Imperial College London

Department of Physics

PhD

Search for Dark Higgs and for Heavy Higgs resonances in top-antitop quark pair production with the CMS Experiment

Author
Samuel Stokholm Baxter

Supervisor
Prof. Oliver Buchmüller,
Imperial College London

Supervisor
Prof. Christian Schwanenberger,
DESY

Abstract

This work covers two searches for exotic physics in proton-proton scattering at CERN's Large Hadron Collider, utilising data from the CMS experiment taken at a centre of mass energy of 13 TeV. The first analysis involves a search for Dark Matter in association with the decay products of a new scalar coupled exclusively to the dark sector called the Dark Higgs. The other search is for a heavy Higgs boson that decays into a top-antitop pair. The analysis on the Dark Higgs signal utilises the relations between the scalar transverse momentum sum of jets, H_{T} and the negative vector sum of jet momentum $H_{\mathrm{T}}^{\mathrm{miss}}$ to separate signal from background. The result from $35.9\,\mathrm{fb}^{-1}$ of data from 2016 is presented. No deviation from the Standard Model prediction is observed, and the first experimental limits on Dark Higgs models ever are set, with Dark Matter mediator masses of up to 1700 GeV and Dark Matter particle masses up to 300 GeV can be excluded at 95% C.L. In the search for heavy scalar and pseudoscalar Higgs bosons, the invariant top-antitop mass is explored together with an angular observable sensitive to the correlation between the spin of the top quark and the spin of the antitop quark. For this search, $41.5\,\mathrm{fb}^{-1}$ of 2017 data has been used. Expected upper limits on the coupling between the heavy Higgs boson and the top quark are set for different hypotheses for the mass of the heavy Higgs boson, which are consistent with the expected limits from the same analysis on 2016 data.

Contents

List of Figures

List of Tables

Statement of Originality

I hereby declare that this work is my own and that any third party material contained herein is appropriately referenced.

Copyright Statement

1. Introduction

There are many mysteries in nature that are yet to be uncovered and understood, among these are some very fundamental questions such as: Where do we come from? How is everything connected? Will we ever find the theory on everything? Mankind is going to great lengths to try to answer these questions, and the rapid technological development of our day offers new opportunities to explore further and deeper than ever before. One area where this pursuit is particularly manifest is in experimental particle physics, where new discoveries can improve our understanding not only on the interactions between the elementary particles but also space and even the universe.

One of the main branches of experimental particle physics is collider physics, where collisions between accelerated particles at high energy are studied. From these collisions one can potentially discover new particles. Many interesting hypotheses predict particles at higher mass than the ones known today, which require collisions with as high energy as possible for potentially finding any of these particles. One such hypothesis with particular appeal is the assumption that the Dark Matter (DM) predicted by astronomical observations is a weakly interacting massive particle (WIMP). Under this assumption, DM can be produced from collisions via a mediator that can be either spin-1 or spin-0 that can couple to both to DM and to particles included in the Standard Model (SM). Since the DM is assumed to be only weakly interacting, it is expected to escape the detector undetected. To determine the occurrence of particles escaping detection, one looks for missing transverse energy.

The Large Hadron Collider (LHC) at CERN has produced proton-proton collisions at an unprecedented energy scale of 13 TeV. Two experiments at the LHC have collected vast amounts of data at this energy over three years. One of these two experiments is the Compact Muon Solenoid (CMS). Two analyses are presented in this work from which data collected from the CMS in 2016 and 2017, respectively.

The first analysis in this work studies a novel signature for DM production resulting from the radiation of an additional Higgs boson for the dark sector (**Dark Higgs**)

in association with a Z' mediator decaying into DM particles. In this search, the DM particles recoil against a visibly decaying Dark Higgs boson resulting in a signal composed of the final stages of the Dark Higgs decay products associated with missing transverse energy. This is in contrast to more conventional searches where the missing transverse energy is associated with the DM recoil against an SM state from initial state radiation. This search is particularly interesting because the presence of such Dark Higgs boson both accounts for the mass generation of the particles in the dark sector and enables the possibility of relaxing the cosmological constraints originating from the DM relic abundance.

