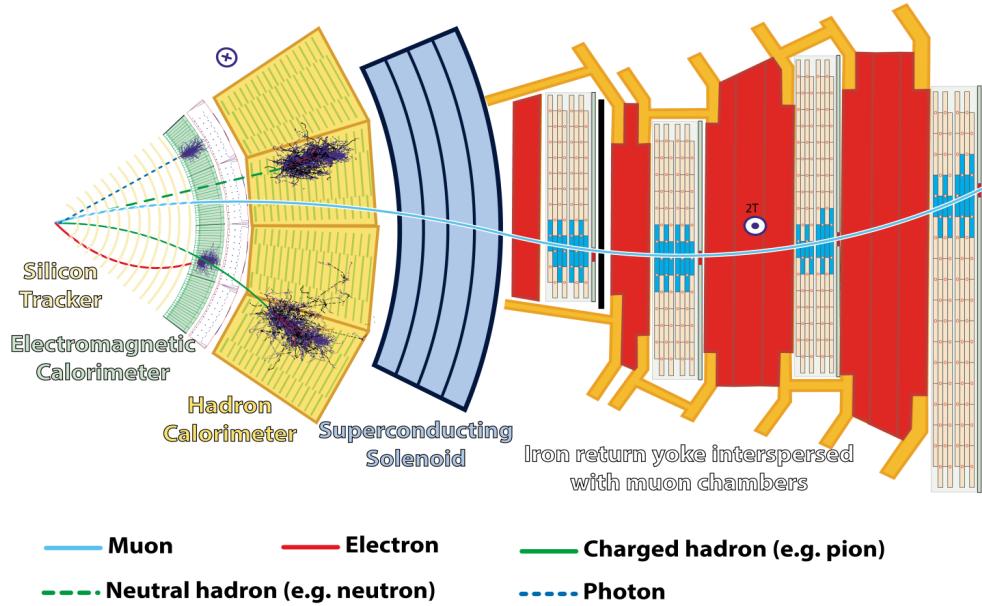


Figure 3.1: Transverse section of CMS showing the different tracks expected by different kind of particles in the detector.

considered to be the origin of the most interesting pp collision from which many different tracks are emitted.

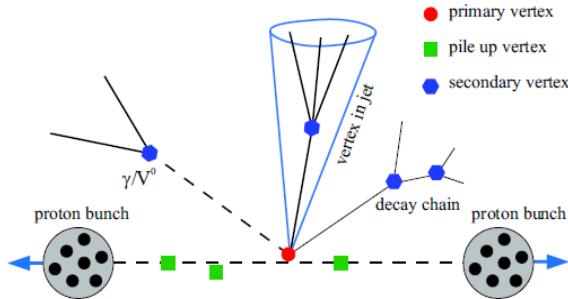


Figure 3.2: Different kind of vertices typically observed in a pp collision in the LHC.

3.3 Leptons reconstruction

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

3.3.1 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 their easy identification.

The typical signature of a muon consists in 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 into three different PF algorithms, able to reconstruct different kind of muons [95].

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 DTs, CSCs and RPCs and tries to reconstruct a vector of trajectory in each case using a Kalman Filter (KF) filter [97]. These segments are then combined in the best statistical way possible in order to form a candidate track for each muon of the event, by extrapolating the innermost vectors to the layer surface of the next chamber and by comparing it with the local track segment. The trajectory parameters are then computed and the process continues until reaching the outermost chamber, before being performed in the reverse order to estimate the innermost track parameters as well.

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

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 such muons on the other hand is able to propagate tracks identified in the inner silicon tracker (having a momentum $p > 2.5$ GeV and $p_T > 0.5$ GeV) to the muon system itself in order to try and find corresponding segments in the different muon chambers (these tracks are therefore said to be built *inside-out*). An extrapolated track and a segment are only matched if the difference between their positions in the x coordinate is smaller than 3cm 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 muons tracks, since a single hit in any of the muon chambers is enough for the candidate to be considered a

valid tracker muon, even though hadron shower remnants can for example quite easily reach the innermost muon station. The momentum assigned to such muons is the same as the one measured by the silicon tracker track itself.

Global muons

Finally, these muons are built *outside-in* since they are obtained 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 at high p_T , > 200 GeV) by exploiting the information from both the inner tracker and the muon system, while at low momentum, the best momentum resolution for muons is obtained from the inner silicon tracker directly.

Using this strategy, about 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed either as global or tracker muons, as seen in Figure 3.3.

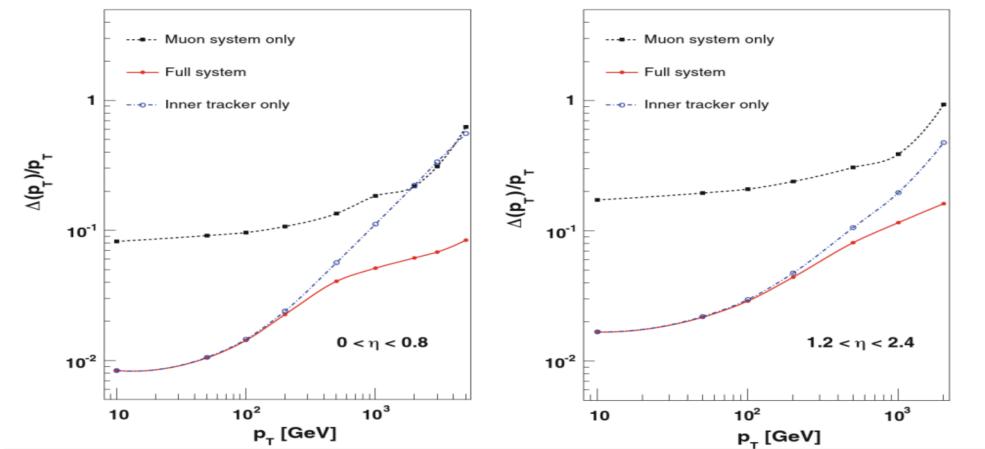


Figure 3.3: Muon p_T resolution obtained in simulation in the barrel (on the left) and endcap (on the right) for different kind of reconstructed muons [98].

