



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 en español**



# 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 both the personal and professional points of view, and something I will definitely benefit from in the future.

I would also like to personally thank Jónatan Piedra for supervising this thesis, but also for all your help and patience during all these years, especially at the beginning. You taught me a lot and I cannot count the number of bugs and issues you helped me solve, so I am thankful you kept your door open at all times anyway.

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

I would also like to thank all the people who helped me at some point: Alicia, Rocío, Pupi, Chus and Celso, always there to help me fight the administrative related issues I faced. This work would not have been possible either without the help of all the people involved in this particular analysis, such as Deborah and Dominic. Thank you as well to all the persons involved in the development of the Latino framework, which quickly became a great tool to work with. Citng them all would take too long, but I am also deeplygrateful to all the students and seniors from the University of Oviedo for all these interesting meetings, making the wednesday morning special each week.

Thank you also to my (PhD) coworkers who fought alongside me. Pedro, Nicolò, Celia, Clara, Andrea and Pablo, you made the students office a better place to work in and I wish you all the best, even though I am sure you will achieve great things in the future.

Of course, I would especially like to thank my parents Philippe and Chantal, and my brothers Antoine and Lucas for giving me the opportunity to come and live in Spain for so long. I never could have reached this far without your presence and support that I felt daily, even more than a thousand kilometers away. Thank you to Inés as well: you are the reason why I came to Spain in the first place and I can now say this was one of the best decisions I ever made. You know how I feel about you, you brought me more than you can imagine and I will never forget all these great moments we spent together. Even though one chapter ends, all I hope now is to find a way to stay in touch with you and your family for many more years.



# Acronyms used

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

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

# Contents

<b>1 Objects reconstruction</b>	<b>1</b>
1.1 Particle Flow (PF) algorithm . . . . .	1
1.2 Primary Vertex (PV) definition . . . . .	2
1.3 Leptons reconstruction . . . . .	3
1.3.1 Muons . . . . .	4
1.3.2 Electrons . . . . .	6
1.4 Jets reconstruction . . . . .	8
1.4.1 B-tagging . . . . .	10
1.5 Missing Transverse Energy (MET) . . . . .	11
1.6 Top reconstruction . . . . .	12
1.6.1 Numerical and analytical top reconstruction . . . . .	13
1.6.2 Top reconstruction with additional Dark Matter (DM) . . . . .	15
1.6.3 Top reconstruction in practice . . . . .	16
<b>2 Data, signals and backgrounds</b>	<b>19</b>
2.1 The Monte Carlo (MC) simulation method . . . . .	19
2.2 Files format . . . . .	23
2.3 Analysis code . . . . .	24
2.4 Data samples . . . . .	24
2.5 Signal samples . . . . .	24
2.6 Backgrounds prediction . . . . .	25
2.6.1 Top production . . . . .	26
2.6.2 Drell-Yan estimation . . . . .	28
2.6.3 $t\bar{t} + V$ . . . . .	29
2.6.4 Non prompt contamination . . . . .	29
2.6.5 Smaller backgrounds . . . . .	34
2.6.6 Weights and corrections applied . . . . .	35
<b>3 Event selection</b>	<b>37</b>
3.1 Objects selection . . . . .	37

3.1.1	Triggers selection . . . . .	37
3.1.2	Electrons selection . . . . .	39
3.1.3	Muons selection . . . . .	40
3.1.4	Jet selection . . . . .	42
3.2	Signals regions . . . . .	43
3.2.1	$t/\bar{t}$ +DM region . . . . .	44
3.2.2	$t\bar{t}$ +DM region . . . . .	44
3.3	Control regions . . . . .	47
3.3.1	Drell-Yan (DY) control region . . . . .	47
3.3.2	Top control region . . . . .	47
3.3.3	Same sign control region . . . . .	47
3.4	Background-signal discrimination . . . . .	51
3.4.1	Discriminating variables . . . . .	51
3.4.2	Neural network . . . . .	52
<b>4</b>	<b>Results and interpretations</b>	<b>53</b>
4.1	Systematics and uncertainties . . . . .	53
4.2	Results . . . . .	53
<b>5</b>	<b>Conclusions</b>	<b>55</b>
5.1	Future prospects . . . . .	55
<b>Appendices</b>		<b>57</b>
<b>A</b>	<b>Samples used</b>	<b>59</b>
A.1	Data samples . . . . .	59
A.2	Signal samples . . . . .	59
A.3	Backgrounds samples . . . . .	59
<b>B</b>	<b>Neural network optimization</b>	<b>67</b>
<b>Bibliography</b>		<b>73</b>



# Chapter 1

## Objects reconstruction

As we just saw, the Compact Muon Solenoid (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 1.1. 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 1.3), the jets (Section 1.4), the MET (Section 1.5) and the top reconstruction (Section 1.6) of the different  $pp$  collisions recorded.

### 1.1 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 Large Hadron Collider (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 ??: 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 1.1. Basically, the different kind of particles produced are going to interact with different parts of the detector and combining the information given by the tracker and the rest of the subsystems then allows to unequivocally identify each particle. This is usually done in a specific order in order to be as efficient as possible:

1. First of all, the most energetic **Primary Vertex (PV)** is identified by taking into account the Pile up (PU) and by assigning the different tracks to the different  $pp$  collisions happening during a single bunch-crossing.  
All the particles originating from less energetic PV or from a secondary vertex of interaction are typically ignored and the corresponding hits in the tracker left by such particles can therefore be removed, leaving less hits available for the clustering algorithm later on, allowing for a more efficient reconstruction of the following objects.
2. Then, **muons** are the easiest particles to identify since they are at first order the only particles leaving many hits in the muon chambers placed on the outside of the detector. Each muon identified is associated to its track in the tracker, where all the hits matching a muon can therefore be removed to simplify the following reconstruction steps.
3. **Electrons** do have a charge, so they are visible by the tracker and by the Electromagnetic Calorimeter (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 Hadronic Calorimeter (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.

## 1.2 Primary Vertex (PV) definition

Different kinds of vertices originating from a single  $pp$  collision can usually be defined, as shown in Figure 1.2: 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

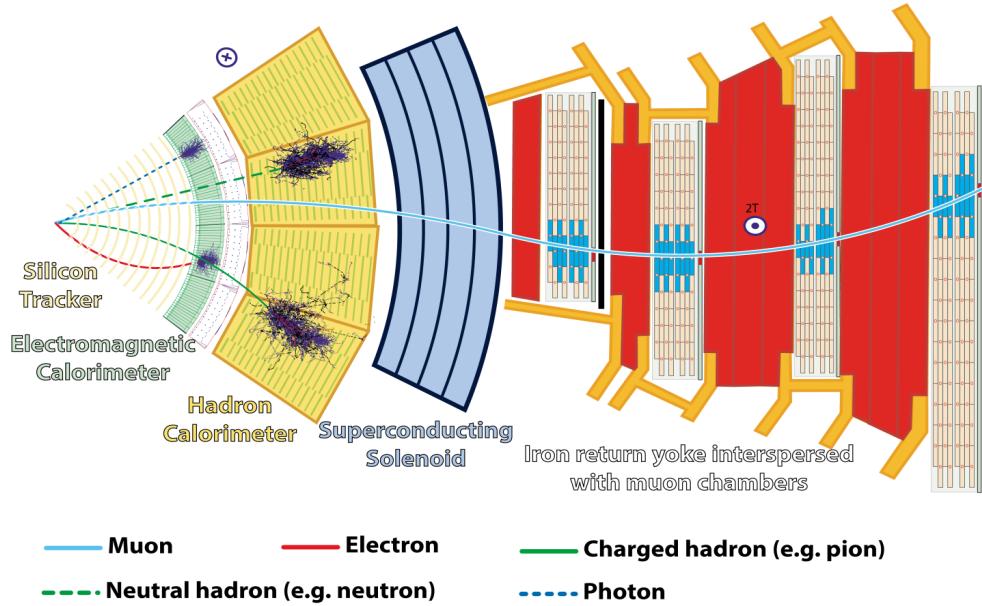


Figure 1.1: Transverse section of CMS showing the different tracks expected by different kind of particles in the detector.

largest  $p_T^2$  summed over all the physics objects of the event is then assumed to be the PV, as it is considered to be the origin of the most interesting  $pp$  collision from which many different tracks are emitted.

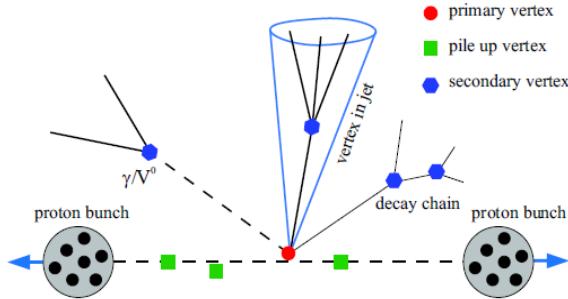


Figure 1.2: Different kind of vertices typically observed in a  $pp$  collision in the LHC.

### 1.3 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.

### 1.3.1 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 Drift tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (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.

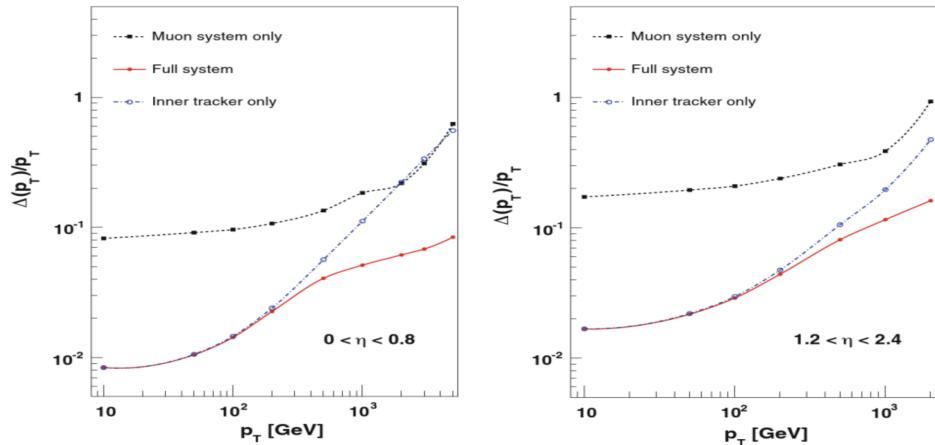


Figure 1.3: Muon  $p_T$  resolution obtained in simulation in the barrel (on the left) and endcap (on the right) for different kind of reconstructed muons [98].

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  $(\eta, \phi)$  plane, as shown in Figure 1.4) 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.

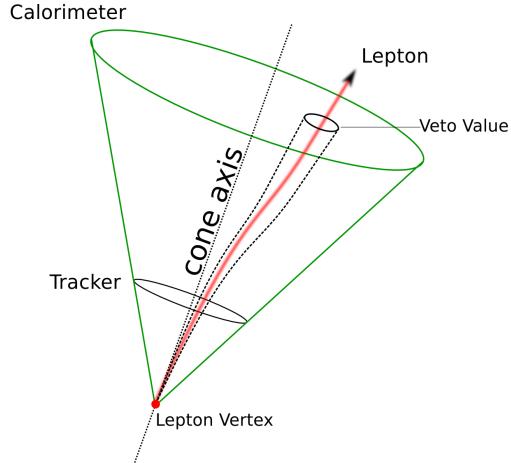


Figure 1.4: Lepton isolation cone typically used to enhance the prompt leptons purity.

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.

### 1.3.2 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  $\phi$  axis because of the deviation implied by the solenoid) then need to be collected as well and associated to the correct electron of the event.

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

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

MeV in the endcaps).

The excited contiguous crystals found are then grouped into clusters, themselves considered candidates for the final global cluster, the SC, if their energy is higher than another given threshold ( $E_{\text{cluster}} > 350$  MeV in the barrel,  $E_{\text{cluster}}^T > 1$  GeV in the endcaps) [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  $\chi^2$  penalty to each distribution, depending for example on the number of missing hits. This fit is also able to take into account sudden changes in the curvature radius caused by an eventual bremsstrahlung photon emission.

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

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

This complete electron workflow explained here in a simplified way can be summarized in Figure 1.5.

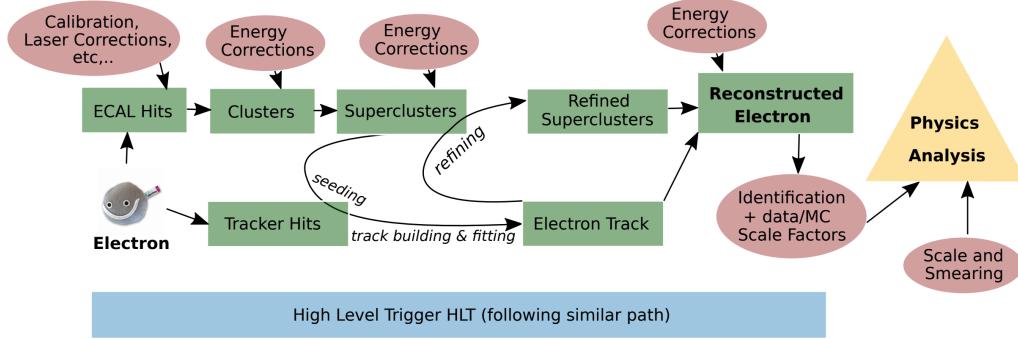


Figure 1.5: Schematic representation of the full electron reconstruction workflow in CMS [100].

## 1.4 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 1.6.

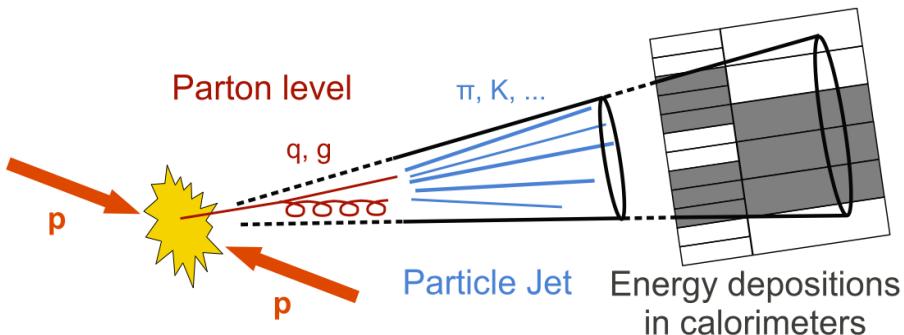


Figure 1.6: Schematic representation of the typical development of a jet within the CMS detector.

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 1.1, these distances

can be computed using different variables such as  $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i + \phi_j)^2$ , the distance between  $i$  and  $j$  in the  $(\eta, \phi)$  space, the  $p_T^2$  of each particle and  $R$ , 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.

$$\begin{cases} d_{ij} = \min \left( \frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \frac{\Delta R_{ij}^2}{R^2} \\ d_{iB} = \frac{1}{p_{T,i}^2} \end{cases} \quad (1.1)$$

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 4.1.

The efficiency of the PF algorithm for jet identification and reconstruction has been checked using simulation, as shown in Figure 1.7. 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  $p_T$ .

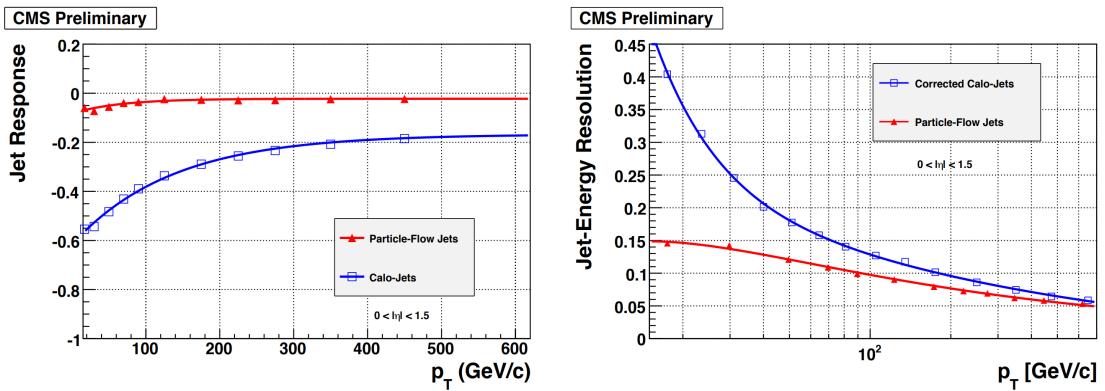


Figure 1.7: Comparison of the jet response (on the left) and jet energy resolution (on the right) for dijets simulated events in the barrel for jets reconstructed using only the calorimeters (in blue) and jet candidates from the PF algorithm (in red) [102].

### 1.4.1 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 3.

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 1.8 ; 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.