The second search involves an additional Higgs particle that has similar properties to the Higgs boson discovered at the LHC in 2012 except for a higher mass. This additional Higgs particle (**Heavy Higgs**) can serve as a possible mediator to DM production, but it is not the link to DM that is exploited in this search. Under the assumption that the new scalar boson will couple most strongly to the heaviest known fermion, the top quark, which is assumed to have less than half the mass of the Heavy Higgs. These assumptions allow for the Heavy Higgs to decay into a top-antitop pair that will give rise to a resonance in the mass distribution of top-antitop pairs produced at the LHC. The result from this analysis on 2016 data, not presented here, has drawn a lot of attention due to a local excess over the SM prediction having been reported [1]. The analysis presented in this work is an update with the 2017 data set.

The outline of this work is as follows: Chapter 2 contains an overview of the SM with a focus on its shortcomings, in particular its lack of a candidate for a DM particle. Furthermore, an introduction to the Dark Higgs and Heavy Higgs models are given.

In Chapter 3, the experimental setup including the LHC and the CMS detector is described. Chapter 4 contains a description of event reconstruction.

Chapter 5 documents the Monte Carlo event simulation with particular emphasis on the validation of the signal generation.

The analyses are described in Chapter 6 for the Dark Higgs search and in Chapter 7 for the search for Heavy Higgs production.

In conclusion of this thesis a summary is given in Chapter 8.

Bibliography

[1] **CMS Collaboration**, *'Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at* $\sqrt{s} = 13$ TeV*'*,
JHEP 04 (2020) 171, doi:10.1007/JHEP04(2020)171, arXiv:1908.01115v2.

2. Theoretical Overview

This chapter provides key concepts from existing theory that provide a foundation for the two possible extensions to the established theory called the Dark Higgs Model and the Heavy Higgs Model. The physical consequences of these two models form the basis for the analyses conducted in this work.

2.1. The Standard Model and Beyond

The Standard Model (SM) is a gauge invariant quantum field theory that describes the interactions between the known elementary particles through the electromagnetic, weak and strong forces. It contains a unified gauge theory between the electromagnetic and weak interactions in an $SU(2) \times U(1)$ gauge group [1][2] and quantum chromodynamics (QCD) describing interactions of the strong force in an $SU(3)$ gauge group [3]. This theory offers accurate predictions for elementary particle interactions that are being verified with increasing accuracy. The recent discovery of the Higgs particle has completed the theory in the sense that all the particles predicted by the theory have been discovered [4][5].

2.1.1. The Higgs Mechanism

A key feature in the SM is the Higgs mechanism, which enables the existence of massive gauge bosons [6][7][8] and is furthermore used to enable the fermions to acquire mass [9]. This Higgs mechanism is here demonstrated using the formalism from [10]. One starts with the Higgs potential:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2, \tag{2.1}$$

where we set $\mu^2 > 0$ and $\lambda > 0$. We take the field: $\phi(x) = U(x)\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$, where $h(x)$ is a real scalar field with $\langle h(x) \rangle = 0$, and $U(x)$ is a unitary gauge

transformation under $SU(2)$ to make the most general case with a complex doublet field. One can transform away $U(x)$ so that the field is reduced to real parameters. The acquisition of a non-zero minimum of the Higgs potential in a random direction in the complex plane is called spontaneous symmetry breaking. The potential in eq. (2.1) is minimised when we have:

$$v = \sqrt{\frac{\mu^2}{\lambda}}, \tag{2.2}$$

which is the vacuum expectation value (vev) of the Higgs field. When evaluating the potential in therms of the vev and the fluctuating field, one obtains the Higgs mass, $m_h = \sqrt{2}\mu = \sqrt{2\lambda}v$. When considering the Lagrangian:

$$\mathcal{L} = |D_\mu \phi|^2 - V(\phi), \tag{2.3}$$

where we use the covariant derivative:

$$D_\mu = \partial_\mu(-igA^a_\mu \tau^a - i\frac{1}{2}g'B_\mu)\phi, \tag{2.4}$$

where g and g' are the couplings to the two gauge fields A and B and τ^a is related to the Pauli matrices σ^a.