Once reconstructed, candidates are required to pass some selection criteria and are then fed to the actual PF algorithm itself to start the global reconstruction of the event. This selection consists mainly in applying identification and isolation (evaluated relative to its p_T by summing up the energy in geometrical cones of radius $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ surrounding the muon in the (η, ϕ) plane, as shown in Figure 3.4) criteria in order to enhance the purity of the reconstructed prompt muons (for example, coming from the decay of a W boson) by rejecting muons coming from the decay of heavy flavour quarks, typically surrounded by a large amount of hadronic activity. The calculation of the isolation is typically performed considering only the PV since higher levels of PU are expected to bias this measurement by increasing the hadronic activity.

Different identification Working Points (WPs) can then be defined for the offline analyses, from veto to tight, in order to reject more or less contamination from misidentified leptons, keeping in mind that a tighter selection will also have an impact on the efficiency of the selection. The loose and tight WP have then been defined in order to respectively achieve 95% and 98% efficiencies.

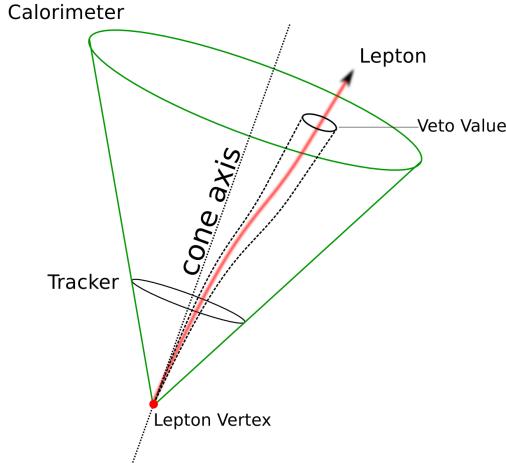


Figure 3.4: Lepton isolation cone typically used to enhance the prompt leptons purity.

These efficiencies are tunes using simulated $Z \rightarrow \mu^+\mu^-$ events with a $p_T > 20$ GeV, while the efficiency to reject muons in jets is done using simulated QCD and W+jets processes [95].

Our particular muon definition is based on the tight WP provided centrally by the muon Physics Object Group (POG), but has been tweaked based on the requirements of our group and will be detailed in Chapter ??.

3.3.2 Electrons

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

It is usually a bit harder to reconstruct electrons than muons mainly because electrons do interact with the tracker and this interaction therefore needs to be modeled to understand the exact behaviour of such particles: this interaction is for example responsible for the emission of secondary bremsstrahlung photons crashing into the ECAL but not coming from the PV. In fact, it is estimated that in CMS between 33% and 86% of the energy of an electron is actually radiated before it reaches the ECAL, depending mostly on its pseudorapidity [99]. In order to measure precisely the energy of an electron, all the photons emitted by bremsstrahlung before reaching the ECAL (usually, along the ϕ axis because of the deviation implied by the solenoid) then need to be collected as well and associated to the correct electron of the event.

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

1. A **clustering algorithm** is first of all defining the so-called **Super Cluster (SC)**. Its goal is to reconstruct the particle showers individually by identifying a seed crystal for the cluster, defined as the crystal collecting the most energy, since the energy deposited in the ECAL is usually spread into several different crystals because of the electromagnetic shower effect discussed in Section 2.2.2 and because of the bremsstrahlung emission of photons due to the interaction with the tracker. The algorithm therefore searches for eventual crystals around this seed whose energy detected would be superior to 2σ of the electronic noise and matching some quality criteria ($E_{\text{seed}} > 230$ MeV in the barrel, $E_{\text{seed}} > 600$ MeV and $E_{\text{seed}}^T > 150$

MeV in the endcaps).

The excited contiguous crystals found are then grouped into clusters, themselves considered candidates for the final global cluster, the SC, if their energy is higher than another given threshold ($E_{\text{cluster}} > 350$ MeV in the barrel, $E_{\text{cluster}}^T > 1$ GeV in the endcaps) [99]. The SC energy is then given by the sum of the energies of all its constituent clusters, while its position is calculated as the energy-weighted mean position of the different clusters.

2. Once the SC identified, **electron tracker tracks** are reconstructed using a procedure a bit different than the usual KF reconstruction method for all the tracks of the silicon tracker [97] because of the large radiative losses for electrons in the tracker material.

This reconstruction is known to be very time consuming, so a good identification of potential electron seeds has to be performed as the method efficiency greatly relies on this first identification. Two different strategies can be used to perform this seeding (even though the electron seeds found using the two algorithms are usually combined afterwards):

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

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

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

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

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

This complete electron workflow explained here in a simplified way can be summarized in Figure 3.5.

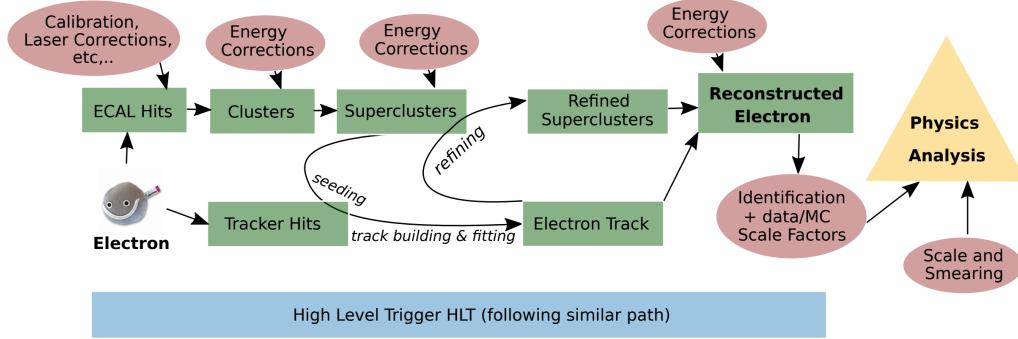


Figure 3.5: Schematic representation of the full electron reconstruction workflow in CMS [100].