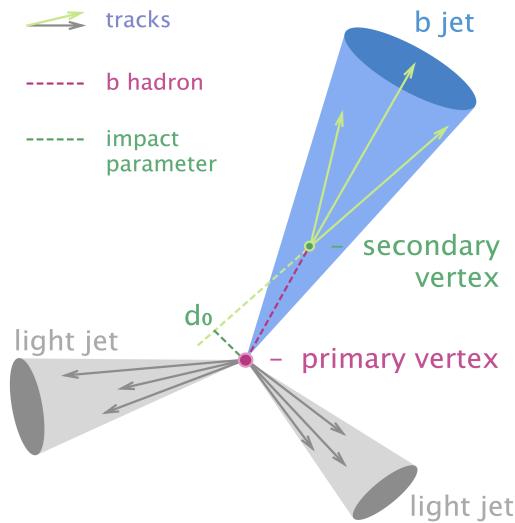


Figure 1.8: Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex.

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 1.9, 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.

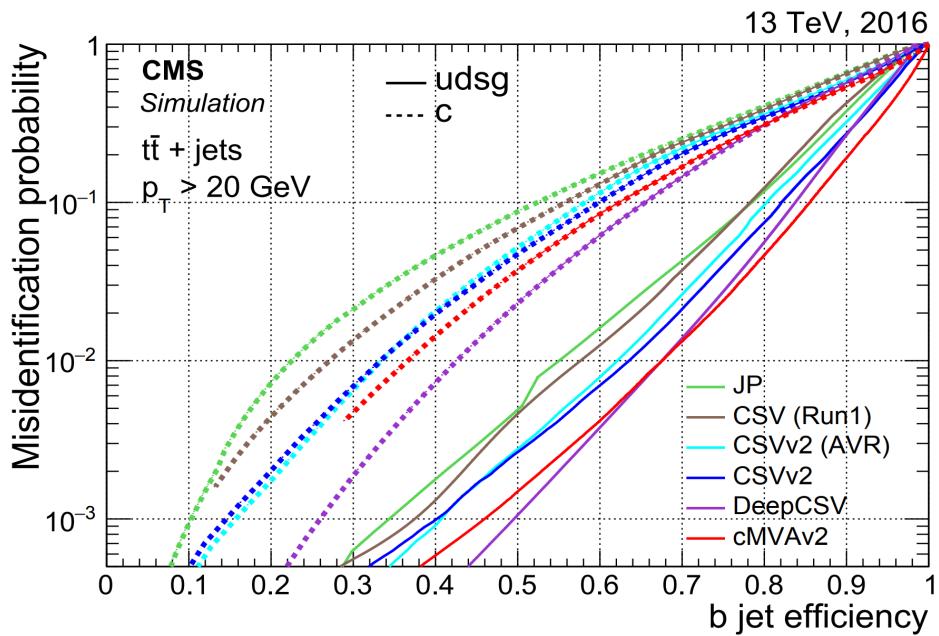


Figure 1.9: B-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [103].

## 1.5 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 1.10:

- 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, Standard Model (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 1.2 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

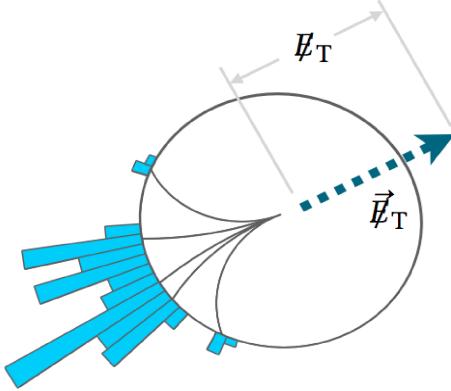


Figure 1.10: Schematic representation of the Missing Transverse Energy (MET).

searching for new Beyond the Standard Model (BSM) physics, which is not expected to interact with the detector.

$$\vec{p}_T^{\text{miss}} = \cancel{E}_T = - \sum_j \vec{p}_{T,j} \quad (1.2)$$

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 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 1.11.

## 1.6 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 2.6.1 is still an extremely important part of this analysis. The reconstruction of the top quark from channels containing leptons is typically quite challenging, given the fact that neutrinos are not directly observed, meaning that the sum of neutrino momenta

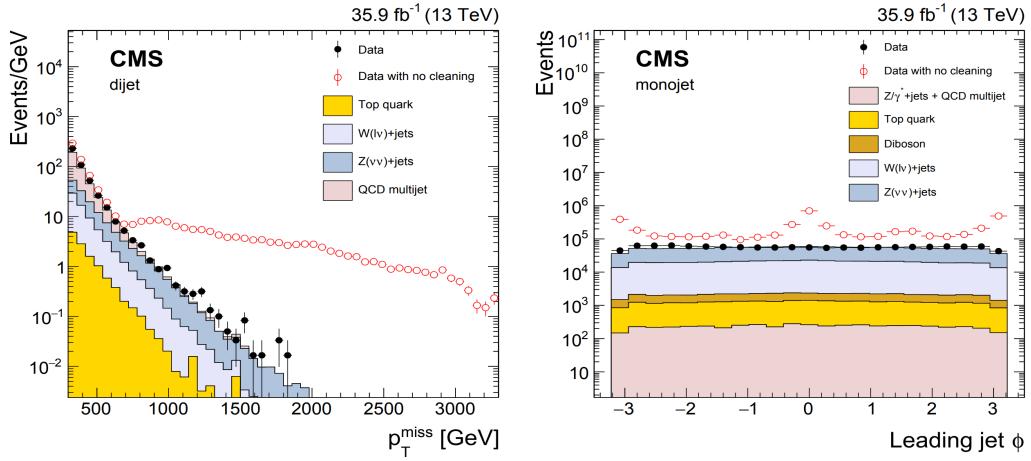


Figure 1.11: MET and jet  $\phi$  distributions with and without MET filters applied [104].

can only be inferred from the total momentum imbalance which frequently has a bad resolution. Additionally, the determination of the individual momenta of each neutrino require advanced computation techniques that will be briefly introduced in this section.

As previously explained, because of the high mass of the top quark, it is expected to decay before reaching the detector, almost 100% of the time into a  $b\bar{b}$  pair of quarks, which can be identified using the b-tagging method explained in Section 1.4.1, 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)).

### 1.6.1 Numerical and analytical top reconstruction

Two main methods exist in order to solve this reconstruction problem, now shortly described:

- The **Sonnenschein numerical method** first of all relies on the kinematics of the system and on the expression of the four-momenta of the different particles involved in the top quark decay chains, as expressed in Equations 1.3a to 1.3c, if we assume that the MET of the event is coming only from the two neutrinos produced.

$$\begin{cases} p_x^{\text{miss}} = p_{\nu_x} + p_{\bar{\nu}_x} \\ p_y^{\text{miss}} = p_{\nu_y} + p_{\bar{\nu}_y} \end{cases} \quad (1.3a)$$

$$\begin{cases} m_{W^+}^2 = (E_{l+} + E_\nu)^2 - (\vec{p}_{l+} + \vec{p}_\nu)^2 \\ m_{W^-}^2 = (E_{l-} + E_{\bar{\nu}})^2 - (\vec{p}_{l-} + \vec{p}_{\bar{\nu}})^2 \end{cases} \quad (1.3b)$$

$$\begin{cases} m_t^2 = (E_b + E_{l+} + E_\nu)^2 - (\vec{p}_b + \vec{p}_{l+} + \vec{p}_\nu)^2 \\ m_{\bar{t}}^2 = (E_{\bar{b}} + E_{l-} + E_{\bar{\nu}})^2 - (\vec{p}_{\bar{b}} + \vec{p}_{l-} + \vec{p}_{\bar{\nu}})^2 \end{cases} \quad (1.3c)$$

In this case, 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 [105], leading to a quartic equation in one of the unknowns, analytically solvable but quite ambiguous given the variable number of solutions of such equation.

- The **Betchart analytical method** on the other hand is able to describe the decay  $t \rightarrow b + l + \nu$  using a geometric approach. This method was chosen in this analysis because it offers the following advantages:
  - With this method, the invariant mass constraints from the top quark and the W boson are both exact and do not suffer the same kind of ambiguity as observed previously.
  - As we will now see, the solution set for each neutrino momentum in this case in an ellipse that can be described precisely. This provides us with a more precise measurement of the  $p_T$  of the mediator of the interaction, while also giving us information about the precision of this measurement.
  - The results obtained here can be useful for other event topologies featuring similar kinematic constraints as well.

Basically, this methods relies on two observations constraining the geometrical shape of the W boson momentum vector. First of all, the decay of top quark constrains this vector to an ellipsoidal surface of revolution about an axis coincident with the bottom quark momentum. The decay of the W boson itself on the other hand additionally constrain this vector to another ellipsoidal surface of revolution about an axis matching the momentum of the resulting charged lepton. The W boson momentum vector will then be defined by the intersection of the surfaces given by these two constraints, resulting in an ellipse in the phase space. The neutrino momentum vector can then be expressed as a translation of this ellipse, using a parametric expression in the laboratory coordinate system.

In the two neutrinos final state, it is then possible to show that the elliptical solution sets for the neutrino momenta ( $\nu_\perp, \bar{\nu}_\perp$ ) respective to the two top quarks decaying to leptons are given by Equation 1.4 [106], where  $N_\perp$  is the solution ellipse in the transverse place.

$$\begin{cases} \nu_\perp^T N_\perp \nu_\perp = 0 \\ \bar{\nu}_\perp^T \bar{N}_\perp \bar{\nu}_\perp = 0 \end{cases} \quad (1.4)$$

Given the fact that the measured components ( $\not{x}, \not{y}$ ) of the MET are the sum of the  $\nu_\perp$  and  $\bar{\nu}_\perp$  components, they can be related by Equation 1.5.

$$\bar{\nu}_\perp = \begin{pmatrix} -1 & 0 & \not{x} \\ 0 & -1 & \not{y} \\ 0 & 0 & 1 \end{pmatrix} \nu_\perp \equiv \Gamma \nu_\perp \quad (1.5)$$

The solutions for the momenta of the neutrinos will then be given by the intersections of these two ellipses giving either zero, two or four solution pairs ( $\mathbf{p}_\nu, \mathbf{p}_{\bar{\nu}}$ ), as shown in Figure 1.12. If the two ellipses do not intersect, a  $\chi^2$  method can be used to check the compatibility between the solution obtained and the standard  $t\bar{t}$  process, the best solution being defined as the point of closest approach between the ellipses.

So far we have seen two methods able to reconstruct the individual momentum of both neutrinos. However, both these methods usually assume that the MET is only coming from these neutrinos and this assumption is no longer verified for our signals, for which the DM particles produced

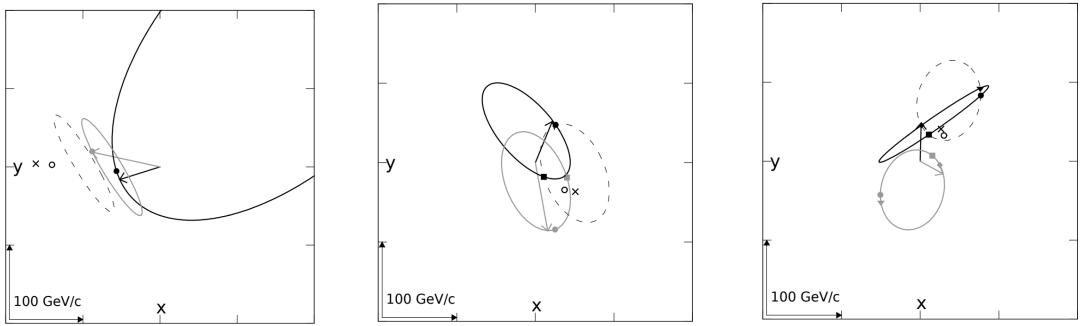


Figure 1.12: Three events constraining the neutrino and antineutrino momenta (black and grey arrows, respectively) resulting in 0 (on the left), 2 (on the center) or 4 (on the right) solutions. The dashed ellipses is obtained by using the additional constraint according to which measured MET is equal to the sum of neutrino transverse momenta [106].

will contribute by a significant amount to the global MET, resulting in a reconstruction efficiency of only  $\sim 75\%$ , compared to the  $\sim 96\%$  obtained considering only the standard  $t\bar{t}$  process. The method then needs to be slightly adapted to our particular case, as we will now see.

### 1.6.2 Top reconstruction with additional DM

In this particular case, including an additional contribution  $\phi_{\perp} = (\phi_x, \phi_y, 0)^T$  to the MET is equivalent to slightly modifying the Equation 1.5 to obtain Equation 1.6.

$$\bar{\nu}_{\perp} = \begin{pmatrix} \bar{\nu}_x \\ \bar{\nu}_y \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & \not{x} \\ 0 & -1 & \not{y} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_x + \phi_x \\ \nu_y + \phi_y \\ 1 \end{pmatrix} \equiv \Gamma(\nu_{\perp} + \phi_{\perp}) \quad (1.6)$$

By substituting these values into Equation 1.4, we then get a new relation between the two ellipses for the neutrino,  $N_{\perp}$ , and for the antineutrino,  $\bar{N}_{\perp}$ , given by Equation 1.7.

$$\nu_{\perp}^T \Gamma^T \bar{N}_{\perp} \Gamma \nu_{\perp} + \nu_{\perp}^T \Gamma^T \bar{N}_{\perp} \Gamma \phi_{\perp} + \phi_{\perp}^T \Gamma^T \bar{N}_{\perp} \Gamma \nu_{\perp} + \phi_{\perp}^T \Gamma^T \bar{N}_{\perp} \Gamma \phi_{\perp} = 0 \quad (1.7)$$

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

- First of all, the so-called **dark  $p_T$**  is a variable estimating the value of the  $p_T$  of the eventual mediator of the interaction. Indeed, as we just saw, if the system considered does not admit any solution, then there is no intersection between the ellipses formed in the MET phase space. However, in this case, the distance between the ellipses gives us a way to approximate by how much we miss a perfect standard  $t\bar{t}$  reconstruction, and it is actually possible to show that distance between the center of both ellipses is related to the  $p_T$  of the mediator of the interaction [107]. This method being completely analytical, estimated the distance between the ellipses is trivial and is giving us the first background-signal discriminating variable that

will be developed later in Section 3.4.

- The second interesting variable is the so-called **overlapping factor**  $R$ , defined in Equation 1.8, where  $l_1$  and  $l_2$  are the two sizes of the ellipses measured along the axis joining their centers, and  $d$  is the distance between these centers.

$$R = \frac{l_1 + l_2}{d} \quad (1.8)$$

By its definition, this factor is able to take into account not only the distance between these two ellipses but also their respective sizes and this is extremely important since, at the end of the day, we want an event having small ellipses far away from each other to have a much higher weight than an event featuring two incredibly large ellipses and therefore less significative when trying to distinguish the background and signal processes. Both these extreme cases are represented in Figure 1.13.

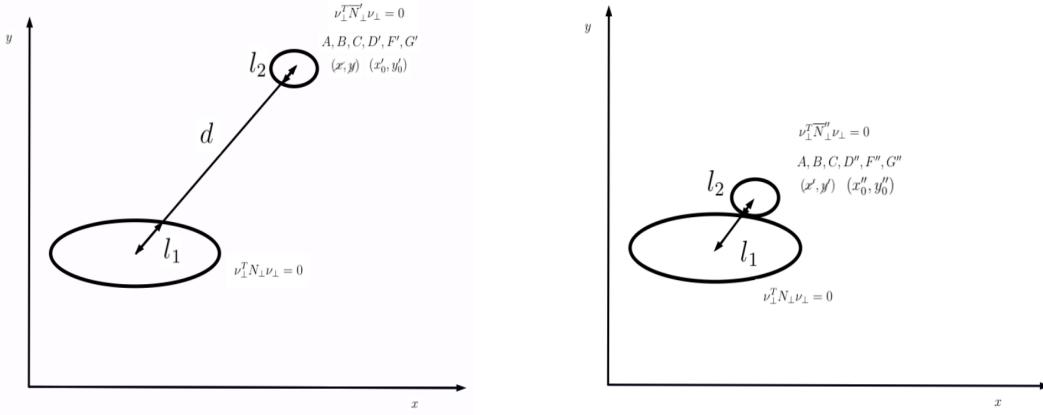


Figure 1.13: Schematic representation of the two extreme cases that can be observed when defining the overlapping factor: at the reconstruction of a system with (on the left) and without (on the right) the presence of DM [107].

### 1.6.3 Top reconstruction in practice

Even though the equations seem quite well defined and solvable, several complications quickly appear when solving this problem in practice. 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, by categorizing the event depending on the number of b-jets observed:

- 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 3.
- 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. When taking into account all these possible combinations, a reconstruction efficiency of  $\sim 96\%$  has been achieved when considering SM  $t\bar{t}$  MC samples.

This efficiency can be slightly increased by performing a **smearing** method, taking into account imperfectly measured kinematics by repeating the reconstruction 100 times for each event, by changing each time the jets and leptons energy and directions within their respective uncertainties.

Add details about the smearing once known



# Chapter 2

## Data, signals and backgrounds

In order to find a possible hint of the production of DM in the LHC collisions considering our signal models of interest, briefly described in Section ??, the data collected needs to be compared with Monte Carlo (MC) simulations produced in a central way for each SM process. Indeed, any deviation of the data observed with respect to what we expect to see, obtained from these MC simulations, might be the sign of some BSM physics. All of the steps needed to mathematically simulate the  $pp$  collisions of the LHC and to take into account the effect of the detector on the particles produced will first of all be introduced in Section 2.1.