The kinetic term from eq. 2.3 will then contain the masses of the gauge bosons as follows:

$$\mathcal{L}_K = \frac{1}{2}(\partial_\mu \phi)^2 + \left[m_W^2 W^{\mu+} W^-_\mu + \frac{1}{2}m_Z^2 Z^\mu Z_\mu\right] \cdot \left(1 + \frac{h}{v}\right)^2 \tag{2.5}$$

Equivalent terms for fermions are given as:

$$\mathcal{L}_f = -m_f f \bar{f}\left(1 + \frac{h}{v}\right). \tag{2.6}$$

These terms contain mass terms for the fermions and give rise to the Higgs-fermion couplings in the form: $g_{hff} = \frac{m_f}{v}$.

2.1.2. Charge-Parity

Interactions in the SM are generally invariant under a combination of charge conjugation and parity, called CP-invariance. Under charge conjugation, the particles

in the system are replaced with their antiparticles. For particles indistinguishable from their antiparticles, a quantum number called C-parity is assigned. Parity transformation reflects the system in space, which leads to the reversal of chirality of particles in the system. Each particle has a quantum number for parity, which is established by analysing a change in parity of a system caused by the given particle. Fermions are assigned parity of $P = 1$ and antifermions have $P = -1$, and a fermion-antifermion pair will have a P value of -1 in the bound state with lowest energy, implying zero angular momentum. In general, the parity of a set of two particles is the product of the parities of the two particles times $(-1)^L$, where L is the orbital angular momentum of the two particles. Similarly for charge conjugation where the product of the C-parities if any is taken and multiplied by $(-1)^{S+L}$, where L is the orbital angular momentum and S is the total spin of the system. For CP-invariance to be valid, the initial state of a process needs to have the same CP quantum number as the final state. There are some special cases of SM processes involving decays of neutral heavy mesons, where final states with a different CP number than the initial state have been observed. These processes are therefore CP-violating and involve mixing via flavour changing weak interactions in loop diagrams as seen in fig. 2.1 for the neutral kaon. A CP-violating phase is included in the matrix describing the mixing between quarks (CKM) [11] as a result. The quark flavour mixing occurs when the quarks interchange between their mass eigenstate and their flavour eigenstate.

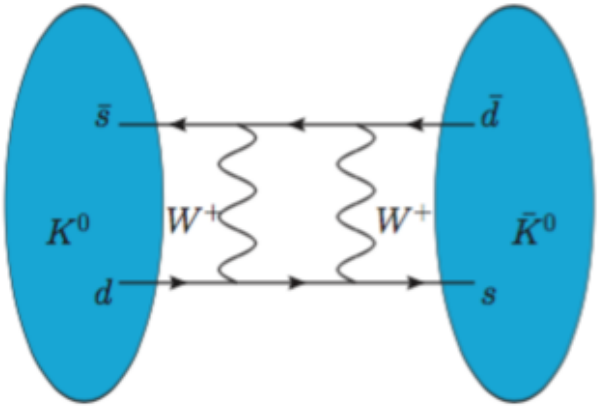


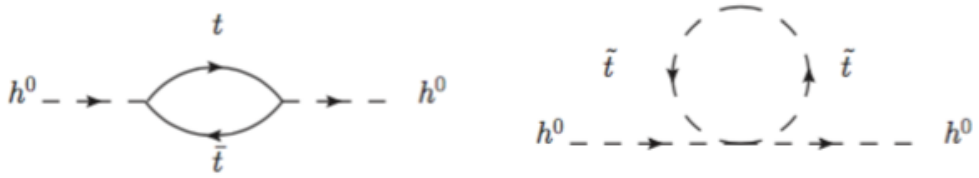
Figure 2.1.: Feynman diagram for mixing between K^0 and $\bar{K}^0$.

CP-violation is considered a key requirement for obtaining the observed baryon asymmetry of the early universe, and the CP-violating phase of the CKM matrix in the SM leads to an insufficient amount of CP-violation to account for the

production of the observed baryon asymmetry under the conditions of the early universe [12].