3.4 Jets reconstruction

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

This practically means that once a single quark is produced, it will start losing energy by forming new $q\bar{q}$ pairs, themselves forming additional $q\bar{q}$ pairs. This chain continues until the resulting pairs of quarks have such a low energy that they can start combining into colourless hadrons. This is called the *hadronization* process and the actual result of the apparition of a quark is a shower 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 3.6.

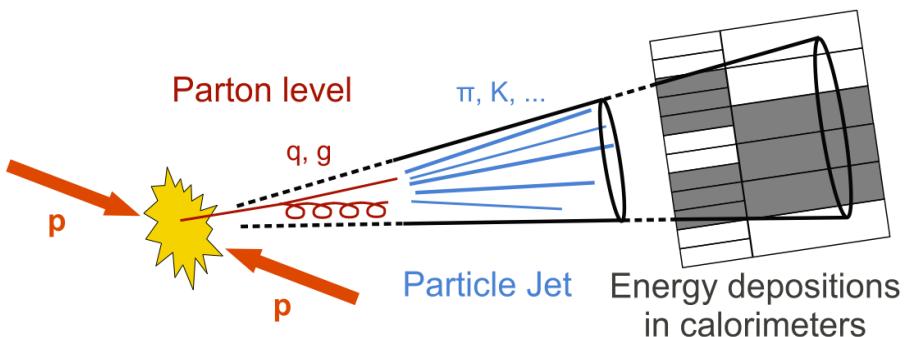


Figure 3.6: Schematic representation of the typical development of a jet within the CMS detector.

Several algorithms can be used to reconstruct the jets by linking the information coming from the tracker and the calorimeters, but the most used tool in CMS is the so-called anti- k_T algorithm, able to cluster all the charged and neutral hadrons along with the eventual non-isolated photons or lepton produced and merge them into a single jet [101]. 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 3.1, these distances

can be computed using different variables such as $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i + \phi_j)^2$, the distance between i and j in the (η, ϕ) space, the p_T^2 of each particle and R , the clustering algorithm radius parameters 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 is summed to get the momentum of the jet itself.

$$\begin{cases} d_{ij} = \min \left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \frac{\Delta R_{ij}^2}{R^2} \\ d_{iB} = \frac{1}{p_{T,i}^2} \end{cases} \quad (3.1)$$

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

- If d_{ij} is smaller 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 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. This typically introduces a source of systematic uncertainty that will be taken into account and discussed in Section ??.

The efficiency of the PF algorithm for jet identification and reconstruction has been checked using simulation, as shown in Figure 3.7. This study clearly shows that between 95 and 97% of the energy of the PF jet candidates can be reconstructed, compared to a 40-60% reconstruction efficiency using only the calorimeters data, and that this algorithm also leads to a gain in resolution up to a factor 3, depending on the jet p_T .

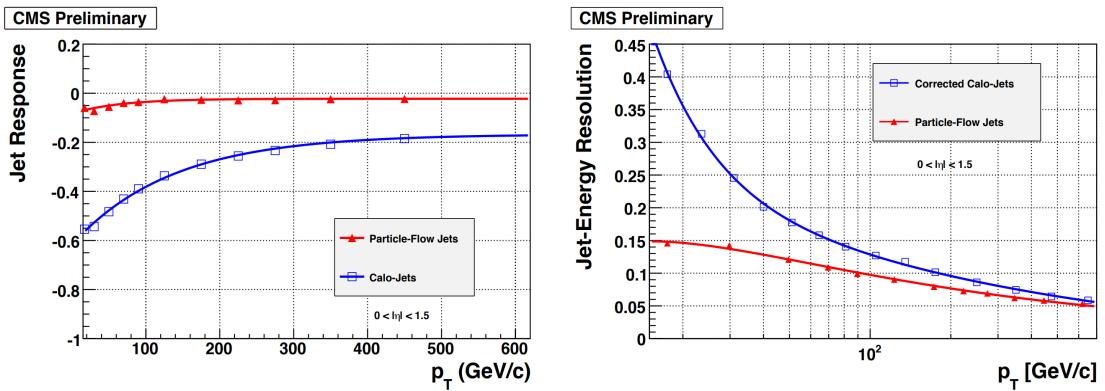


Figure 3.7: Comparison of the jet response (on the left) and jet energy resolution (on the right) for dijets simulated events in the barrel for jets reconstructed using only the calorimeters (in blue) and jet candidates from the PF algorithm (in red) [102].

3.4.1 B-tagging

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

This specific kind of jets can be distinguished from other jets because of the relatively long lifetime of the bottom quark ($\tau \sim 1.5\text{ps}$) that produces in the detector a secondary vertex displaced by a few millimeters with respect to the PV, as shown in Figure 3.8 ; and this gives a perfect way to discriminate b-jets and 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.