Then, the different formats of files available to perform the analysis and the code used will be briefly introduced in Sections 2.2 and 2.3 and the different data samples collected during the Run II of operation of the LHC will be then detailed in Section 2.4, while the signal models and samples considered in this particular analysis along with the MC samples used for the simulation of the different backgrounds will be introduced in Sections 2.5 and 2.6 respectively.

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

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

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

To make the generation of such simulations a bit easier, the description of a typical  $pp$  collision can usually be divided into several steps that will now be described, as shown with the color code used in Figure 2.1. The typical approximations used to make this kind of simulation possible from the computational point of view will also be briefly introduced at this point.

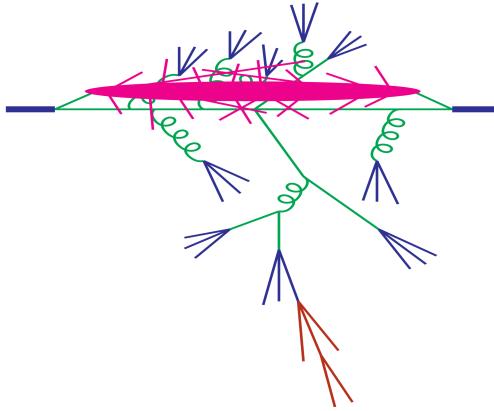


Figure 2.1: Structure of a  $pp$  collision and different steps of the MC simulation used by the event generators, such as the parton shower (in green), the UE (in pink), the hadronization (in blue) and the decay of unstable particles (in red) [108].

### Hard scattering

A typical  $pp$  collision at a center of mass energy  $\sqrt{s}$  is usually described by an event generator as the interaction between a parton  $i$  coming from one proton with a parton  $j$  coming from the other, leading to the production of a final state  $A$ , made out of  $n$  different particles. The total cross section of such process can be expressed with Equation 2.1 [109].

$$\sigma_A(s) = \sum_{i,j} \iint dx_1 dx_2 f_i(x_1, \mu^2) f_j(x_2, \mu^2) \hat{\sigma}_{ij \rightarrow A}(\hat{s}, \mu^2) \quad (2.1)$$

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

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

Many algorithms have been developed in order to select a hard process  $ij \rightarrow A$  and determine its kinematics by solving this equation using different methods. The samples used in this work have actually been produced at different orders and by different hard scattering generators, such as MADGRAPH [111] (at LO) and POWHEG [112] and MC@NLO [113] (at NLO).

### Parton showers

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

a large acceleration to the partons involved, which then tend to emit Quantum ChromoDynamics (QCD) radiation under the forms of gluons, just like accelerated electric charges do by emitting photons. However, the gluons emitted do have a color charge and can therefore emit further radiation until reaching such a low energy that they are able to form colourless hadrons, as discussed in Section 1.4. This process typically leads to the creation of the so-called **parton showers**, approximate higher-order real-emission corrections to the hard scattering, that need to be simulated by the event generators as well since they are an important part of the kinematics of the collision.

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

## Underlying Event (UE)

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

These secondary collisions typically lead to the production of extra hadrons and therefore need to be simulated as well by events generators, usually by distributing the partons of the incoming protons in an area of  $1\text{fm}^2$ : an increased UE will be obtained when the so-called impact parameter, the distance between the parton and the centre of this area, is decreased, making the collision mostly central and almost head-on [114]. The UE is typically well simulated using softwares such as Herwig [115] and PYTHIA [116]. The spectrum for the generation of some variables in a top enriched sample can be found in Figure 2.2.

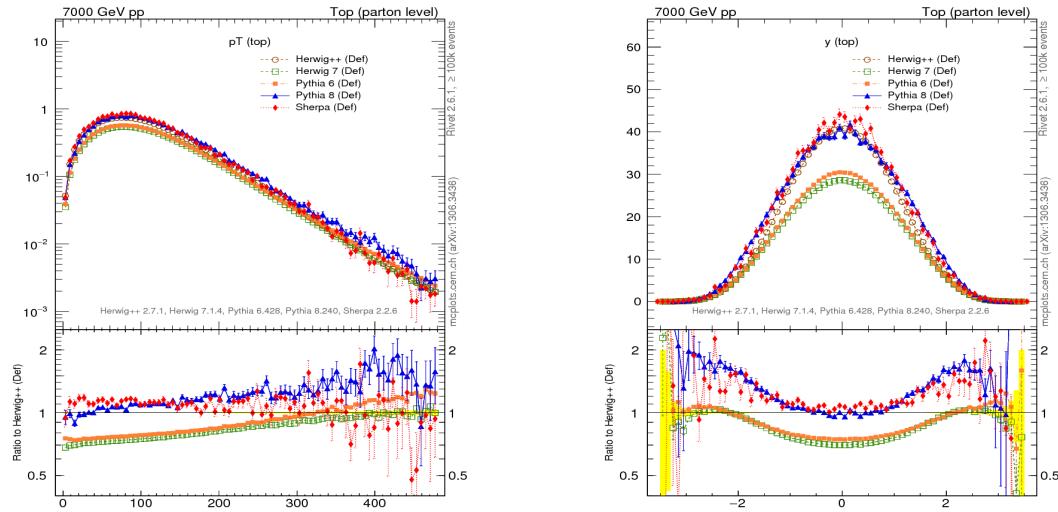


Figure 2.2: Top  $p_T$  (on the left) and rapidity (on the right) distributions obtained using different MC generators [117].

## **Hadronization**

Once all the primary and secondary collisions simulated, it is time for the event generators to simulate the **hadronization** and binding processes of the different coloured partons emitted into colourless hadrons, as explained in Section 1.4. This hadronization process happen at low energies, when the perturbation theory becomes invalid and the dynamics enter a non-perturbative phase, which leads to the formation of the observed final-state hadrons. Non-perturbative calculations then have to be used by the event generators in order to simulate this effect.

## **Unstable particle decays**

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

## **Detector simulation**

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

This is typically done by the GEANT4 software [118], able to model different effects, such as:

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

This modeling accounts for all the cracks and for the disposition of the subsystems inside of the CMS detector. This software is for example able to model the interaction of the electrons with the tracker, responsible for the emission of bremsstrahlung photons, as explained in Section 1.3.2.

The results of the comparison between the output of two different versions of the GEANT4 software and prototypes of the CMS calorimeter in the test beam facility at European Council for Nuclear Research (CERN) lead to comparable results, as shown in Figure 2.3.

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

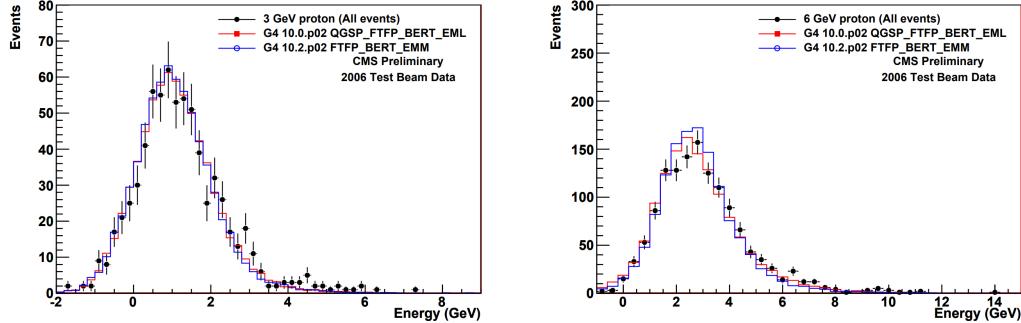


Figure 2.3: Proton energy distribution at 3 (on the left) and 6 (on the right) GeV compared for the test beam data (in black) and two different GEANT4 versions [119].

## 2.2 Files format

Once recorded (or simulated), the data (or MC) still needs to go under a complete post-processing in order to change its format and reduce the total size of the samples to be considered in the analyses. Different types of analysis are expected to need different levels of data reduction, so the data is usually accessible at different levels [120]:

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

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

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

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

## 2.3 Analysis code

The code used for the event generation, simulation and reconstruction is the version 10\_4\_X of the official software of the CMS collaboration, called CMSSW [123]. This software contains the CMS Event Data Model (EDM) which is able to describe every event as a C++ object containing all the RAW and reconstructed information related to the collision. These objects are stored using the ROOT file format [124], an analysis package written in C++.

Once all the different samples produced centrally up to the nanoAOD stage, another framework was put in place in order to do a post-processing of such samples, by selecting objects interesting for different dileptonic analyses, reducing therefore even more the size of the samples to be considered by selecting only events having 2 tight leptons. This selection will be detailed in Chapter 3. This *Latino* framework, written in python, is common to several different analyses and has been developed by tens of different people over the past few years, providing several tools to produce samples, read the files, apply different corrections to the MC samples and produce the histograms needed to perform a search such as this one.

## 2.4 Data samples

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

During this period, an integrated luminosity of  $35.9 \pm 0.9 \text{ fb}^{-1}$  (2016) [125],  $41.5 \pm 1.0 \text{ fb}^{-1}$  (2017) [126] and  $59.7 \pm 1.5 \text{ fb}^{-1}$  (2018) [127] has been collected, resulting in a total dataset of  $137.1 \pm 2.0 \text{ fb}^{-1}$  recorded by the CMS detector and ready to be analyzed. This data has been obtained by combining a set of single and double lepton triggers that will be described in Section 3.1.1 by taking care of avoiding any eventual double counting due to events present in different triggers. All the data samples considered for this analysis are listed in Appendix A.1.

## 2.5 Signal samples

Two different sets of MC signal samples have been produced centrally for this analysis, corresponding to the  $t+DM$  (along with its different channels of production) and to the  $t\bar{t}+DM$  signal. Different mass points were produced in both cases, considering different dark matter masses, from 1 to 55 GeV, and different scalar or pseudoscalar mediator masses, ranging from 50 to 500 GeV. In this context, 17 different mass points have been produced for each mediator, as listed in Appendix A.2.

The impact on the kinematics (in this particular case, on the spectrum of PUPPI MET) of these different mass points available can be observed in Figures 2.4 and 2.5, considering scalar and pseudoscalar mediators respectively. As expected from Table A.4, we can see on one hand first of all in these figures that the higher the mediator mass is, the lowest is its spectrum because of the lower cross section associated to the model. On the other hand, we can also observe that the mass of the DM itself is not impacting in any way this spectrum.

To be updated once the  $t + DM$  samples are actually available

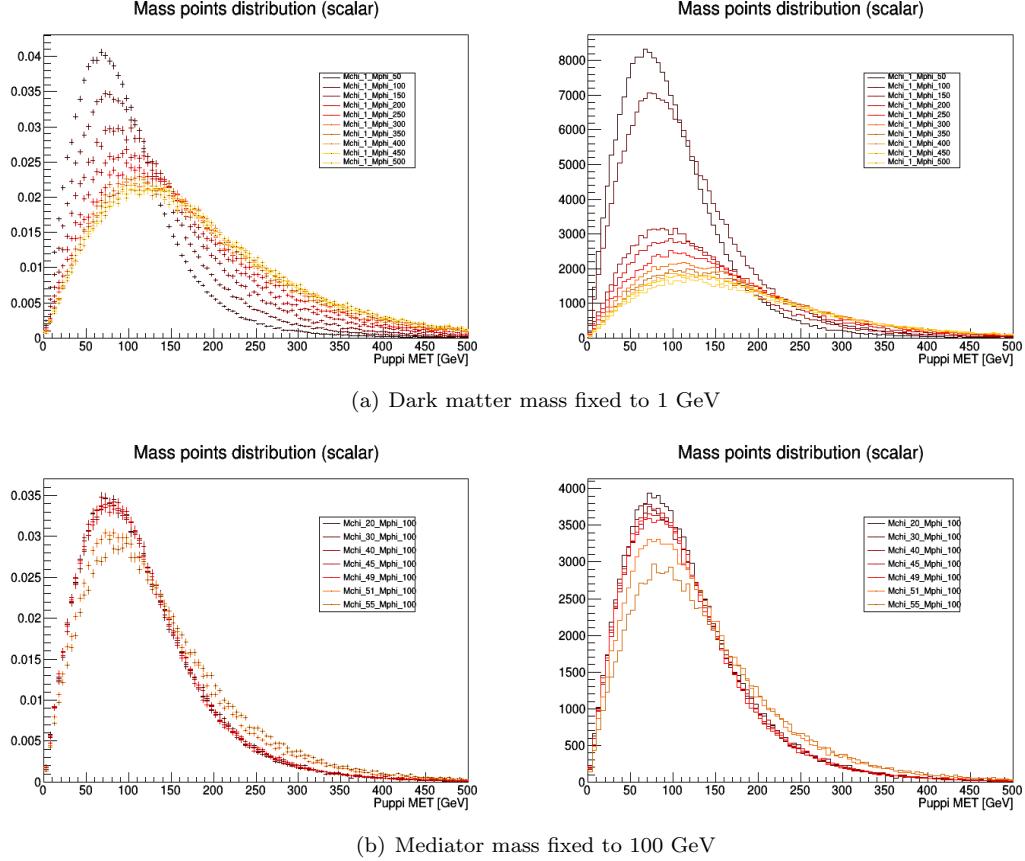


Figure 2.4: PUPPI MET spectrum considering different **scalar** mediator (on the top) and dark matter (on the bottom) masses, with (on the left) and without normalization (on the right).

## 2.6 Backgrounds prediction

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

- The major background for the  $t\bar{t}$ +DM analysis, the SM  $t\bar{t}$ , kinematically really close to the signal searched for (Section 2.6.1).
- On the other hand, the major background for the  $t/\bar{t}$ +DM analysis is the single top production, which has an even higher cross section than the  $t\bar{t}$  (Section 2.6.1).
- Then, mainly because of its huge cross section at 13 TeV, as shown in Figure 2.6, the DY process is usually important to consider. Even though quite reduced in the Signal Regions (SRs) because of the cuts applied, a specific Control Region (CR) will be dedicated to check this background, taken directly from MC (Section 2.6.2).
- The  $t\bar{t} + V$  ( $t\bar{t} + Z$  and  $t\bar{t} + W$ ) may have a kinematics even closer to our  $t\bar{t}$ +DM signal than the  $t\bar{t}$  process and is therefore extremely important in our signal regions, even though its low cross section does limit its impact (Section 2.6.3).
- Finally, the non-prompt background is another important piece of this analysis mainly because of the particular data-driven method used to compute them (Section 2.6.4).

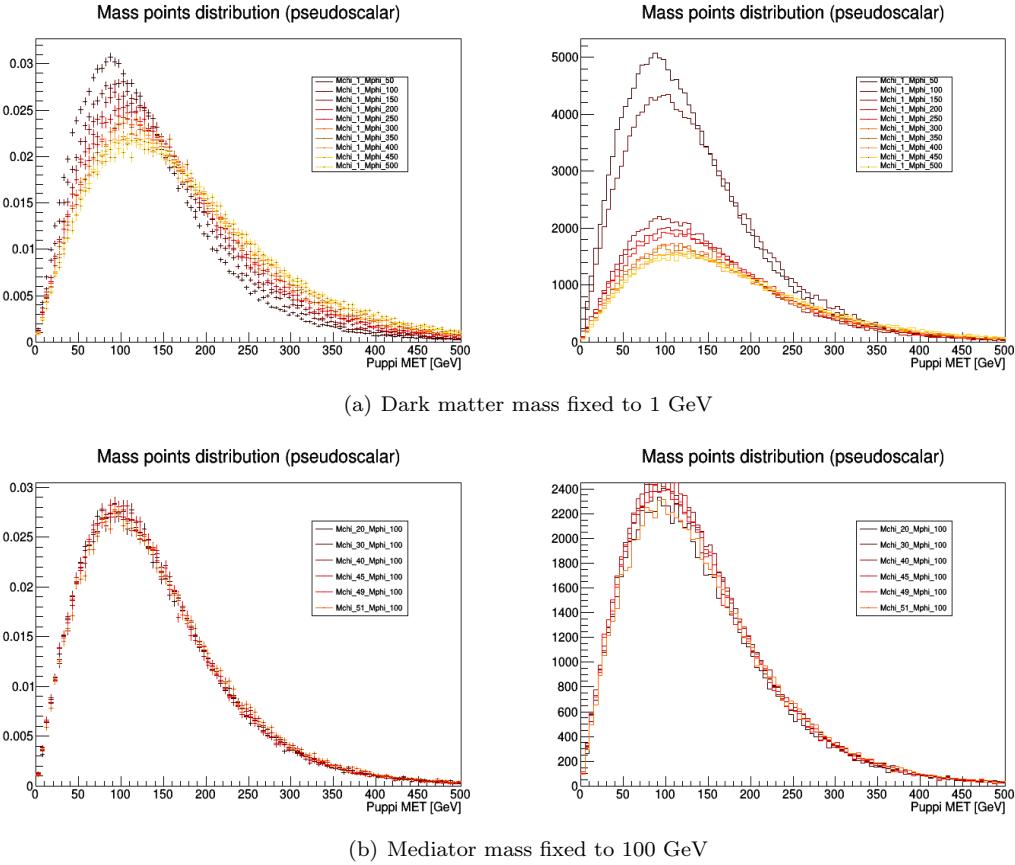


Figure 2.5: PUPPI MET spectrum considering different **pseudoscalar** mediator (on the top) and dark matter (on the bottom) masses, with (on the left) and without normalization (on the right).