2.1.3. The Hierarchy Problem

One aspect of the SM which is considered a weakness is the fact that a slight perturbation of the Higgs coupling can lead to changes in the Higgs mass that are several orders of magnitude larger than the Higgs mass itself [13][14][15]. This is due to the fact that the Higgs will couple to any massive particle that exists, impacting the Higgs mass through higher order Feynman diagrams involving loops of particles coupled to the Higgs, for example fermions like the top quark (see fig. 2.2a). These loops give contributions to the squared Higgs mass of the order of the squared mass of the particle in the loop. With the known SM particles, this is not an issue, but if one introduces a new particle of any form at a sufficient mass scale to unify the SM with gravity, one encounters a problem with the Higgs mass in the sense that the Higgs particle is much lighter than the new particle at the unification scale $\Lambda \sim 10^{16}$ GeV. The resulting renormalisation of the Higgs mass term, $-\mu^2$ will have to result from taking a value of the order Λ^2. To land on the Higgs mass, one ends with a parameter which is constrained within the order of 10^{-28}, which is a precision considered unnatural.



(a) (b)

Figure 2.2.: Feynman diagrams for one-loop corrections to the Higgs energy with the top quark (a) and the supersymmetric counterpart, the stop (b).

One needs a mechanism to avoid the need for this fine tuning of the renormalisation, where one of the more appealing ideas is to add new particles that cancel out the contributions to the Higgs mass from the loop diagrams (see fig. 2.2b). This procedure works because boson loops (both scalar and vector) make positive contributions to the Higgs mass whereas the contributions from fermion loops are negative. For each known boson one can then add a fermion with the same mass,

and for each known fermion one adds a scalar boson with the same mass. This is the approach to solving the Hierarchy problem taken in supersymmetric theories (SUSY) [16].

2.1.4. Proton-Proton Collisions

To understand what goes on when two protons collide, one first needs to understand the underlying structure of the proton. The simple model of the proton consists of three quarks (uud) bound together by gluons. Both the gluons and quarks are referred to as partons in collision physics, and a key component for predicting the outcome of proton-proton collisions is an accurate estimate of the parton distribution function (PDF). The outcome that one is mainly interested in is the cross-section for the final state of interest, which can either be hypothetical or previously discovered. The cross-section is the conventional way to express the predicted rate of a process. For proton-proton collisions, the strong force is dominant with QCD being the underlying theory. When the momentum transfer in the process is high enough for perturbation theory to be valid, the process is referred to as hard. Typically there are also interactions involved in the hard processes with low momentum transfer referred to as soft interactions for which perturbation theory breaks down. These soft processes have very unreliable theoretical predictions and are still necessary to account for when comparing theoretical predictions to experimental data.

For hard processes, the cross-section for the proton-proton collision to produce a given final state X is given in the following equation at leading order [17]:

$$\sigma_{AB} = \int \mathrm{d}x_a \mathrm{d}x_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \cdot [\hat{\sigma}_0 + \alpha_s(\mu_R^2)\hat{\sigma}_1 + ...]_{ab\to X}, \qquad (2.7)$$

where A and B are the two protons, a is the interacting parton from A, b is the interacting parton from B, $f_{a/A}$ and $f_{b/B}$ are the PDFs for A and B respectively. The factorisation scale μ_F characterises the momentum transfer of the process $ab \to X$ which is taken into account in the PDFs. The running coupling α_s depends on the renormalisation scale μ_R, which generally is assumed to take the same value as μ_F. The terms involving α_s are radiative corrections given as a logarithmic expansion, the accuracy and complexity of which is stated as (n times next to) leading logarithm (NNLL). The scales μ_F and μ_R can assume alternative values that are equally valid within a reasonable range which is typically set at a factor two. The choice of these scales have a slight impact on the final cross-section. The variable x is the momentum fraction of the given parton.

The PDFs are determined with fits to experimental data aided by theoretical assumptions. In these experiments, the parton structure of the proton has been examined at different energies with different probes. Among the probes used are positrons, photons, antiprotons and protons. Each type of probes has given significant contributions to our knowledge of the parton structure of the proton. Currently it is possible to obtain accuracies up to next to next to leading order (NNLO) thanks to the vast amounts of data collected from collider experiments over the past decades. In the meantime the theoretical understanding of QCD processes has increased to NNLO accuracy in calculations to make better use of the accuracy obtained from data [18].