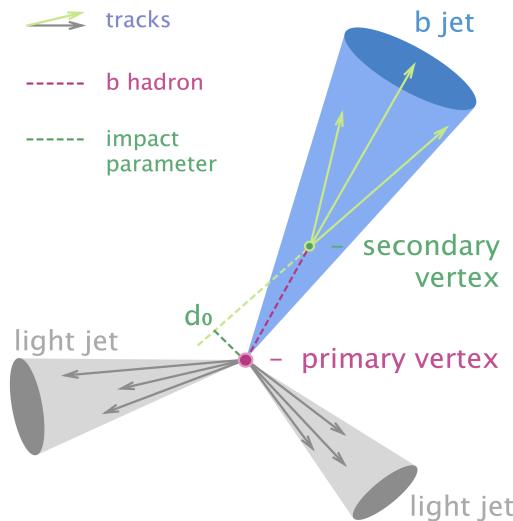


Figure 3.8: Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex.

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

In this analysis, the typical deep Combined Secondary Vertex (CSV) algorithm able to combine the information on the secondary vertex with the one on the track impact parameters and based on a Deep Neural Network (DNN) previously trained, has been used to identify such b-jets. The performance of this method can be observed in Figure 3.9, where we can see that this deep CSV algorithm is one of the best algorithms able to identify b-jets, depending on the phase space, while keeping a relatively low misidentification rate for light-flavor jets (u, d, s and gluons).

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

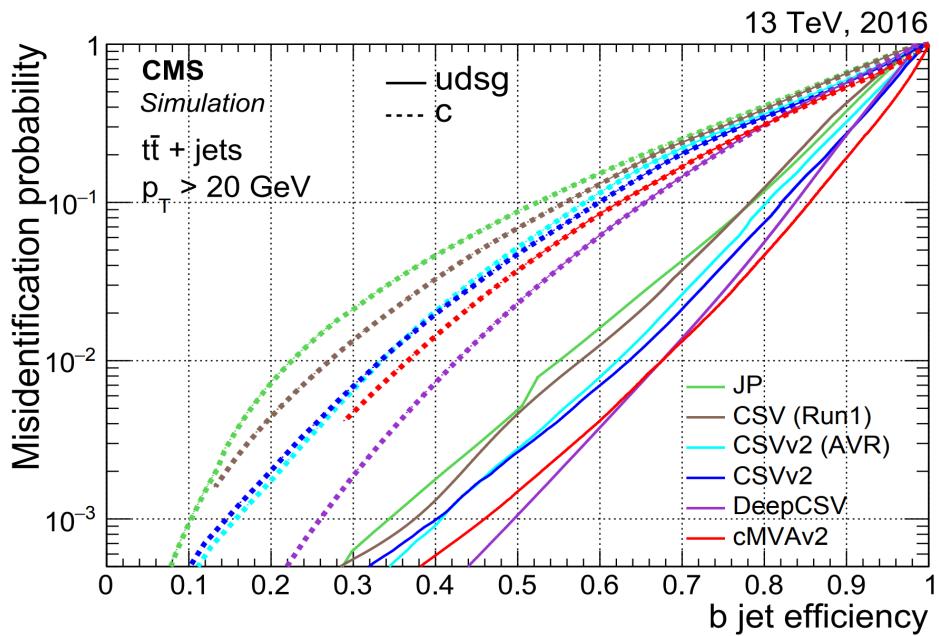


Figure 3.9: B-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [103].

3.5 Missing Transverse Energy (MET)

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

However, this statement is not totally true since we are aware of several effects that could induce an imbalance in this transverse momentum, as shown in Figure 3.10:

- Even though the CMS detector has been carefully designed, 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 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, leading to an instrumental MET in some cases.
- Finally, the eventual exotic weakly interacting particles produced, such as Dark Matter (DM), is typically expected to leave some MET in the detector as well.

The Missing Transverse Energy (MET) variable, defined in Equation 3.2 as the negative sum of the transverse momentum of all the particles j of the event, is able to take into account this eventual imbalance in the transverse momentum and is therefore a key variable in most of the analyses

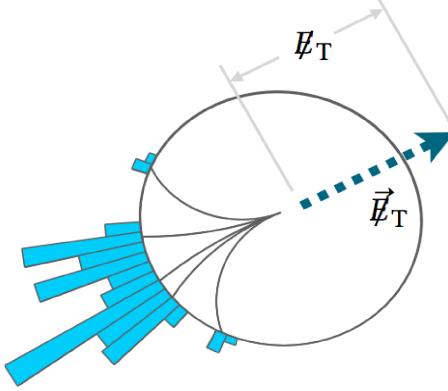


Figure 3.10: Schematic representation of the Missing Transverse Energy (MET).

searching for new BSM physics, which is not expected to interact with the detector.

$$\vec{p}_T^{\text{miss}} = \vec{E}_T = - \sum_j \vec{p}_{T,j} \quad (3.2)$$

Different algorithms can be used in order to reconstruct this variable, the most famous being [104]:

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

Several corrections also need to be applied to this spectrum to filter anomalous high MET events arising because of a variety of reconstruction failures induced by the detector due to several effects, such as the electronic noise and eventual dead cells in the calorimeters or the presence of an eventual beam halo particles from the LHC itself, leading to a global miscalculation of the final energy of the event. These filters are extremely important, especially in the end of the MET spectrum, as observed in Figure 3.11.

3.6 Top reconstruction

Although not formerly a part of the PF algorithm and done offline, the kinematic reconstruction of the SM $t\bar{t}$ process detailed in Section ?? is still an extremely important part of this analysis.