Add percentage of each background in the SR once known

Finally, some smaller backgrounds will be introduced in Section 2.6.5, such as the diboson and triboson production, and the weights and corrections applied to all these MC samples will be detailed in Section 2.6.6.

### 2.6.1 Top production

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

They also have a kinematics quite close to the one expected for our signal: some additional MET is expected because of the production of a pair of DM particles, which might lead to a distribution  $\phi$  distribution as well, but other than that achieving some discrimination between these processes is quite complex. This process mostly relies on the use of Machine Learning (ML) techniques and will be fully detailed in Section 3.4.

Typically, given the importance of such processes, a specific CR to check for the validity of the CR of such processes is defined, as explained in Section 3.3.

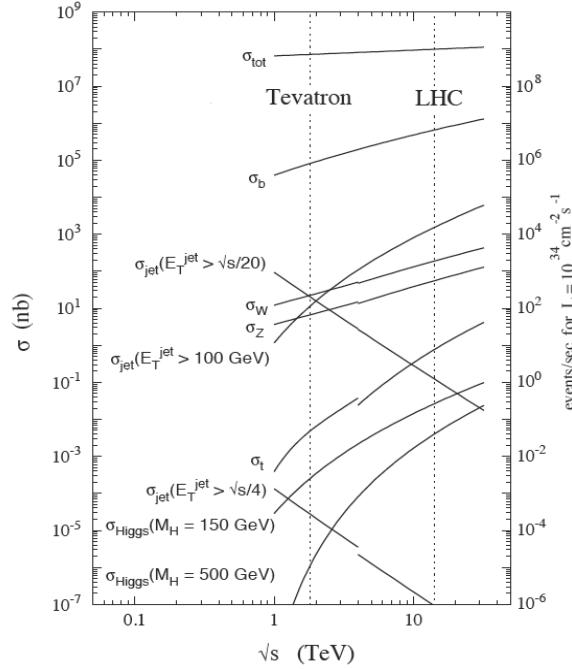


Figure 2.6: Production cross section of the most common SM processes considering different center of mass energies, such as the 13 TeV of the LHC.

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

Different Feynman diagrams contribute to this process at Leading Order (LO) in a hadron collider, as shown in Figure 2.7. This background is estimated directly from MC and, due to its importance, dedicated CRs have been designed in which to check for the data/MC agreement.

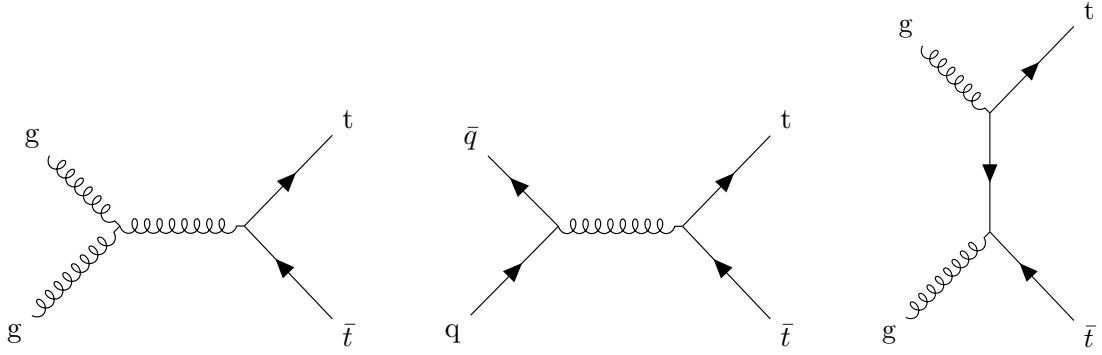


Figure 2.7: Main feynman diagrams for the production of the SM  $t\bar{t}$  process.

### Single top

Different Feynman diagrams also account for this process in the s-channel (Figure 2.8), t-channel and tW production mode (Figure 2.9).

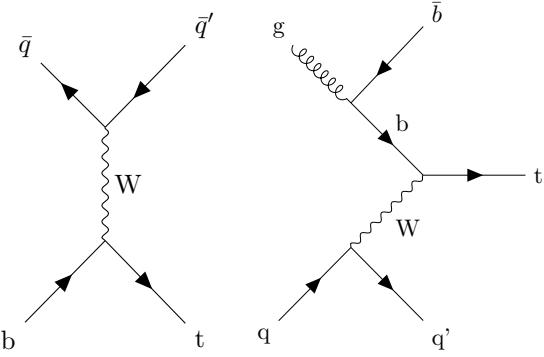


Figure 2.8: Feynman diagrams for s-channel production mode of a single top quarks.

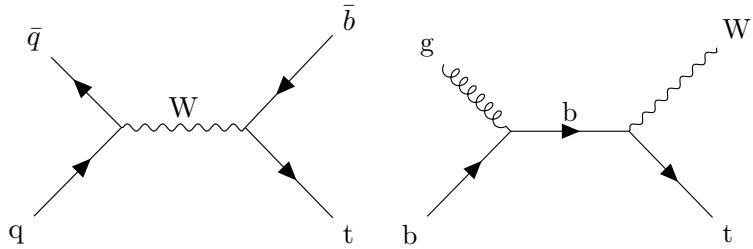


Figure 2.9: Feynman diagrams for t-channel (on the left) and tW (on the right) production modes of a single top quarks.

### Top decay

As previously mentioned, the top is the heaviest particle of the SM and is expected to decay inside of the beam pipe itself, usually into a bottom quark, giving us a b-jet, and a W boson; this boson can decay itself into different channels even though only its leptonic decay is considered in this particular case. The decay considered of the top/antitop produced is represented in Figure 2.10.

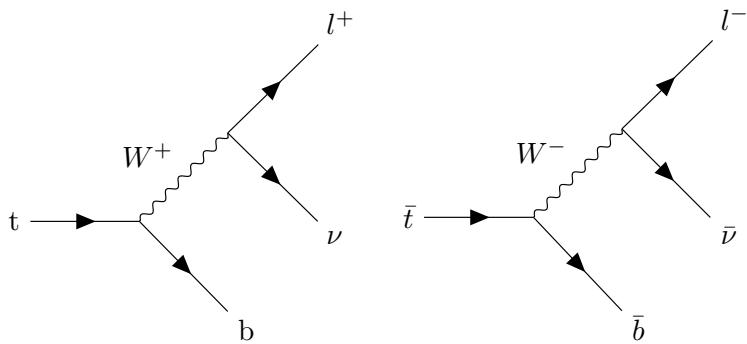


Figure 2.10: Feynman diagrams for the leptonic decay of the top (on the left) and antitop (on the right) quarks.

### 2.6.2 Drell-Yan estimation

As previously mentioned, most of the DY, produced through the Feynman diagram represented in Figure 2.11, is not expected to survive the selection applied to the analysis but because of the huge cross section of this process, two to three orders of magnitude larger than the production of top, the contribution of this process in the SRs is still quite important.

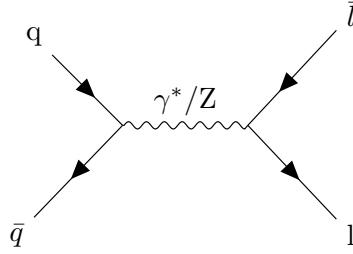


Figure 2.11: Feynman diagram for the DY process involving a virtual  $\gamma^*$  or Z boson.

This background is also estimated from MC as well and is being checked in a specific CR as well, as described in Section 3.3.

### 2.6.3 $t\bar{t} + V$

This background is coming from a usual  $t\bar{t}$  production along with an ISR or Final State Radiation (FSR) production of a W or Z boson, as shown in Figure 2.12. The contribution of this background is also taken directly from MC.

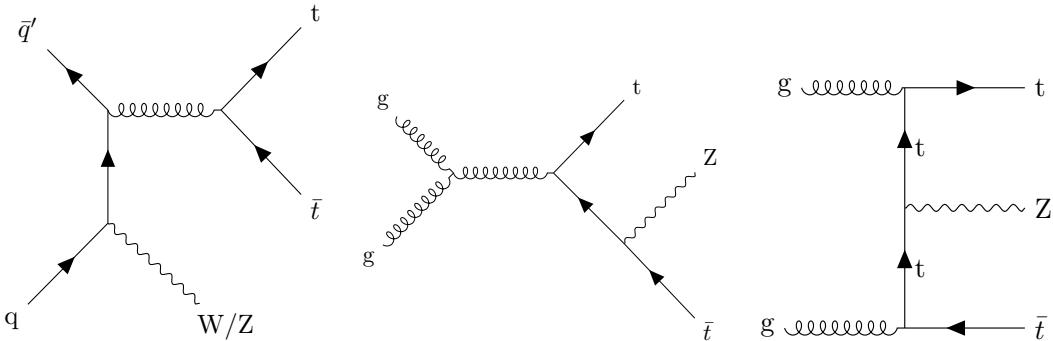


Figure 2.12: Possible Feynman diagrams for the Initial State Radiation (ISR)  $t\bar{t}$  with a W/Z boson (on the left) and for the production of an Final State Radiation (FSR)  $t\bar{t}Z$  (on the center and right).

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

### 2.6.4 Non prompt contamination

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

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

- First of all, a **prompt lepton** is defined as a real lepton, in the sense that the lepton is originating from the PV of a  $pp$  collision.

- The **Prompt Rate (PR)** is defined as the number of prompt leptons passing the tight selection criteria of the analysis over the number of leptons passing the loose selection criteria.
- On the other hand, by **fake** or **non-prompt lepton**, we usually refer to truly **fake leptons**, such as jets misidentified by the detector as leptons, as shown in Figure 2.13, and real leptons coming from eventual heavy flavor decays.
- The **Fake Rate (FR)** is then defined similarly to the prompt rate but considering this time fake leptons only for the tight-to-loose ratio. This ratio therefore corresponds to the probability for a fake lepton to be considered as a real lepton in the analysis.

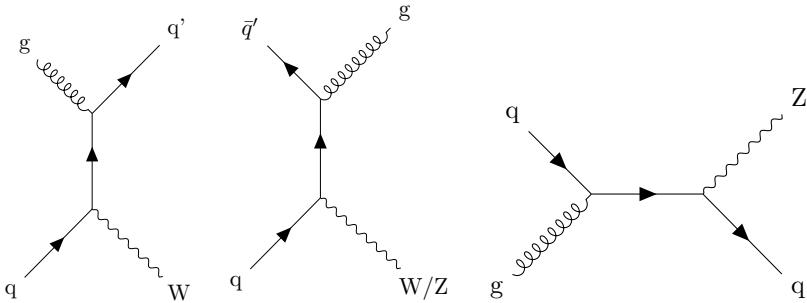


Figure 2.13: Possible Feynman diagrams for the production of a W/Z boson with a jet.

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

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

### Fake Rate (FR) computation

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

- |   |  |
|---|--|
| <ul style="list-style-type: none"> <li>• Exactly 1 lepton</li> <li>• <math>p_T &gt; 13</math> (10) GeV for <math>e</math> (<math>\mu</math>)</li> <li>• <math> \eta  &lt; 2.5</math> (2.4) for <math>e</math> (<math>\mu</math>)</li> </ul> | <ul style="list-style-type: none"> <li>• <math>mtw1 &lt; 20</math> GeV</li> <li>• PuppiMET <math>&lt; 20</math> GeV</li> <li>• PassJets</li> </ul> |
|---|--|

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

Using this method, the jet that fakes a lepton is actually the one recoiling against ( $\Delta R > 1.0$ ) the jet used for the systematics, as shown in Figure 2.14.

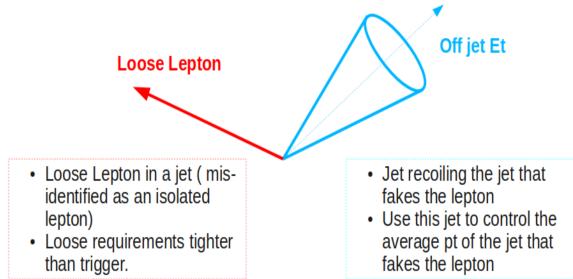


Figure 2.14: Schematic representation of the two jets used for the systematics and for the jet faking a lepton in the tight-to-loose datadriven method.

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

$$\text{Muon triggers} = \begin{cases} \text{HLT\_Mu8\_TrkIsoVVL (if } p_T < 20 \text{ GeV)} \\ \text{HLT\_Mu17\_TrkIsoVVL (if } p_T \geq 20 \text{ GeV)} \end{cases} \quad (2.2a)$$

$$\text{Electron triggers} = \begin{cases} \text{HLT\_Ele8\_CaloIdL\_TrackIdL\_IsoVL\_PFJet30 (if } p_T < 25 \text{ GeV)} \\ \text{HLT\_Ele23\_CaloIdL\_TrackIdL\_IsoVL\_PFJet30 (if } p_T \geq 25 \text{ GeV)} \end{cases} \quad (2.2b)$$

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

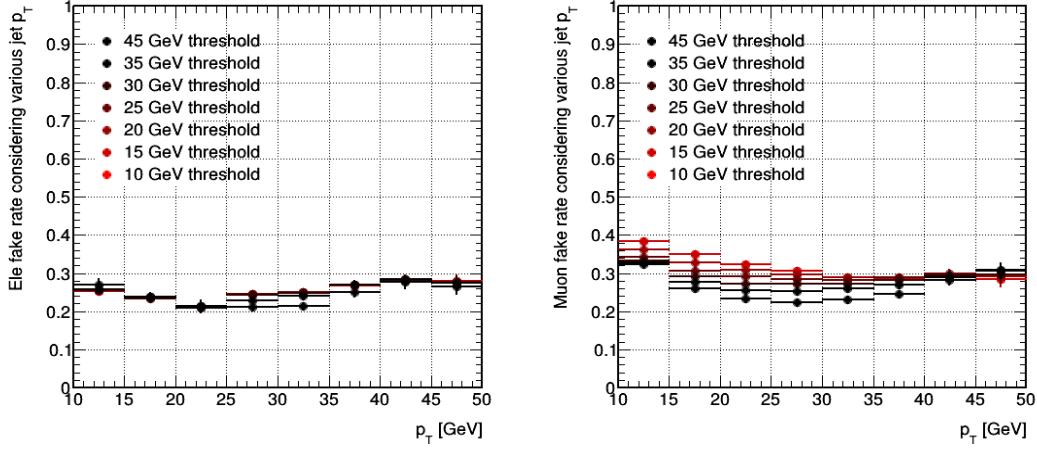
Since both the FR and PR heavily depend on the kinematics of the event and on the Working Point (WP) chosen for the leptons of the analysis, they are computed separately depending on the flavor of the lepton and 2D histograms (accounting for the  $p_T$  and  $\eta$  of the event) need to be created at this stage to calculate this factor, for a given input jet  $E_T$  threshold; 1D histograms corresponding to the projections of these 2D histograms along both their axes are also defined at this point, as shown in Figure 2.15.

### Prompt Rate (PR) computation

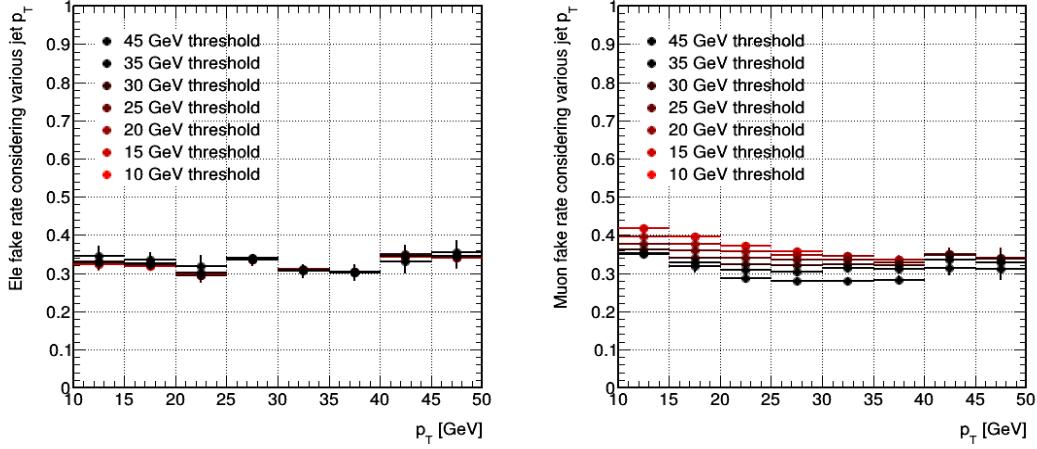
The PR, taking into account the real lepton contamination in the CR defined, is also important to calculate, even though the objects WP are usually chosen in such a way that this ratio is quite close to 1 and can therefore be ignored.