2.1.5. The Higgs Particle

Since the discovery of the Higgs particle [5][4], the nature of this new particle has been investigated further to check that it indeed behaves according to the SM predictions. The SM predicts that the Higgs particle is a spin zero fundamental particle with CP=1. These features have been proven experimentally, giving more clarity on the nature of the Higgs sector. Other interesting features of the Higgs particle involve Higgs production and decay channels that have well defined SM predictions. So far, these predictions are consistent with the observations [19][20][21]. The Higgs mass is observed to be $125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV, and the recommended value for the Higgs mass in calculations is 125 GeV, which is assumed for the cross-sections stated below [22].

Higgs Production

There are several production channels for the Higgs particle in proton-proton collisions, each with distinctive features that help establish the couplings of the Higgs particle to vector bosons and the top quark. Some of the rarer channels have proven helpful in isolating Higgs events with final states that are otherwise challenging to separate from background.

The leading Higgs production channel is called gluon fusion (shown in fig. 2.3a) where two gluons fuse together in a fermion loop that produces the Higgs particle. It has a predicted cross-section for a Higgs particle at $\sqrt{s} = 13$ TeV of $48.58\ ^{+2.22}_{-3.27}(\text{theory}) \pm 1.56(\text{PDF} + \alpha_s)$ pb at NNNLO [23]. Another Higgs production channel is called vector boson fusion (VBF), where the Higgs particle is

produced from a pair of emitted vector bosons from quarks (shown in fig. 2.3b). In addition to the Higgs signal, VBF produces two jets that make it distinguishable from gluon fusion. It is significantly rarer than gluon fusion with a cross-section at 13 TeV of 3.782 $^{+0.43\%}_{-0.33\%}$(QCD scale) $\pm$ 2.1%(PDF + α_s) pb [22].

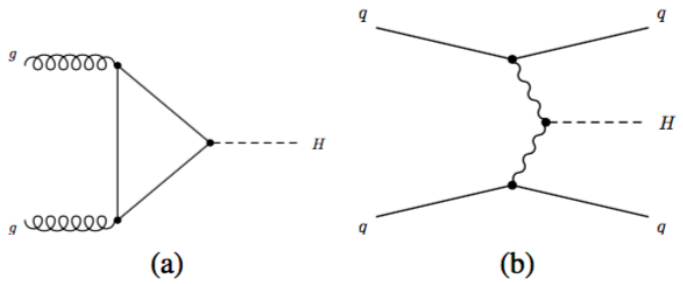


Figure 2.3.: Leading order Feynman diagrams for Higgs production via gluon fusion (a) and vector boson fusion (b). Figure taken from source [19].

Production of a Higgs boson with an additional vector boson (VH) is the leading category of Higgs production with additional heavy particles. The recoil from the vector boson and the additional final state particles make this channel the main contributor in the $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ searches [19]. The leading VH process is the production of a W boson in addition to the Higgs particle (WH) which originate from two quarks in the initial state (see fig. 2.4a). The cross-section for this channel is predicted for 13 TeV to be 1.373 $^{+0.5\%}_{-0.7\%}$(QCD scale)$\pm$1.9%(PDF+α_s) pb. The remaining VH process produces an additional Z boson (ZH), which can either happen through gluon fusion (figg. 2.4b and c) or via quarks (fig. 2.4a). The predicted inclusive cross-section for ZH at 13 TeV is 0.8837 $^{+3.8\%}_{-3.1\%}$(QCD scale) $\pm$ 1.6%(PDF + α_s) pb [22].