As previously explained, because of the high mass of the top quark, it is expected to decay before reaching the detector, almost 100% of the time into a $b\bar{b}$ pair of quarks, which can be identified

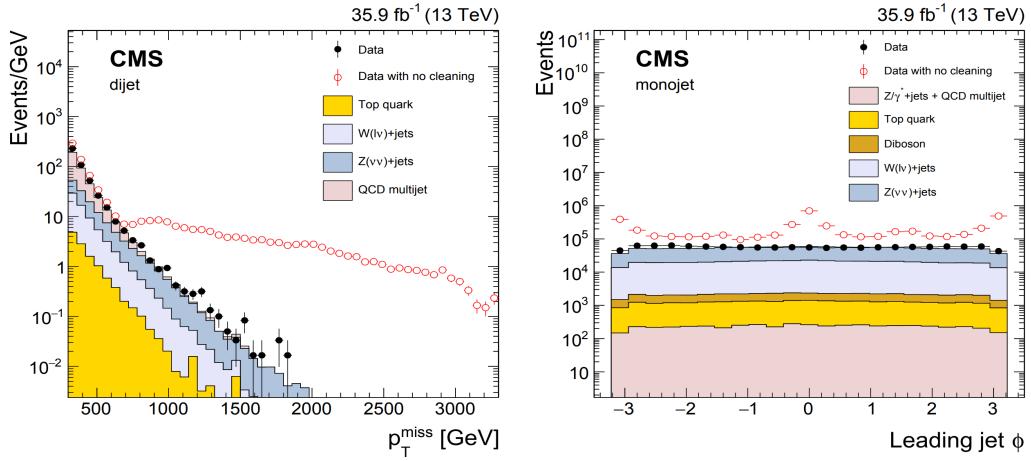


Figure 3.11: MET and jet ϕ distributions with and without MET filters applied [104].

using the b-tagging method explained in Section 3.4.1, and a W boson, unstable as well (in our particular case, only the W decay to a lepton and a neutrino is considered for reasons explained in Section ??, even though this decay has a low Branching Ratio (BR)).

The ideal kinematics of such an event can be expressed using the four-momenta of the different particles involved, as expressed in Equations 3.3a to 3.3c, if we assume that the MET of the event is coming only from the two neutrinos produced.

$$\begin{cases} p_x^{\text{miss}} = p_{\nu_x} + p_{\bar{\nu}_x} \\ p_y^{\text{miss}} = p_{\nu_y} + p_{\bar{\nu}_y} \end{cases} \quad (3.3a)$$

$$\begin{cases} m_{W+}^2 = (E_{l+} + E_\nu)^2 - (\vec{p}_{l+} + \vec{p}_\nu)^2 \\ m_{W-}^2 = (E_{l-} + E_{\bar{\nu}})^2 - (\vec{p}_{l-} + \vec{p}_{\bar{\nu}})^2 \end{cases} \quad (3.3b)$$

$$\begin{cases} m_t^2 = (E_b + E_{l+} + E_\nu)^2 - (\vec{p}_b + \vec{p}_{l+} + \vec{p}_\nu)^2 \\ m_{\bar{t}}^2 = (E_{\bar{b}} + E_{l-} + E_{\bar{\nu}})^2 - (\vec{p}_{\bar{b}} + \vec{p}_{l-} + \vec{p}_{\bar{\nu}})^2 \end{cases} \quad (3.3c)$$

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

We therefore have 6 equations to solve and exactly 6 unknowns corresponding to the three momentum components of each neutrino produced, a problem that can in principle be solved, leading to a quartic equation in one of the unknowns analytically solvable but quite ambiguous given the variable number of solutions of such equation [105]. Solving this problem considering SM $t\bar{t}$ is in this sense extremely helpful since it gives us access to the p_T of the neutrinos produced.

3.6.1 Top reconstruction in practice

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

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

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

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

3.6.2 Top reconstruction with additional DM

Appendices

Appendix A

Samples used

A.1 Data samples

All the data samples considered for this analysis are listed in Tables A.1, A.2 and A.3. The luminosity of each dataset has been computed using the Brilcalc tool provided by CMS [131], while the number of generated events has been obtained using the CERN official Data Aggregation System (DAS).

A.2 Signal samples

The MC signal samples have been produced centrally, considering different dark matter and mediators (scalar, pseudoscalar) masses, for both the $t+DM$ and $t\bar{t}+DM$ analyses. All the mass points considered along with their respective cross sections can be found in Tables ?? ($t+DM$ analysis) and A.4. The mass points generated are the same for 2016, 2017 and 2018.

To be updated once the tDM signals are actually available

A.3 Backgrounds samples

All the background MC samples considered for this analysis are listed in Tables A.5, A.6 and A.7 for 2016, 2017 and 2018 respectively.

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

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

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

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

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

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

Mass point	Cross-section [pb]
Scalar mediators	
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_50	7.136
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_100	1.686
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_150	0.5619
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_200	0.2526
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_250	0.1337
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_300	0.08323
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_350	0.05696
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_400	0.03561
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_450	0.02326
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_500	0.01614
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_20_mPhi_100	1.687
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_30_mPhi_100	1.691
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_40_mPhi_100	1.709
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_45_mPhi_100	1.728
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_49_mPhi_100	1.747
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_51_mPhi_100	0.02068
TTbarDMJets_Dilepton_scalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_55_mPhi_100	0.01038
Pseudoscalar mediators	
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_50	0.759
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_100	0.4889
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_150	0.3269
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_200	0.242
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_250	0.1707
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_300	0.1222
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_350	0.07111
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_400	0.03719
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_450	0.02489
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_1_mPhi_500	0.01783
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_20_mPhi_100	0.4881
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_30_mPhi_100	0.4896
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_40_mPhi_100	0.4893
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_45_mPhi_100	0.4903
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_49_mPhi_100	0.4931
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_51_mPhi_100	0.04549
TTbarDMJets_Dilepton_pseudoscalar_LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8_mChi_55_mPhi_100	0.0183

Table A.4: Signal samples mass points considered for the $t\bar{t}+DM$ signal used in this analysis.

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

Table A.5: Main 2016 MC simulations for the different background processes considered for this analysis and their respective cross sections.

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

Table A.6: Main 2017 MC simulations for the different background processes considered for this analysis and their respective cross sections.