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

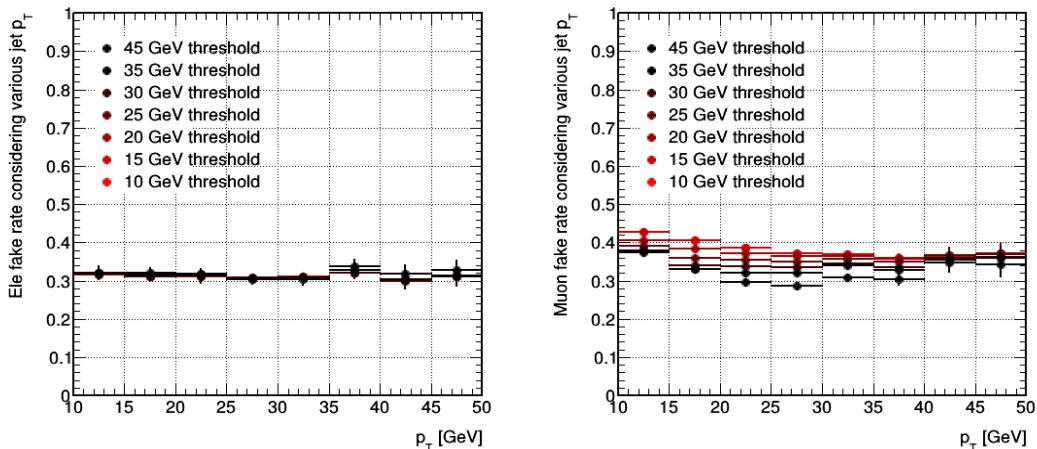
The results obtained in this case have been represented in Figure 2.16



(a) 2016 electron and muon fake rates

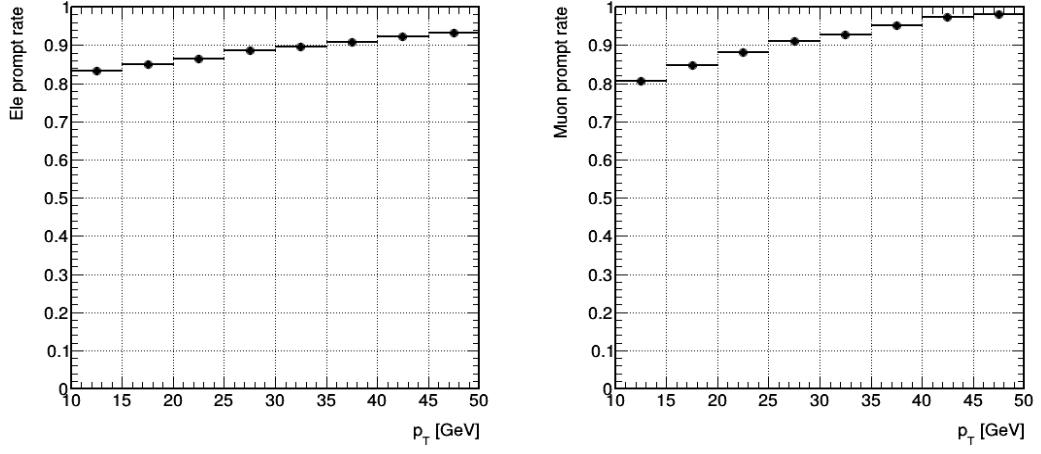


(b) 2017 electron and muon fake rates

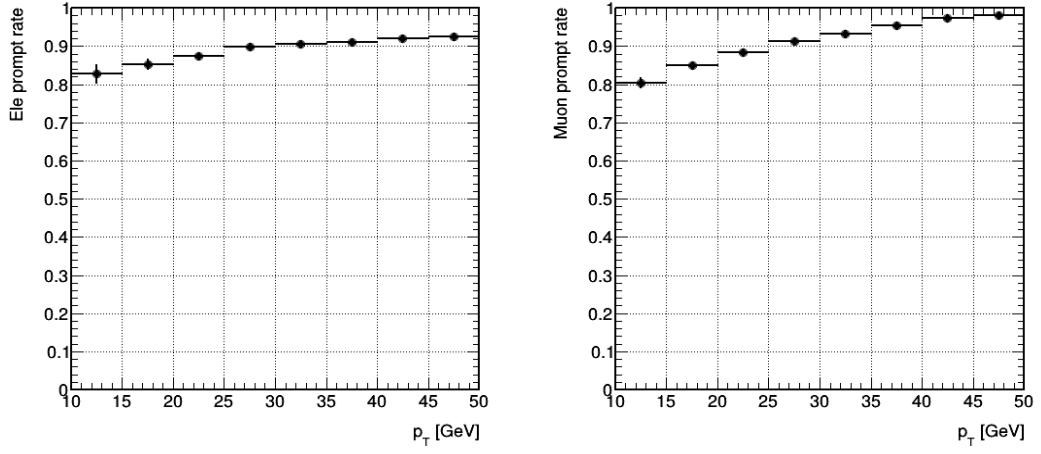


(c) 2018 electron and muon fake rates

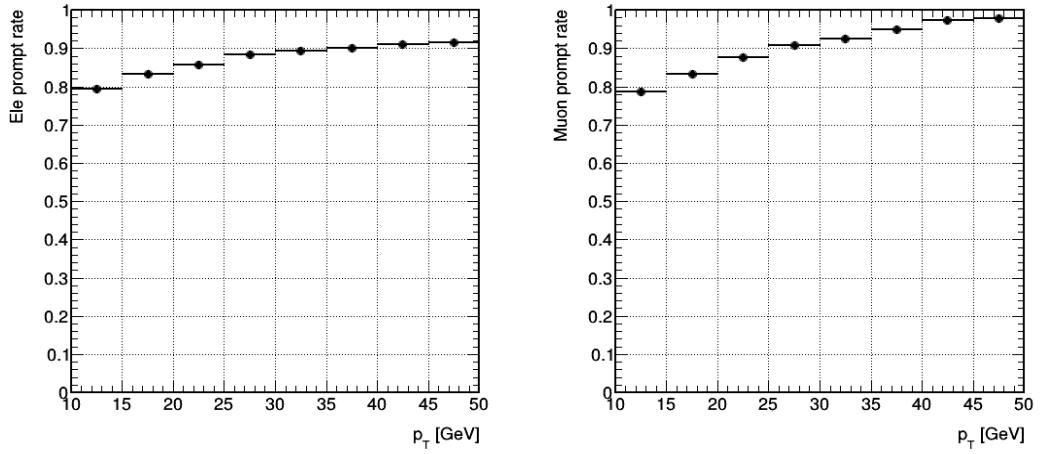
Figure 2.15: Electron (on the left) and muon (on the right) FR obtained in a QCD enriched region for different jet  $E_T$  thresholds for 2016, 2017 and 2018 with respect to the  $p_T$  of the lepton.



(a) 2016 electron and muon prompt rates



(b) 2017 electron and muon prompt rates



(c) 2018 electron and muon prompt rates

Figure 2.16: Electron (on the left) and muon (on the right) PR obtained in a Z+jets enriched region by a general tag and probe method for 2016, 2017 and 2018 with respect to the  $p_T$  of the lepton.

## Fake weight calculation

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

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

It is then possible to see in Equation 2.3 that, if  $p$  is the PR and  $f$  the FR previously calculated.

$$\left\{ \begin{array}{l} N_l = N_{pp} + N_{fp} + N_{ff} = N_{t2} + N_{t1} + N_{t0} \\ N_{t0} = (1-p)^2 N_{pp} + (1-p)(1-f)N_{fp} + (1-f)^2 N_{ff} \\ N_{t1} = 2p(1-p)N_{pp} + (f(1-p) + p(1-f))N_{fp} + 2f(1-f)^2 N_{ff} \\ N_{t2} = p^2 N_{pp} + pfN_{fp} + f^2 N_{ff} \end{array} \right. \quad (2.3)$$

These equations can be inverted in order to represent the unknowns with respect to the known variables, as shown in Equation 2.4, giving us a way to apply the weights previously calculated to this particular l2loose region.

$$\begin{pmatrix} N_{pp} \\ N_{fp} \\ N_{ff} \end{pmatrix} = \frac{f-p}{-(p-f)^3} \cdot \begin{pmatrix} f^2 & -f(1-f) & (1-f)^2 \\ -2fp & p(1-f) + f(1-p) & -2(1-p)(1-f) \\ p^2 & -p(1-p) & (1-p)^2 \end{pmatrix} \cdot \begin{pmatrix} N_{t0} \\ N_{t1} \\ N_{t2} \end{pmatrix} \quad (2.4)$$

## Same sign control region

Now that we have a way to estimate the kinematics and yields of the non-prompt background in a different region, a same sign CR enriched in fakes is usually defined in order to check this background. This will be done and detailed in Section 3.3.3 of this work.

### 2.6.5 Smaller backgrounds

Even though quite negligible, some additional backgrounds still need to be considered, such as the dibosons (WW, WZ and ZZ) and tribosons (WWW, WWZ, WZZ, ZZZ) productions, as shown in Figure 2.17. Backgrounds such as the  $W\gamma$  and Higgs related productions are also considered at this stage. All these smaller backgrounds are taken directly from MC.

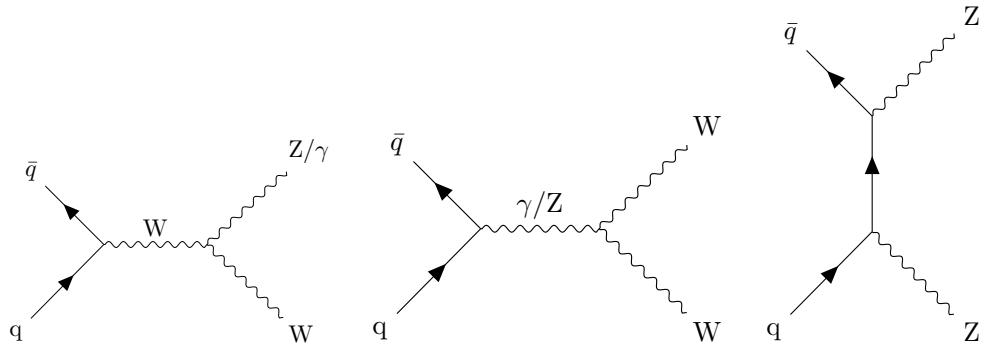


Figure 2.17: Possible Feynman diagrams for smaller backgrounds of this analysis: WW (on the left),  $W\gamma$  and  $WZ$  (on the center) and ZZ (on the right).

### 2.6.6 Weights and corrections applied

Several different weights and Scale Factors (SF) usually need to be applied to the different MC processes in order to account for several effects observed in data but not accounted for during the generation of the MC simulation, such as the efficiency of the selection of the different objects, which will be discussed in Chapter 3, usually directly provided by the different Physics Object Groups (POGs) for their own default objects definitions.

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

#### Prefiring corrections

In 2016 and 2017 an issue, known as the **prefiring**, causing highly energetic readout from jets, photons and electrons in the ECAL endcap to be assigned by the Level-1 Trigger (L1) to the previous bunch crossing was discovered. To make up for this difference, a weight  $(1 - x)$  is usually applied to all MC events, where  $x$  is the probability of an event to be prefire [128].

The effect this correction has on the data/MC agreement in a 2017 top enriched control region is shown in Figure 2.18.

#### Top $p_T$ reweighting

Previous studies of the  $t\bar{t}$  generator typically considered in the physics analysis predict a harder top quark  $p_T$  spectrum than the one observed in data [129].

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

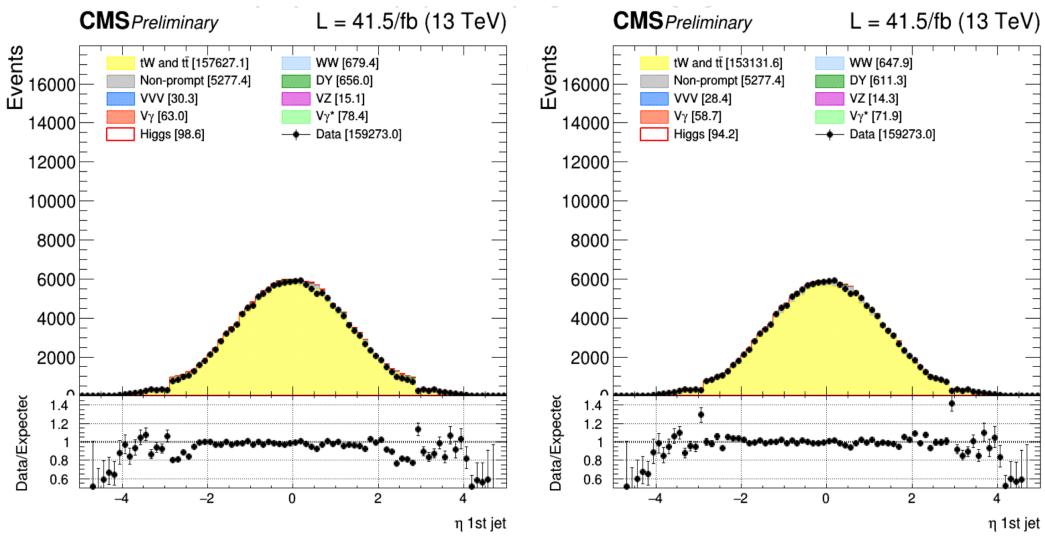


Figure 2.18: Data/MC agreement with (on the right) and without (on the left) prefiring corrections in a 2017 top control region.

# Chapter 3

## Event selection

This Chapter will be dedicated to the analysis itself, by defining first of all the different objects actually used in this case, along with the actual selection that has been applied to enhance the quality of such objects in this particular search in Section 3.1. Then, the different Signal Regions (SRs) defined in which a high purity of signal is expected are defined in Section 3.2 while all the different Control Regions (CRs) defined in order to check the behavior of the MC simulation performed for the major backgrounds on this analysis, such as the single top or SM  $t\bar{t}$  production, will be introduced in Section 3.3.

Finally will come a description about the different variables expected to naturally introduce some discrimination of the  $t/\bar{t}$  and  $t\bar{t}+{\rm DM}$  signals with respect to the different backgrounds in Section 3.4, along with a global description of the Machine Learning (ML) techniques employed in order to optimize the discriminating power of these variables in the best way possible.

### 3.1 Objects selection

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

First of all, the different triggers used to collect the data will be detailed in Section 3.1.1. Then, the leptons used in this analysis will be introduced in Sections 3.1.2 (for electrons) and 3.1.3 (for muons). Finally, given the nature of the DM signal searched for, a complete description of the jets selected in the analysis will be necessary and performed in Section 3.1.4.

#### 3.1.1 Triggers selection

The triggers, described in Section ??, and particularly the trigger paths chosen are an important part of each analysis since they will describe the kind of data that can be collected and therefore analyzed. The triggers used in this analysis for the years 2016, 2017 and 2018 can be found in Tables 3.1, 3.2 and 3.1 respectively.

Dataset	Run range	<b>High-Level Trigger (HLT) trigger path</b>
SingleMu	[273158,284044]	HLT_IsoMu24_v* HLT_IsoTkMu24_v*
SingleEle	[273158,284044]	HLT_Ele27_WPTight_Gsf_v* HLT_Ele25_eta2p1_WPTight_Gsf_v*
DoubleEG	[273158,284044]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
DoubleMu	[273158,281612] [281613,284044]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v* HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
MuonEG	[273158,278272] [278273,284044]	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 3.1: 2016 trigger paths considered for this analysis.

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

These triggers have been studied in order to make sure that they are efficient enough in the  $p_T$  region of the leptons of the analysis to avoid any undesired effect due to the turn-on of any trigger. These trigger efficiencies, calculated using a general tag and probe method and found for example for different runs of the 2017 data taking period in Figure 3.1 for a DoubleEG trigger, are then used to reweight the simulated samples.

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

Table 3.2: 2017 trigger paths considered for this analysis.

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

Table 3.3: 2018 trigger paths considered for this analysis.

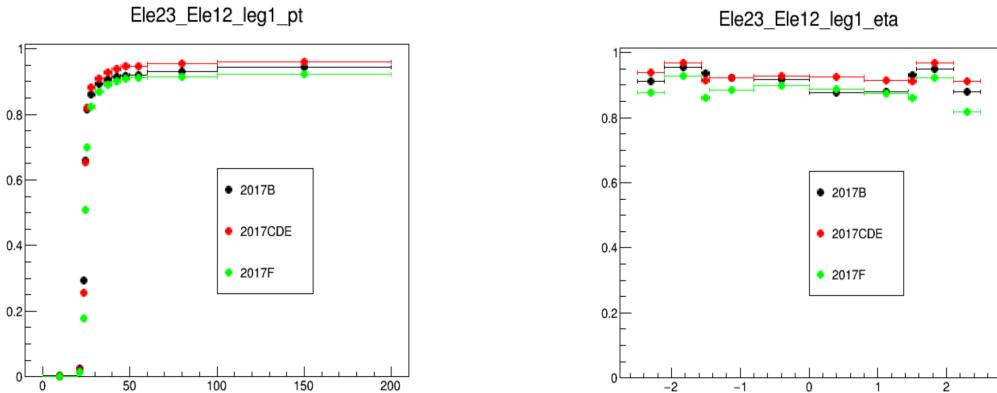


Figure 3.1: DoubleEG trigger efficiencies with respect to the  $p_T$  (on the left) and  $\eta$  (on the right), computed using a tag and probe method, for the 2017 data taking period.

### 3.1.2 Electrons selection

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

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

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

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

Two WPs are then given directly by the CMS EGamma POG [130], corresponding to an electron selection efficiency of 80 and 90% respectively. For this analysis, the tight POG *mva\_90p\_Iso2016* (for 2016) and *mvaFall17V1Iso\_WP90* (for 2017 and 2018) WPs have been chosen for the electron definition, along with additional quality cuts defined in Tables 3.4 and 3.5 (as previously mentioned, these cuts sometimes differ quite a bit depending on whether the electron interacts with the ECAL barrel ( $|\eta| < 1.479$ ) or one of the endcaps ( $|\eta| \geq 1.479$ )).

The electron efficiencies computed for this particular selection can be seen in Figure 3.2.

### 3.1.3 Muons selection

The selection applied to muons is mostly based on the Muon POG, providing references efficiencies for standard selection, recommendations for the tight selection [131], with a few additional cuts on both the impact parameters.

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

- The PF muon reconstructed is a global muon.

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

Table 3.4: Quality cuts applied to define a 2016 tight electron in this analysis.

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

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

- $p_T > 10$  GeV,  $|\eta| < 2.4$  and  $|d_z| < 0.1$  cm
- $|d_{xy}| < 0.01$  cm (if  $p_T < 20$  GeV) or  $|d_{xy}| < 0.02$  cm (if  $p_T \geq 20$  GeV)
- Tight muon isolation requirement ( $< 0.15$ ) with  $\Delta\beta$  correction and in a cone size  $\Delta R < 0.4$ , as defined in Equation 3.1, in order to reduce the number of muons coming from the hadronization process of bottom and charm quarks.

$$\text{ISO} = \frac{\sum p_T^{\text{ch. had. (PV)}} + \max(0, \sum E_T^{\text{neut. had.}} + \sum E_T^\gamma - 0.5 \times \sum p_T^{\text{ch. had. (PU)}})}{p_T(\mu)} \quad (3.1)$$

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

Table 3.5: Quality cuts applied to define a 2017/2018 tight electron in this analysis.

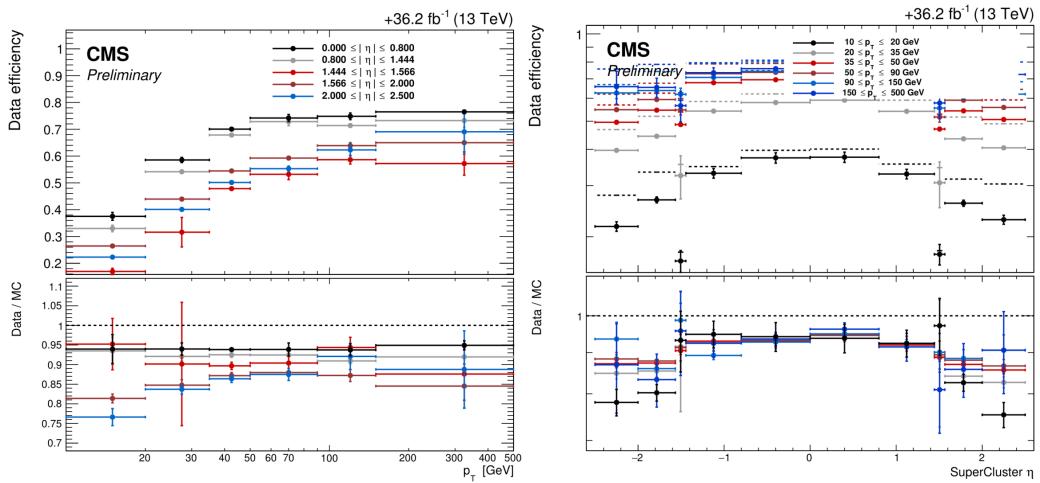


Figure 3.2: Tight electron efficiencies for this analysis, based on the Egamma POG `mva_90p_Iso2016` WP, for the data taking period of 2016.

This selection leads to efficiencies higher than 80% for all the muon momenta and pseudorapidities range, as shown in Figure 3.3.

### 3.1.4 Jet selection

Jets are an important part of this analysis as well given the final state searched for in this case. As explained in Section 1.4, the jets are clustered from the PF candidates using the anti-kT algorithm (with a typical distance parameter  $R = 0.4$ ).

A selection is then applied on these jets in order to separate noise from real jets. In this analysis, this selection follows the tight WP definition given by the CMS JET/MET POG [132], whose selection depends on the pseudorapidity of the jet and on the year of data taking, as shown in Tables 3.6 and 3.7. The tight WP has been chosen since it offers an efficiency higher than 98-99% for all the jets and a background rejection higher than 98% for jets having  $|\eta| < 3.0$ .

A slightly different selection (loose PU jet ID) is applied to jets having a  $p_T > 50$  GeV in order to reject jets coming from PU interactions in 2017 and 2018, as shown in Table 3.8.

The b-jets are then also select using the recommendations given by the B-Tagging and Vertexing POG [133]. The loose deepCSV b-tagging WP is used in this case, as explained in Section 1.4.1, for which the rate for misidentifying a light jet as a b-jet is around 10%. A jet is considered a b-jet

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

Color Definition

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

(a) 2016 muon efficiencies

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

Color Definition

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

(b) 2017 muon efficiencies

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

Color Definition

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

(c) 2018 muon efficiencies

Figure 3.3: Tight muon efficiencies for this analysis, based on the Muon POG tight WP with additional cuts for 2016, 2017 and 2018.

in this analysis if it passes the following cuts:

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

## 3.2 Signals regions

It is important to note that a strict **blinding policy** has been followed for this search, in order to avoid optimizing the analysis based on what has already seen. At first, the data available to be plotted in the following signal regions has therefore been limited to  $1 \text{ fb}^{-1}$  for each year, before the unblinding, allowing us to look at the whole Run II dataset.

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

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

Table 3.6: JET/MET POG tight WP for 2016.

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

Table 3.7: JET/MET POG tight WP for 2017 and 2018.

### 3.2.1 $t/\bar{t}$ +DM region

As shown in Figure 3.4

To be added once defined

### 3.2.2 $t\bar{t}$ +DM region

As shown in Figure 3.5

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

Table 3.8: JET/MET POG tight WP for 2017 and 2018, for jets having a  $p_T > 50$  GeV.

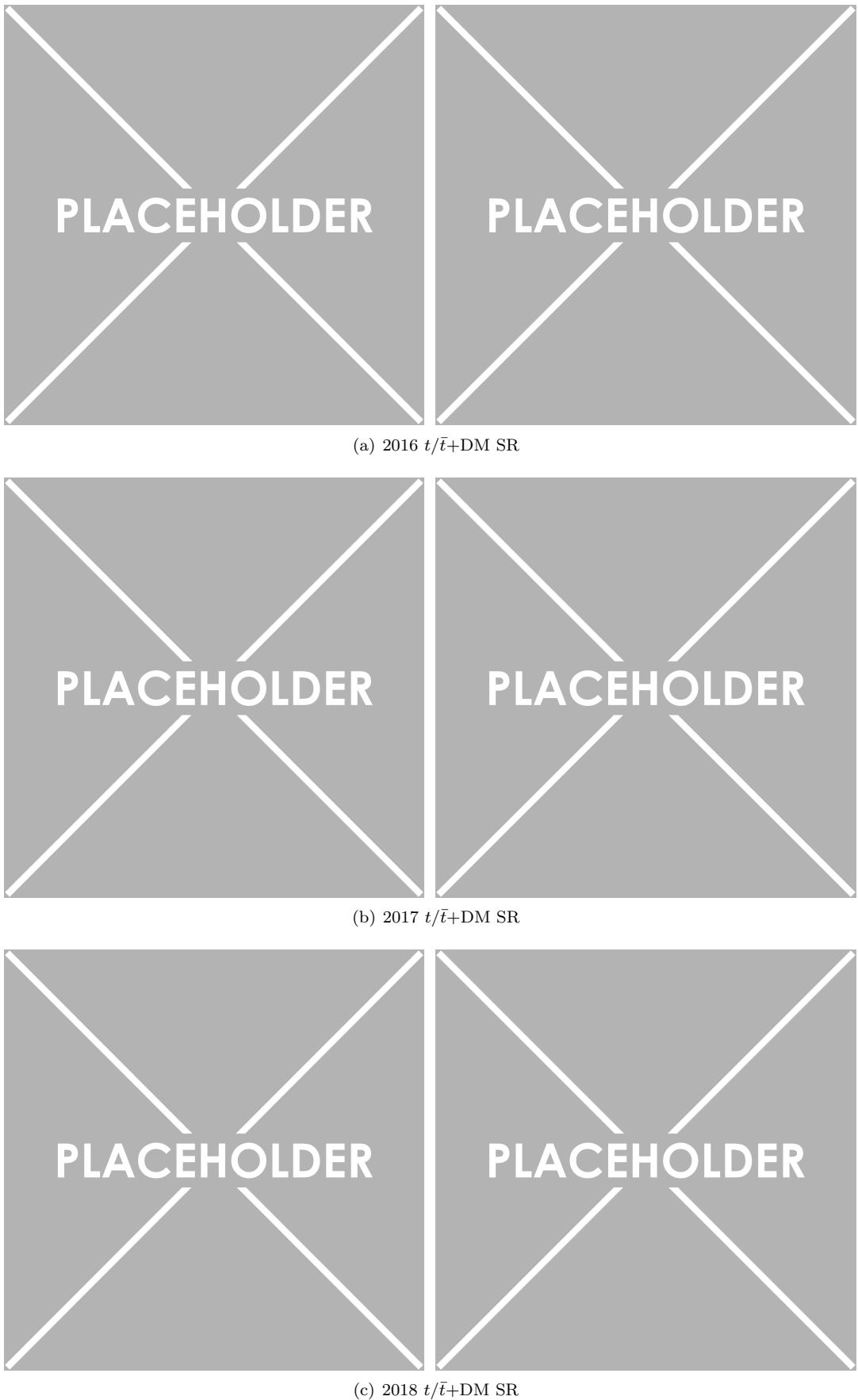


Figure 3.4: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the  $t/\bar{t}+\text{DM}$  process.

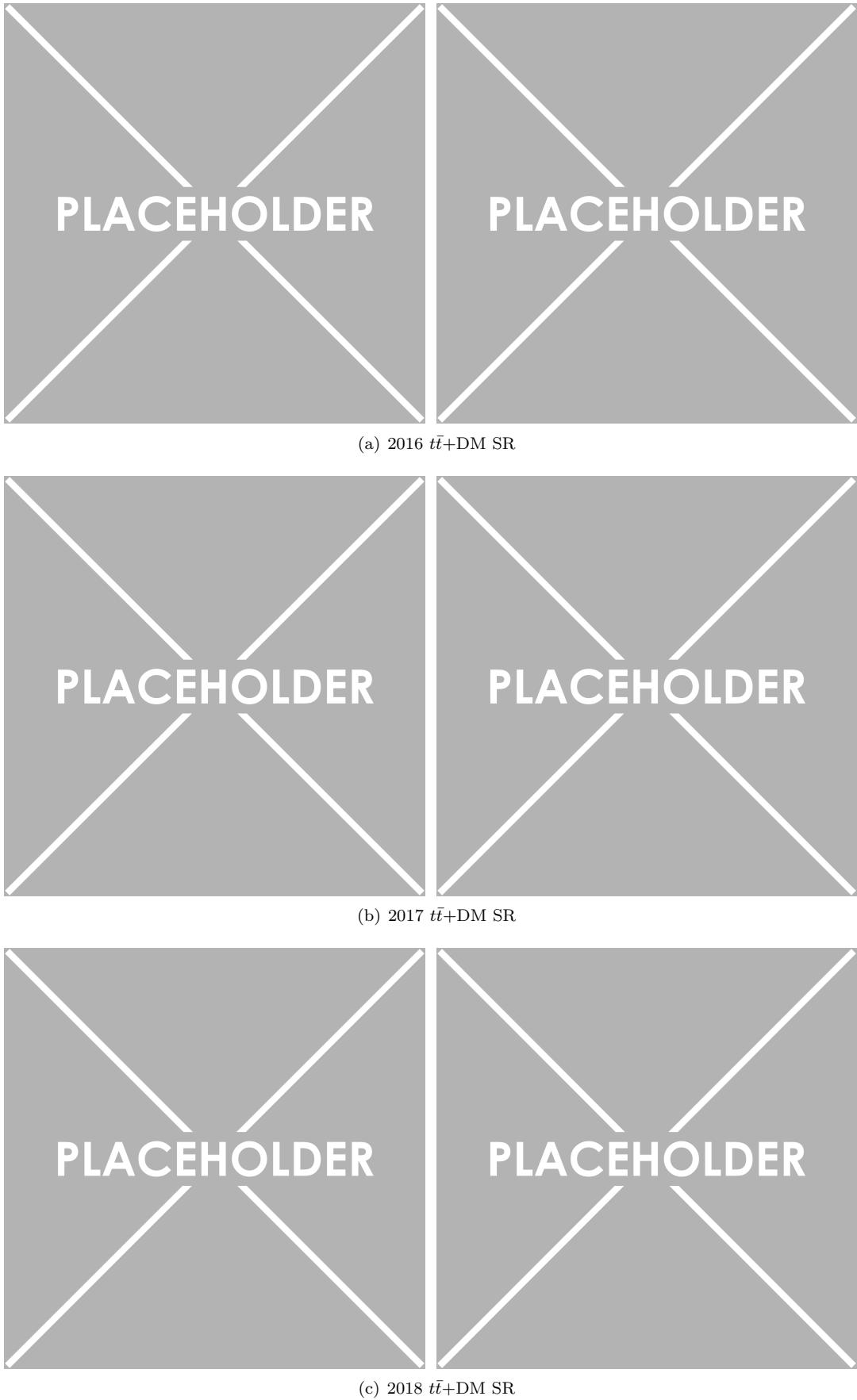


Figure 3.5: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the  $t\bar{t}$ +DM process.

To be added once defined

### 3.3 Control regions

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

In this context, the DY control region will be defined in Section 3.3.1 while the SM  $t\bar{t}$ , WW and ttV processes will be checked in the Sections 3.3.2, ?? and ?? respectively.

#### 3.3.1 DY control region

As shown in Figure 3.6

To be added once defined

#### 3.3.2 Top control region

As shown in Figure 3.7

To be added once defined

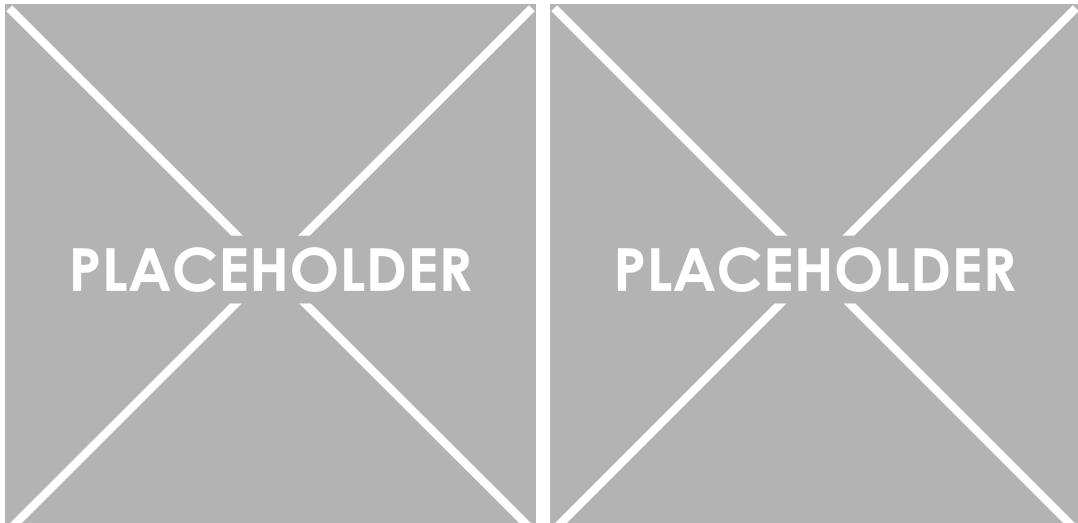
#### 3.3.3 Same sign control region

A same sign control region has also been defined in order to check the non-prompt background, calculated using a data-driven tight-to-loose method described in Section 2.6.4. This CR is defined with the following cuts:

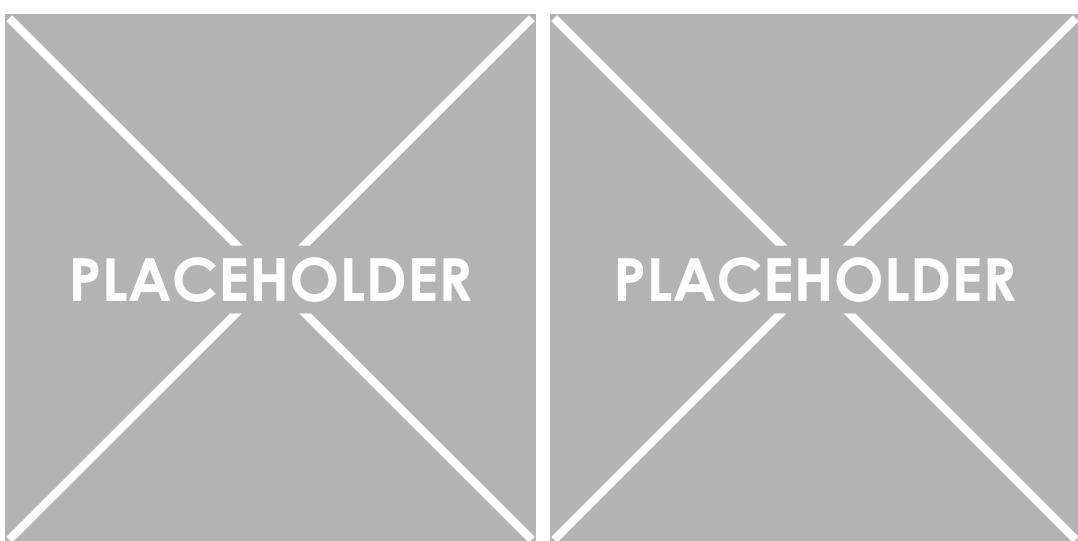
- |   |                       |
|---|-----------------------|
| • Exactly 2 same sign leptons               | • $m_{ll} > 12$ GeV   |
| • $p_{T,1} > 20$ (25) GeV for $e$ ( $\mu$ ) | • PuppiMET $> 20$ GeV |
| • $p_{T,2} > 13$ GeV                        | • $p_T^{ll} > 30$ GeV |
| • $ \eta  < 2.5$ for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

Several regions are then defined according to this cut, depending on the data taking period, on the channel, on the number of jets observed and on the  $p_T$  of the second lepton. Some resulting plots can be found in Figure 3.8.