Process	Sample	Cross section [pb]
DY	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610.0
	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6189.39
TTTo2L2Nu	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	3.360
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	26.38
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	44.33
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60
WW	WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8	12.178
	WWJJToLNuLNu_EWK_TuneCP5_13TeV-madgraph-pythia8	0.4286
	GluGluToWWTo*_TuneCP5_13TeV_MCFM701_pythia8	0.06387
V γ /V γ^*	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	405.271
	ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	131.300
	WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg-pythia8	58.59
VZ	ZZTo2L2Nu_TuneCP5_13TeV_powheg-pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.212
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
VVV	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01398
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
	WWZ_TuneCP5_13TeV-amcatnlo-pythia8	0.16510
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.18331
Non-Prompt	Data-driven (tight-to-loose method)	

Table A.7: Main 2018 MC simulations for the different background processes considered for this analysis and their respective cross sections.

Appendix B

Neural network optimization

List of figures

2.1	LHC injection chain and experiments performed at CERN [89].	6
2.2	Integrated luminosity collected by CMS over its different years of operation so far.	9
2.3	Mean PU distribution and luminosity recorded by CMS over the different years of operation of the LHC.	10
2.4	Schematic representation of the CMS detector, along with all its subdetectors and main characteristics.	11
2.5	Schematic representation of the CMS coordinate system used by convention.	12
2.6	Schematic representation of the CMS tracker, for different pseudorapidity values and along the z-axis [93].	13
2.7	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) [92].	14
2.8	Tracker reconstruction efficiency of muons (on the left) and pions (on the right) calculated in simulation for different pseudorapidities and particle momenta (1, 10 and 100 GeV) [92].	14
2.9	Schematic representation of the sub-systems of the CMS ECAL [92].	15
2.10	Schematic representation of a typical electromagnetic shower and the radiation length X_0 concept [92].	16
2.11	Schematic representation of the HCAL sub-system in CMS [92].	17
2.12	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 [92].	18
2.13	Picture of the solenoid system of CMS being setup in the assembly hall.	19
2.14	Geometrical repartition along the z-axis of the different muons chambers in CMS [94].	20
2.15	Lateral geometrical division of the different DT chambers in one of the 5 wheels of CMS [92].	21
2.16	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) [92].	22

2.17	Location of the new GEM muon subsystem currently being installed in the very-forward region of CMS [92].	23
2.18	Architecture of Level-1 Trigger (L1) of CMS [92].	23
2.19	Architecture of CMS Data Acquisition System (DAQ) system [92].	24
3.1	Transverse section of CMS showing the different tracks expected by different kind of particles in the detector.	27
3.2	Different kind of vertices typically observed in a pp collision in the LHC.	27
3.3	Muon p_T resolution obtained in simulation in the barrel (on the left) and endcap (on the right) for different kind of reconstructed muons [98].	29
3.4	Lepton isolation cone typically used to enhance the prompt leptons purity.	30
3.5	Schematic representation of the full electron reconstruction workflow in CMS [100].	32
3.6	Schematic representation of the typical development of a jet within the CMS detector.	32
3.7	Comparison of the jet response (on the left) and jet energy resolution (on the right) for dijets simulated events in the barrel for jets reconstructed using only the calorimeters (in blue) and jet candidates from the PF algorithm (in red) [102].	33
3.8	Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex.	34
3.9	B-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [103].	35
3.10	Schematic representation of the Missing Transverse Energy (MET).	36
3.11	MET and jet ϕ distributions with and without MET filters applied [104].	37

List of tables

2.1	Expected and observed main parameters of pp operation of the LHC across the different eras of operation [88].	7
2.2	Comparison of the three main sub-systems currently used by CMS in order to identify muons [95].	21
A.1	Datasets collected in 2016 and considered for this analysis.	42
A.2	Datasets collected in 2017 and considered for this analysis.	43
A.3	Datasets collected in 2018 and considered for this analysis.	44
A.4	Signal samples mass points considered for the $t\bar{t}$ +DM signal used in this analysis. .	45
A.5	Main 2016 MC simulations for the different background processes considered for this analysis and their respective cross sections.	46
A.6	Main 2017 MC simulations for the different background processes considered for this analysis and their respective cross sections.	47
A.7	Main 2018 MC simulations for the different background processes considered for this analysis and their respective cross sections.	48

Bibliography

- [1] G. Altarelli, "The Standard Model of Particle Physics", CERN-PH-TH/2005-206, 2005
- [2] CMS Collaboration, "The CMS Experiment at the CERN LHC", JINST 3 S08004, 2008
- [3] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", JINST 3 S08003, 2008
- [4] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", Phys. Rev. Lett. 13, pp. 321-323, 1964
- [5] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett. 13, pp. 508-509, 1964
- [6] 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]
- [7] 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]
- [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. Mu?oz, 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] CMS Collaboration, "Search for the production of dark matter in association with top-quark pairs in the single-lepton final state in proton-proton collisions at $\sqrt{s} = 8$ TeV", JHEP, vol. 6 121, 2015
- [15] CMS Collaboration, "Search for the Production of Dark Matter in Association with Top Quark Pairs in the Di-lepton Final State in pp collisions at $\sqrt{s} = 8$ TeV", CMS-PAS-B2G-13-004, 2014