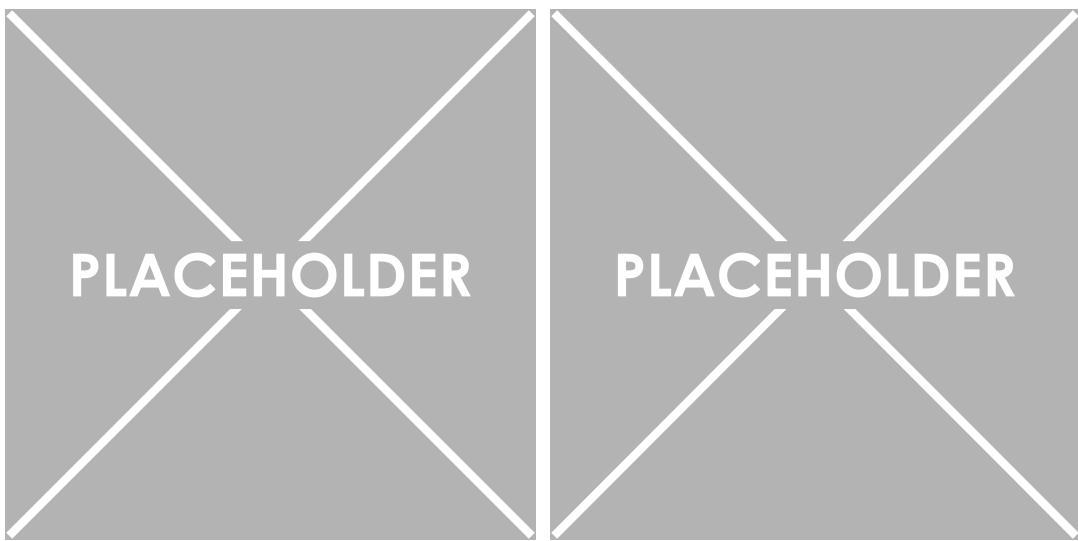
In order to cover for most of the discrepancies between the data and the simulation observed in the distributions of this control region, a flat 30% systematic uncertainty is typically associated to this background, as will be discussed in more details in Section 4.1.



(a) 2016 DY CR

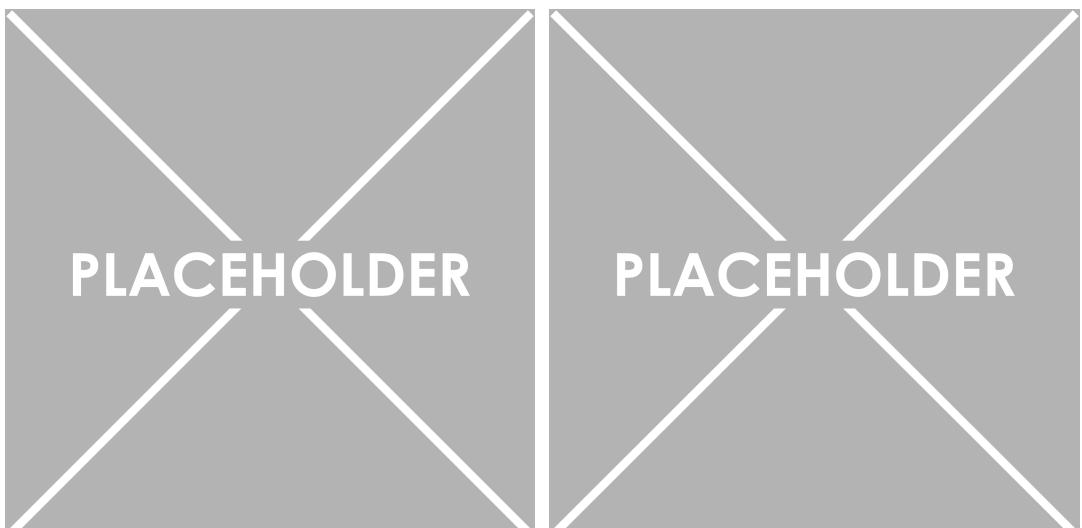


(b) 2017 DY CR

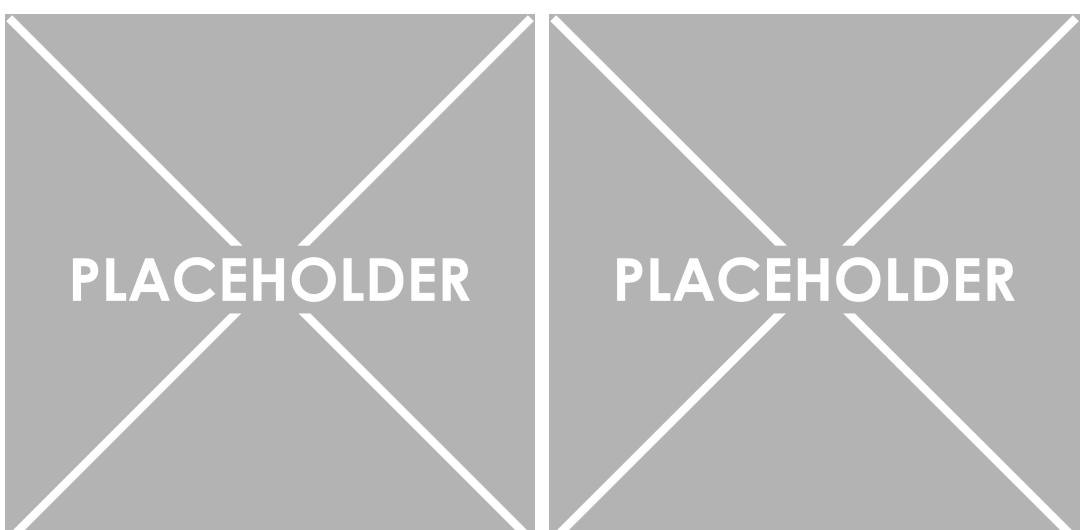


(c) 2018 DY CR

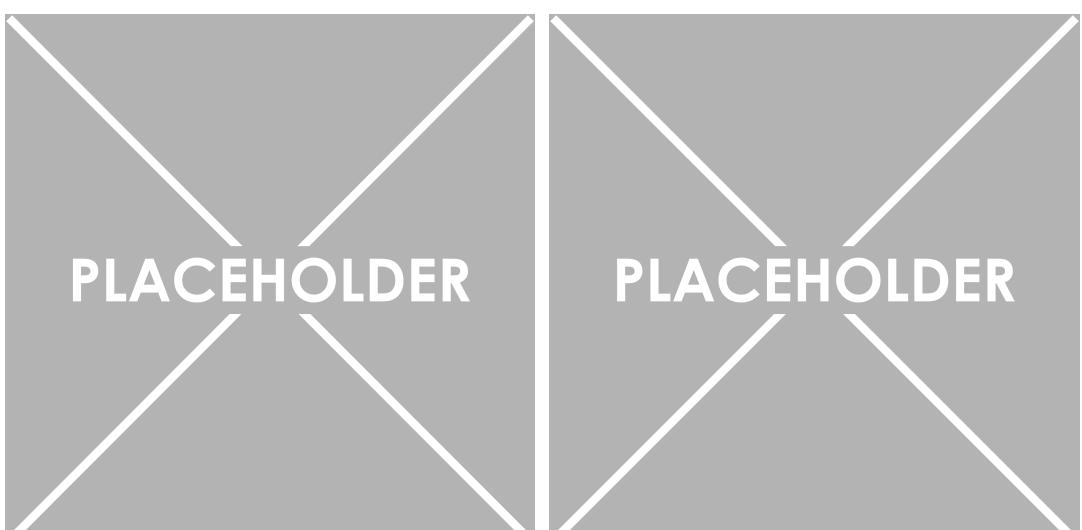
Figure 3.6: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the DY control region defined.



(a) 2016 top CR



(b) 2017 top CR



(c) 2018 top CR

Figure 3.7: Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the top control region defined.

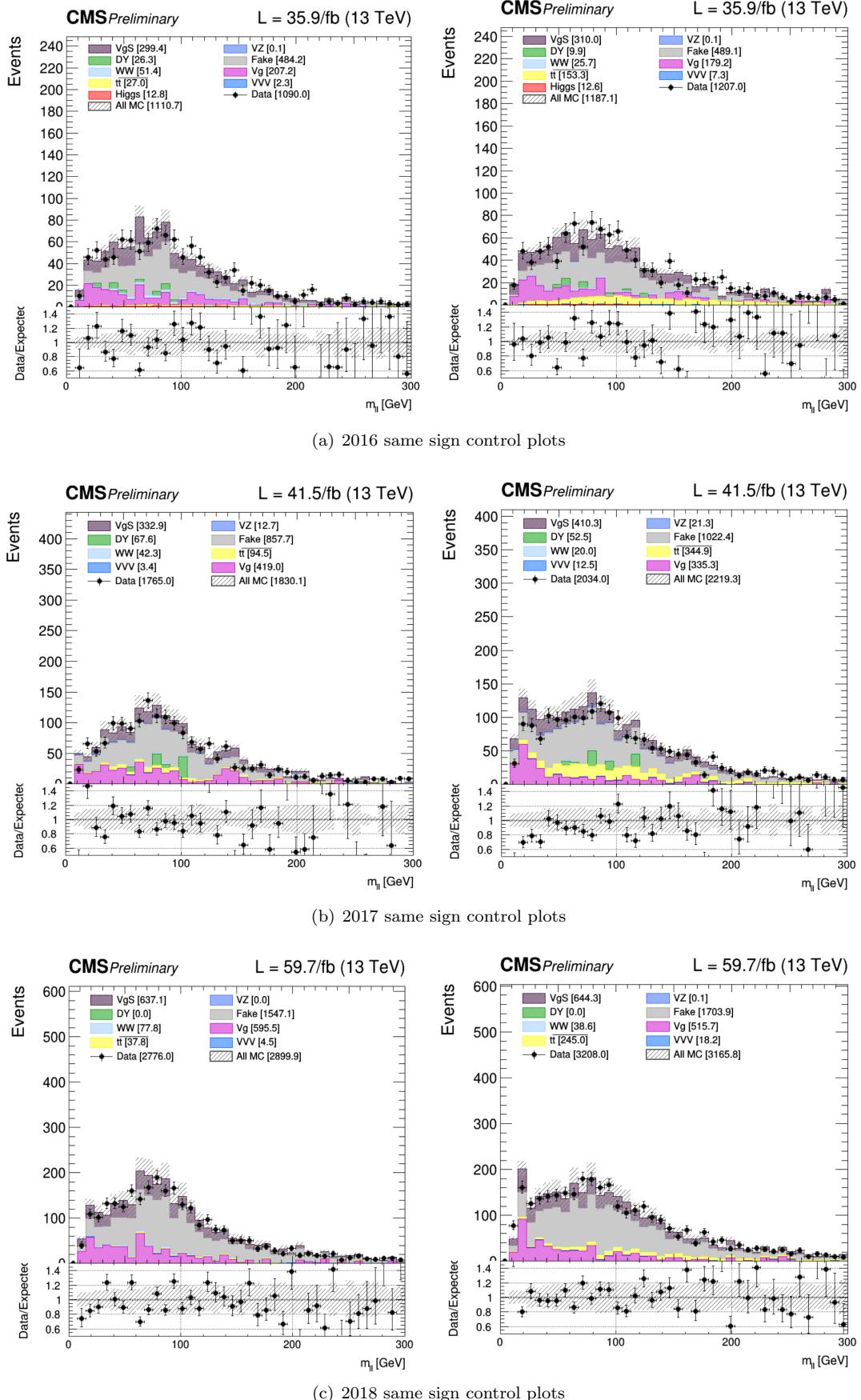


Figure 3.8: Same sign control region  $m_{ll}$  distributions for the years 2016, 2017 and 2018 and for the  $e\mu$  channel and for the 0-jet (on the left) and 1j (on the right) categories.

## 3.4 Background-signal discrimination

### 3.4.1 Discriminating variables

#### Dark $p_T$ and overlapping factor

Already introduced with the top reconstruction method in Section 1.6.2, these variables are of course expected to give some discrimination, since the standard  $t\bar{t}$  process is expected to peak around 0 while...

#### Missing Transverse Energy (MET)

This variable has already been defined in Section 1.5, and corresponds to the imbalance in transverse momentum which can be left by different phenomena, such as the apparition of a SM neutrino or the existence of DM particles, able to escape the detector without being detected.

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

TALK ABOUT SINGLE TOP?

#### Stransverse mass

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

In this particular case, two particles contribute to the presence of Missing Transverse Energy (MET) and the individual contribution of each particle ( $\not{p}_{T_1}$  and  $\not{p}_{T_2}$ ) to this missing energy cannot be inferred. The stransverse mass is then defined according to Equation 3.2, where  $\not{p}_{T_i} = \vec{p}_{T_i}$  is the (visible) transverse momentum of the particle  $i$  and  $\alpha$  is the angle between the visible and invisible  $p_T$  of the decay considered [121].

$$\begin{cases} M_{T2}^2 = \min_{\not{p}_{T_1} + \not{p}_{T_2} = \not{p}_{T_{\text{tot}}}} \left( \max \left( m_T^2(\not{p}_{T_1}, \not{p}_{T_1}), m_T^2(\not{p}_{T_2}, \not{p}_{T_2}) \right) \right) \\ m_T^2(\not{p}_T, \not{p}_T) = 4 |\not{p}_T| |\not{p}_T| \sin^2 \left( \frac{\alpha}{2} \right) \end{cases} \quad (3.2)$$

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

In this particular analysis,  $M_{T2}(ll)$  is calculated from a general algorithm described in [122], since the role of the visible particles is played by the two final state leptons. This variable is expected to

introduce some discrimination because, according to the definition just given, the  $M_{T2}(ll)$  variable for a SM  $t\bar{t}$  process is expected to have an endpoint exactly at the mass of the W boson, while an eventual  $t\bar{t}+{\rm DM}$  signal does not have this limitation in the  $M_{T2}(ll)$  spectrum because of the pair of DM particles produced, which also contributes to the total MET of the event.

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

TALK ABOUT SINGLE TOP?

### 3.4.2 Neural network

# Chapter 4

## Results and interpretations

### 4.1 Systematics and uncertainties

### 4.2 Results



## Chapter 5

# Conclusions

### 5.1 Future prospects



# **Appendices**



# Appendix A

## Samples used

### A.1 Data samples

All the data samples considered for this analysis are listed in Tables A.1, A.2 and A.3. The luminosity of each dataset has been computed using the Brilcalc tool provided by CMS [134], 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 $^{-1}$ ]
<b>Run 2016B</b>		
/DoubleEG/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	143073268 (99.4Gb)	
/DoubleMuon/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	82535526 (53.2Gb)	
/MuonEG/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	32727796 (26.8Gb)	5.8
/SingleElectron/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	246440440 (167.8Gb)	
/SingleMuon/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	158145722 (96.4Gb)	
<b>Run 2016C</b>		
/DoubleEG/Run2016C-Nano1June2019-v1/NANOAOD	47677856 (35.3Gb)	
/DoubleMuon/Run2016C-Nano1June2019-v1/NANOAOD	27934629 (19.7Gb)	
/MuonEG/Run2016C-Nano1June2019-v1/NANOAOD	15405678 (12.8Gb)	2.6
/SingleElectron/Run2016C-Nano1June2019-v1/NANOAOD	97259854 (69.3Gb)	
/SingleMuon/Run2016C-Nano1June2019-v1/NANOAOD	67441308 (42.4Gb)	
<b>Run 2016D</b>		
/DoubleEG/Run2016D-Nano1June2019-v1/NANOAOD	53324960 (39.6Gb)	
/DoubleMuon/Run2016D-Nano1June2019-v1/NANOAOD	33861745 (24.1Gb)	
/MuonEG/Run2016D-Nano1June2019-v1/NANOAOD	23482352 (19.4Gb)	4.2
/SingleElectron/Run2016D-Nano1June2019-v1/NANOAOD	148167727 (104.4Gb)	
/SingleMuon/Run2016D-Nano1June2019-v1/NANOAOD	98017996 (61.3Gb)	
<b>Run 2016E</b>		
/DoubleEG/Run2016E-Nano1June2019-v1/NANOAOD	49877710 (37.9Gb)	
/DoubleMuon/Run2016E-Nano1June2019-v1/NANOAOD	28246946 (20.8Gb)	
/MuonEG/Run2016E-Nano1June2019-v2/NANOAOD	22519303 (19.0Gb)	4.0
/SingleElectron/Run2016E-Nano1June2019-v1/NANOAOD	117321545 (86.5Gb)	
/SingleMuon/Run2016E-Nano1June2019-v1/NANOAOD	90984718 (58.7Gb)	
<b>Run 2016F</b>		
/DoubleEG/Run2016F-Nano1June2019-v1/NANOAOD	34577629 (26.9Gb)	
/DoubleMuon/Run2016F-Nano1June2019-v1/NANOAOD	20329921 (15.3Gb)	
/MuonEG/Run2016F-Nano1June2019-v1/NANOAOD	16002165 (13.6Gb)	3.1
/SingleElectron/Run2016F-Nano1June2019-v1/NANOAOD	70593532 (51.4Gb)	
/SingleMuon/Run2016F-Nano1June2019-v1/NANOAOD	65489554 (42.4Gb)	
<b>Run 2016G</b>		
/DoubleEG/Run2016G-Nano1June2019-v1/NANOAOD	78797031 (61.6Gb)	
/DoubleMuon/Run2016G-Nano1June2019-v1/NANOAOD	45235604 (34.2Gb)	
/MuonEG/Run2016G-Nano1June2019-v1/NANOAOD	33854612 (29.0Gb)	7.6
/SingleElectron/Run2016G-Nano1June2019-v1/NANOAOD	153363109 (109.2Gb)	
/SingleMuon/Run2016G-Nano1June2019-v1/NANOAOD	149912248 (94.6Gb)	
<b>Run 2016H</b>		
/DoubleEG/Run2016H-Nano1June2019-v1/NANOAOD	85388734 (67.7Gb)	
/DoubleMuon/Run2016H-Nano1June2019-v1/NANOAOD	48912812 (37.3Gb)	
/MuonEG/Run2016H-Nano1June2019-v1/NANOAOD	29236516 (26.0Gb)	8.6
/SingleElectron/Run2016H-Nano1June2019-v1/NANOAOD	128854598 (93.8Gb)	
/SingleMuon/Run2016H-Nano1June2019-v1/NANOAOD	174035164 (110.2Gb)	

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

Dataset	Events (size)	$\mathcal{L}$ [fb $^{-1}$ ]
<b>Run 2017B</b>		
/DoubleEG/Run2017B-Nano1June2019-v1/NANOAOD	58088760 (46.6Gb)	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)	
<b>Run 2017C</b>		
/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)	
<b>Run 2017D</b>		
/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)	
<b>Run 2017E</b>		
/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)	
<b>Run 2017F</b>		
/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)	$\mathcal{L} [\text{fb}^{-1}]$
<b>Run 2018A</b>		
/DoubleMuon/Run2018A-Nano25Oct2019-v1/NANO AOD	75499908 (62.6Gb)	
/EGamma/Run2018A-Nano25Oct2019-v1/NANO AOD	327843843 (261.8Gb)	13.5
/SingleMuon/Run2018A-Nano25Oct2019-v1/NANO AOD	241608232 (167.7Gb)	
/MuonEG/Run2018A-Nano25Oct2019-v1/NANO AOD	32958503 (32.3Gb)	
<b>Run 2018B</b>		
/DoubleMuon/Run2018B-Nano25Oct2019-v1/NANO AOD	35057758 (28.3Gb)	
/EGamma/Run2018B-Nano25Oct2019-v1/NANO AOD	153822427 (123.1Gb)	6.8
/SingleMuon/Run2018B-Nano25Oct2019-v1/NANO AOD	119918017 (82.3Gb)	
/MuonEG/Run2018B-Nano25Oct2019-v1/NANO AOD	16211567 (15.8Gb)	
<b>Run 2018C</b>		
/DoubleMuon/Run2018C-Nano25Oct2019-v1/NANO AOD	34565869 (27.6Gb)	
/EGamma/Run2018C-Nano25Oct2019-v1/NANO AOD	147827904 (119.2Gb)	6.6
/SingleMuon/Run2018C-Nano25Oct2019-v1/NANO AOD	110032072 (75.7Gb)	
/MuonEG/Run2018C-Nano25Oct2019-v1/NANO AOD	15652198 (15.3Gb)	
<b>Run 2018D</b>		
/DoubleMuon/Run2018D-Nano25Oct2019_ver2-v1/NANO AOD	168605834 (128.6Gb)	
/EGamma/Run2018D-Nano25Oct2019-v1/NANO AOD	751348648 (583.6Gb)	32.0
/SingleMuon/Run2018D-Nano25Oct2019-v1/NANO AOD	513867253 (344.5Gb)	
/MuonEG/Run2018D-Nano25Oct2019_ver2-v1/NANO AOD	71961587 (68.6Gb)	

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

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

Table A.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 $\gamma$ /V $\gamma^*$	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	405.271
	ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	58.83
	WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg_pythia8	58.59
VZ	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.212
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
VVV	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01398
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
	WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.16510
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.18331
Non-Prompt	Data-driven (tight-to-loose method)	

Table A.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 $\gamma$ /V $\gamma^*$	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	405.271
	ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	131.300
	WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg-pythia8	58.59
VZ	ZZTo2L2Nu_TuneCP5_13TeV_powheg-pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.212
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
VVV	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01398
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
	WWZ_TuneCP5_13TeV-amcatnlo-pythia8	0.16510
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.18331
Non-Prompt	Data-driven (tight-to-loose method)	

Table A.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

1.1	Transverse section of CMS showing the different tracks expected by different kind of particles in the detector. . . . .	3
1.2	Different kind of vertices typically observed in a $pp$ collision in the LHC. . . . .	3
1.3	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]. . . . .	5
1.4	Lepton isolation cone typically used to enhance the prompt leptons purity. . . . .	6
1.5	Schematic representation of the full electron reconstruction workflow in CMS [100]. . . . .	8
1.6	Schematic representation of the typical development of a jet within the CMS detector. . . . .	8
1.7	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]. . . . .	9
1.8	Schematic representation of the production of a b-jet originating from a slightly displaced secondary vertex. . . . .	10
1.9	B-jets identification efficiency and misidentification rate considering different b-taggers, including the deep CSV b-tag used in this analysis [103]. . . . .	11
1.10	Schematic representation of the Missing Transverse Energy (MET). . . . .	12
1.11	MET and jet $\phi$ distributions with and without MET filters applied [104]. . . . .	13
1.12	Three events constraining the neutrino and antineutrino momenta (black and grey arrows, respectively) resulting in 0 (on the left), 2 (on the center) or 4 (on the right) solutions. The dashed ellipses is obtained by using the additional constraint according to which measured MET is equal to the sum of neutrino transverse momenta [106]. . . . .	15
1.13	Schematic representation of the two extreme cases that can be observed when defining the overlapping factor: at the reconstruction of a system with (on the left) and without (on the right) the presence of DM [107]. . . . .	16
2.1	Structure of a $pp$ collision and different steps of the MC simulation used by the event generators, such as the parton shower (in green), the UE (in pink), the hadronization (in blue) and the decay of unstable particles (in red) [108]. . . . .	20
2.2	Top $p_T$ (on the left) and rapidity (on the right) distributions obtained using different MC generators [117]. . . . .	21

2.3	Proton energy distribution at 3 (on the left) and 6 (on the right) GeV compared for the test beam data (in black) and two different GEANT4 versions [119]. . . . .	23
2.4	PUPPI MET spectrum considering different <b>scalar</b> mediator (on the top) and dark matter (on the bottom) masses, with (on the left) and without normalization (on the right). . . . .	25
2.5	PUPPI MET spectrum considering different <b>pseudoscalar</b> mediator (on the top) and dark matter (on the bottom) masses, with (on the left) and without normalization (on the right). . . . .	26
2.6	Production cross section of the most common SM processes considering different center of mass energies, such as the 13 TeV of the LHC. . . . .	27
2.7	Main feynman diagrams for the production of the SM $t\bar{t}$ process. . . . .	27
2.8	Feynman diagrams for s-channel production mode of a single top quarks. . . . .	28
2.9	Feynman diagrams for t-channel (on the left) and tW (on the right) production modes of a single top quarks. . . . .	28
2.10	Feynman diagrams for the leptonic decay of the top (on the left) and antitop (on the right) quarks. . . . .	28
2.11	Feynman diagram for the DY process involving a virtual $\gamma^*$ or Z boson. . . . .	29
2.12	Possible Feynman diagrams for the Initial State Radiation (ISR) $t\bar{t}$ with a W/Z boson (on the left) and for the production of an Final State Radiation (FSR) ttZ (on the center and right). . . . .	29
2.13	Possible Feynman diagrams for the production of a W/Z boson with a jet. . . . .	30
2.14	Schematic representation of the two jets used for the systematics and for the jet faking a lepton in the tight-to-loose datadriven method. . . . .	31
2.15	Electron (on the left) and muon (on the right) FR obtained in a QCD enriched region for different jet $E_T$ thresholds for 2016, 2017 and 2018 with respect to the $p_T$ of the lepton. . . . .	32
2.16	Electron (on the left) and muon (on the right) PR obtained in a Z+jets enriched region by a general tag and probe method for 2016, 2017 and 2018 with respect to the $p_T$ of the lepton. . . . .	33
2.17	Possible Feynman diagrams for smaller backgrounds of this analysis: WW (on the left), W $\gamma$ and WZ (on the center) and ZZ (on the right). . . . .	35
2.18	Data/MC agreement with (on the right) and without (on the left) prefire corrections in a 2017 top control region. . . . .	36
3.1	DougleEG trigger efficiencies with respect to the $p_T$ (on the left) and $\eta$ (on the right), computed using a tag and probe method, for the 2017 data taking period. .	39
3.2	Tight electron efficiencies for this analysis, based on the Egamma POG <i>mva_90p_Iso2016</i> WP, for the data taking period of 2016. . . . .	42
3.3	Tight muon efficiencies for this analysis, based on the Muon POG tight WP with additional cuts for 2016, 2017 and 2018. . . . .	43

3.4	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the $t/\bar{t}+{\rm DM}$ process. . . . .	45
3.5	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the signal region defined to study the $t\bar{t}+{\rm DM}$ process. . . . .	46
3.6	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the DY control region defined. . . . .	48
3.7	Two different variables ( $m_{ll}$ , on the left, and PUPPI MET, on the right) represented in the top control region defined. . . . .	49
3.8	Same sign control region $m_{ll}$ distributions for the years 2016, 2017 and 2018 and for the $e\mu$ channel and for the 0-jet (on the left) and 1j (on the right) categories. . . .	50



# List of tables

3.1	2016 trigger paths considered for this analysis. . . . .	38
3.2	2017 trigger paths considered for this analysis. . . . .	38
3.3	2018 trigger paths considered for this analysis. . . . .	39
3.4	Quality cuts applied to define a 2016 tight electron in this analysis. . . . .	41
3.5	Quality cuts applied to define a 2017/2018 tight electron in this analysis. . . . .	42
3.6	JET/MET POG tight WP for 2016. . . . .	44
3.7	JET/MET POG tight WP for 2017 and 2018. . . . .	44
3.8	JET/MET POG tight WP for 2017 and 2018, for jets having a $p_T > 50$ GeV. . . . .	44
A.1	Datasets collected in 2016 and considered for this analysis. . . . .	60
A.2	Datasets collected in 2017 and considered for this analysis. . . . .	61
A.3	Datasets collected in 2018 and considered for this analysis. . . . .	62
A.4	Signal samples mass points considered for the $t\bar{t}$ +DM signal used in this analysis. .	63
A.5	Main 2016 MC simulations for the different background processes considered for this analysis and their respective cross sections. . . . .	64
A.6	Main 2017 MC simulations for the different background processes considered for this analysis and their respective cross sections. . . . .	65
A.7	Main 2018 MC simulations for the different background processes considered for this analysis and their respective cross sections. . . . .	66



# 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  $13.3 \text{ fb}^{-1}$  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<sup>-1</sup> 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] B.A. Betchart, R. Demina and A. Harel, "Analytic solutions for neutrino momenta in decay of top quarks", 2013 [arXiv: 1305.1878]
- [107] A. Lantero Barreda, "Improvement of the signal-to-background discrimination in search for Dark Matter produced in association with a pair of top-antitop quarks", UNICAN repository, 2018
- [108] M.H. Seymour and M. Marx, "Monte Carlo Event Generators", MCnet-13-05, 2013 [arXiv:1304.6677]
- [109] B. Cabouat, J.R. Gaunt and K. Ostrolenk, "A Monte-Carlo Simulation of Double Parton Scattering", JHEP11(2019)061 [arXiv: 1906.04669]
- [110] R. Placakyte, "Parton Distribution Functions", 2011 [arXiv:1111.5452]
- [111] J. Alwall et all., "MadGraph 5 : Going Beyond", 2011 [arXiv:1106.0522]
- [112] C. Oleari, "The POWHEG-BOX", Nucl.Phys.Proc.Suppl.205-206:36-41 [arXiv:1007.3893]
- [113] S. Frixione et all., "The MC@NLO 4.0 Event Generator", CERN-TH/2010-216 [arXiv: 1010.0819]
- [114] B. Webber, "Parton shower Monte Carlo event generators", Scholarpedia
- [115] M. Bahr et all., "Herwig++ Physics and Manual", Eur.Phys.J.C58:639-707, 2008 [arXiv: 0803.0883]

- [116] T. Sjostrand, "A Brief Introduction to PYTHIA 8.1" *Comput.Phys.Commun.*178:852-867, 2008 [arXiv: 0710.3820]
- [117] A. Karneyeu et all., "MCPLLOTS: a particle physics resource based on volunteer computing", *European Physical Journal C* 74 (2014) [arXiv: 1306.3436]
- [118] V. Lefebure and S. Banerjee, "CMS Simulation Software Using Geant4", CMS-NOTE-1999-072, 1999
- [119] S. Banerjee, "Validation of Geant4 Physics Models Using Collision Data from the LHC", *J. Phys.: Conf. Ser.* 898 042005
- [120] 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
- [121] C.G. Lester and D.J. Summers, "Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders", *Phys.Lett.B*463:99-103, 1999
- [122] C.G. Lester and B. Nachman, "Bisection-based asymmetric MT2 computation: a higher precision calculator than existing symmetric methods", *JHEP03(2015)100*, 1999 [arXiv: 1411.4312]
- [123] K. Bloom, "CMS software and computing for LHC Run 2", *ICHEP 2016* [arXiv: 1611.03215]
- [124] W. Tanenbaum, "A ROOT/IO Based Software Framework for CMS", *ECONFC0303241:TUKT010*, 2003
- [125] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", *CMS-PAS-LUM-17-001*, 2017
- [126] CMS Collaboration, "CMS Luminosity Measurements for the 2017 Data Taking Period", *CMS-PAS-LUM-17-001*, 2018
- [127] CMS Collaboration, "CMS Luminosity Measurements for the 2018 Data Taking Period", *CMS-PAS-LUM-17-001*, 2019
- [128] CMS Twiki, "Reweighting recipe to emulate Level 1 ECAL prefireing"
- [129] 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]
- [130] CMS Twiki, "Electron and Photon Physics Object Offline Guide"
- [131] CMS Twiki, "Baseline muon selections for Run-II", 2019
- [132] CMS Twiki, "Jet Identification"
- [133] CMS Twiki, "Heavy flavor identification at CMS with deep neural networks"
- [134] CMS Collaboration, "BRIL Work Suite"