- [16] "Search for dark matter in events with heavy quarks and missing transverse momentum in pp collisions with the ATLAS detector", Eur. Phys. J. C (2015) 75:92
- [17] ATLAS Collaboration, Search for the Supersymmetric Partner of the Top Quark in the Jets+Emiss Final State at $\sqrt{s} = 13$ TeV", ATLAS-CONF-2016-077
- [18] ATLAS Collaboration, "Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector", ATLAS-CONF-2016-050, 2016
- [19] ATLAS Collaboration, "Search for direct top squark pair production and dark matter production in final states with two leptons in $\sqrt{s} = 13$ TeV pp collisions using 13.3 fb^{-1} of ATLAS data", ATLAS-CONF-2016-076, 2016
- [20] 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]
- [21] CMS Collaboration, Search for dark matter produced in association with heavy-flavor quark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV", Eur. Phys. J. C (2017) 77: 845
- [22] 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]
- [23] 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]
- [24] S. Manzoni, "The Standard Model and the Higgs Boson", Physics with Photons Using the ATLAS Run 2 Data, Springer Theses, 2019
- [25] 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]
- [26] 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
- [27] H. Poincare, "The Milky Way and the Theory of Gases", Popular Astronomy, vol. 14, pp.475-488, 1906
- [28] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln", Helvetica Physica Acta , vol. 6, pp. 110-127, 1933
- [29] S. Van den Bergh, Phys Rev D "The early history of dark matter", Dominion Astrophysical Observatory, 1999
- [30] V.C. Rubin, W.K. Ford, "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions", Astrophysical Journal 159, p. 379, 1970
- [31] A. A. Penzias, R.W. Wilson, "A Measurement of Excess Antenna Temperature at 4080 Mc/s", Astrophysical Journal 142, pp. 419-421
- [32] D.J. Fixsen, "The temperature of the cosmic microwave background", Astrophysical Journal, 2009
- [33] Planck Collaboration, "Planck 2018 results. I. Overview and the cosmological legacy of Planck", 2018 [arXiv: 1807.06205]

- [34] R. Tojeiro, "Understanding the Cosmic Microwave Background Temperature Power Spectrum", 2006
- [35] Planck Collaboration, "Planck 2018 results. VI. Cosmological parameters", 2018 [arXiv: 1807.06209]
- [36] "Astrophysical Constants and Parameters", 2019
- [37] D. Clowe et all., "A Direct Empirical Proof of the Existence of Dark Matter", *Astrophysical Journal Letters* 648, 2006
- [38] L. Heurtier, H. Partouche, "Spontaneous Freeze Out of Dark Matter From an Early Thermal Phase Transition", CPHT-RR065.112019 [arXiv: 1912.02828]
- [39] 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]
- [40] C.S. Frenk, S.D.M. White, "Dark matter and cosmic structure", *Annalen der Physik*, p. 22 , 2012 [arXiv: 1210.0544]
- [41] R. Kirk, "Dark matter genesis"
- [42] M. Drewes et all., "A White Paper on keV Sterile Neutrino Dark Matter", 2016 [arXiv: 1602.04816]
- [43] C. Alcock et all., "The MACHO Project: Microlensing Results from 5.7 Years of LMC Observations", *Astrophys.J.* 542 (2000) 281-307
- [44] 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)
- [45] EROS and MACHO collaborations, "EROS and MACHO Combined Limits on Planetary Mass Dark Matter in the Galactic Halo", 1998
- [46] Particle Data Group, "Neutrino Cross Section Measurements", PDG 2019
- [47] K. McFarland, "Neutrino Interactions", 2008 [arXiv: 0804.3899]
- [48] E. Morgan, "Aspects of WIMP Dark Matter Searches at Colliders and Other Probes", Springer theses, 2016
- [49] F. Couchot et all., "Cosmological constraints on the neutrino mass including systematic uncertainties", *A & A* 606, A104 (2017)
- [50] E. Bulbul et all., "Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters", 2014 [arXiv: 1402.2301]
- [51] 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]
- [52] 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]
- [53] 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]
- [54] D. Wu, "A Brief Introduction to the Strong CP Problem", Superconducting Super Collider Laboratory, 1991

- [55] R.D. Peccei, H.R. Quinn, "CP Conservation in the Presence of Pseudoparticles", Phys. Rev. Lett. 38, 1440, 1977
- [56] P.W. Graham et all., "Experimental Searches for the Axion and Axion-like Particles", Annual Review of Nuclear and Particle Science 65, 2015 [arXiv: 1602.00039]
- [57] CAST collaboration, "New CAST limit on the axion-photon interaction", Nature Physics 13, pp. 584-590 (2017)
- [58] S.K. Vempati, "Introduction to MSSM", 2012 [arXiv: 1201.0334]
- [59] B. Penning, "The Pursuit of Dark Matter at Colliders - An Overview", 2017 [arXiv: 1712.01391]
- [60] M. Schumann, "Direct Detection of WIMP Dark Matter: Concepts and Status", J. Phys. G46 (2019) no.10, 103003 [arXiv: 1903.03026]
- [61] S.C. Martin et all., "The RAVE survey: constraining the local Galactic escape speed", Mon.Not.Roy.Astron.Soc.379:755-772, 2007
- [62] K. Freese, M. Lisanti, C. Savage, "Annual Modulation of Dark Matter: A Review", [arXiv: 1209.3339v3]
- [63] T.M. Undagoitia and L. Rauch, "Dark matter direct-detection experiments", J. Phys. G43 (2016) no.1, 013001 [arXiv: 1509.08767]
- [64] 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]
- [65] J.M. Gaskins, "A review of indirect searches for particle dark matter", Contemporary Physics, 2016 [arXiv: 1604.00014]
- [66] F.S. Queiroz, "Dark Matter Overview: Collider, Direct and Indirect Detection Searches", Max-Planck Institute of Physics
- [67] LAT collaboration, "Constraints on Dark Matter Annihilation in Clusters of Galaxies with the Fermi Large Area Telescope", JCAP 05(2010)025 [arXiv: 1002.2239]
- [68] A.A. Moiseev et all., "Dark Matter Search Perspectives with GAMMA-400", 2013 [arXiv: 1307.2345]
- [69] L. Covi et all., "Neutrino Signals from Dark Matter Decay", JCAP 1004:017, 2010 [arXiv: 0912.3521]
- [70] 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]
- [71] J. Abdallah et all., "Simplified Models for Dark Matter Searches at the LHC", Phys. Dark Univ. 9-10 (2015) 8-23 [arXiv: 1506.03116]
- [72] 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]
- [73] 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]
- [74] 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]

- [75] 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]
- [76] 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]
- [77] 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]
- [78] 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]
- [79] 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]
- [80] 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]
- [81] C. Munoz, "Models of Supersymmetry for Dark Matter", FTUAM 17/2, IFT-UAM/CSIC-17-005, 2017 [arXiv: 1701.05259]
- [82] CMS Collaboration, "Searches for invisible decays of the Higgs boson in pp collisions at $s_{\text{sqrt}s} = 7, 8,$ and 13 TeV", JHEP 02 (2017) 135 [arXiv: 1610.09218]
- [83] J. Alimena et all., "Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider", 2019 [arXiv: 1903.04497]
- [84] 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]
- [85] M. Tanabashi et al., Particle Data Group, Phys. Rev. D98, 030001 (2018)
- [86] R. Schicker, "The ALICE detector at LHC", 2005
- [87] LHCb Collaboration, "LHCb Detector Performance", Int. J. Mod. Phys. A 30, 1530022 (2015) [arXiv: 1412.6352]
- [88] J.T. Boyd, "LHC Run-2 and Future Prospects", 2020
- [89] E. Gschwendtner, "AWAKE, A Particle-driven Plasma Wakefield Acceleration Experiment", CERN Yellow Report CERN 2016-001, pp.271-288 [arXiv: 1705.10573]
- [90] M. Thomson, "Modern Particle Physics", Cambridge University Press, 2013
- [91] G. Apollinari et all., "High Luminosity Large Hadron Collider HL-LHC", CERN Yellow Report CERN-2015-005, pp.1-19 [arXiv: 1705.08830]
- [92] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004
- [93] CMS Collaboration, "Precision measurement of the structure of the CMS inner tracking system using nuclear interactions", JINST 13 (2018) P10034 [arXiv: 1807.03289]
- [94] M.S. Kim, "CMS reconstruction improvement for the muon tracking by the RPC chambers", 2013 JINST 8 T03001 [arXiv: 1209.2646]

- [95] 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]
- [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] F. Beaudette, "The CMS Particle Flow Algorithm", 2014 [arXiv: 1401.8155]
- [103] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", JINST 13 (2018) P05011 [arXiv: 1712.07158]
- [104] 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]
- [105] L. Sonnenschein, "Analytical solution of ttbar dilepton equations", Phys.Rev.D73:054015, 2016
- [106] M.H. Seymour and M. Marx, "Monte Carlo Event Generators", MCnet-13-05, 2013 [arXiv:1304.6677]
- [107] B. Cabouat, J.R. Gaunt and K. Ostrolenk, "A Monte-Carlo Simulation of Double Parton Scattering", JHEP11(2019)061 [arXiv: 1906.04669]
- [108] R. Placakyte, "Parton Distribution Functions", 2011 [arXiv:1111.5452]
- [109] J. Alwall et all., "MadGraph 5 : Going Beyond", 2011 [arXiv:1106.0522]
- [110] C. Oleari, "The POWHEG-BOX", Nucl.Phys.Proc.Suppl.205-206:36-41 [arXiv:1007.3893]
- [111] S. Frixione et all., "The MC@NLO 4.0 Event Generator", CERN-TH/2010-216 [arXiv: 1010.0819]
- [112] B. Webber, "Parton shower Monte Carlo event generators", Scholarpedia
- [113] M. Bahr et all., "Herwig++ Physics and Manual", Eur.Phys.J.C58:639-707, 2008 [arXiv: 0803.0883]
- [114] T. Sjostrand, "A Brief Introduction to PYTHIA 8.1" Comput.Phys.Commun.178:852-867, 2008 [arXiv: 0710.3820]
- [115] A. Karneyeu et all., "MCPlots: a particle physics resource based on volunteer computing", European Physical Journal C 74 (2014) [arXiv: 1306.3436]

- [116] V. Lefebure and S. Banerjee, "CMS Simulation Software Using Geant4", CMS-NOTE-1999-072, 1999
- [117] S. Banerjee, "Validation of Geant4 Physics Models Using Collision Data from the LHC", J. Phys.: Conf. Ser. 898 042005
- [118] A. Rizzi, G. Petrucciani and M. Peruzzi, "A further reduction in CMS event data for analysis: the NANO AOD format", J. Phys.: Conf. Ser. 214 06021
- [119] 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
- [120] K. Bloom, "CMS software and computing for LHC Run 2", ICHEP 2016 [arXiv: 1611.03215]
- [121] W. Tanenbaum, "A ROOT/IO Based Software Framework for CMS", ECONF C0303241:TUKT010, 2003
- [122] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS-PAS-LUM-17-001, 2017
- [123] CMS Collaboration, "CMS Luminosity Measurements for the 2017 Data Taking Period", CMS-PAS-LUM-17-001, 2018
- [124] CMS Collaboration, "CMS Luminosity Measurements for the 2018 Data Taking Period", CMS-PAS-LUM-17-001, 2019
- [125] CMS Twiki, "Reweighting recipe to emulate Level 1 ECAL prefiring"
- [126] 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]
- [127] CMS Twiki, "Electron and Photon Physics Object Offline Guide"
- [128] CMS Twiki, "Baseline muon selections for Run-II", 2019
- [129] CMS Twiki, "Jet Identification"
- [130] CMS Twiki, "Heavy flavor identification at CMS with deep neural networks"
- [131] CMS Collaboration, "BRIL Work Suite"