Higgs production with additional heavy quarks (top or bottom) provides a way to probe the Higgs couplings to quarks, especially the production with top quarks is interesting since the Higgs is too light to decay into top quarks. Therefore the coupling to top quarks is probed in the associated production rather than Higgs decays. The most abundant of these production events involve production of a quark-antiquark pair ($t\bar{t}$ or $b\bar{b}$) shown in fig. 2.5. These processes are called ttH and bbH respectively. They have predicted cross-sections of 0.5071 $^{+5.8\%}_{-9.2\%}$(QCD scale) $\pm$ 3.6%(PDF + α_s) pb for $t\bar{t}$ and 0.488 $^{+20.2\%}_{-23.9\%}$(QCD scale + PDF + α_s) pb for $b\bar{b}$ [22]. Observation of this Higgs production channel has been achieved in 2018 [24][25].

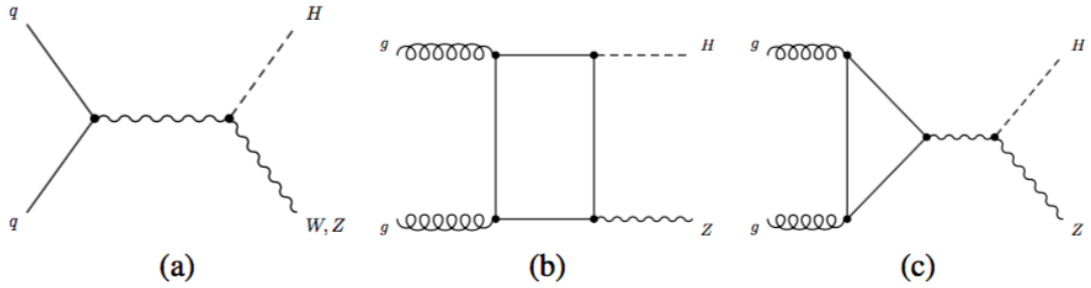


Figure 2.4.: Leading order Feynman diagrams for Higgs production with an additional vector boson originating from two quarks (a) and gluon fusion producing a Higgs and Z boson (b) and gluon fusion producing a Z boson radiating off a Higgs (c). Figure taken from source [19].

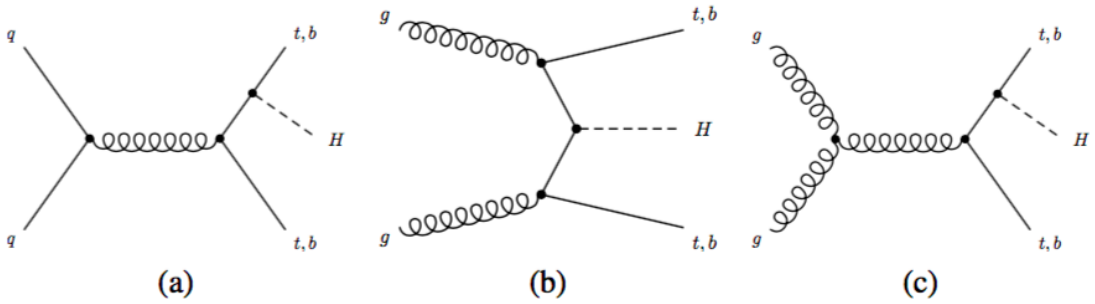


Figure 2.5.: Leading order Feynman diagrams for Higgs production with an additional $t\bar{t}/b\bar{b}$ pair originating from two quarks (a) and gluon fusion producing a Higgs and $t\bar{t}/b\bar{b}$ pair (b) and gluon fusion producing a $t\bar{t}/b\bar{b}$ pair radiating off a Higgs (c). Figure taken from source [19].

Higgs production with a single top quark is also possible with either an additional quark or W boson. Feynman diagrams for these processes are shown in fig. 2.6. The predicted cross-section is $0.07425\ ^{+6.5\%}_{-14.9\%}(\text{QCD scale} + \text{FS}) \pm 3.7\%(\text{PDF} + \alpha_s)$ for t-channel production of a quark and top quark emitting a Higgs (see fig. 2.6b). FS represents the uncertainty due to the choice of flavour scheme, which influences how the Feynman diagrams are constructed and which quarks to consider massless. For the t-channel production of a top and W with either of them radiating off a Higgs, the predicted cross-section is $0.01517\ ^{+4.9\%}_{-6.7\%}(\text{QCD scale}) \pm 6.3\%(\text{PDF} + \alpha_s)$ (see figg. 2.6c and 2.6d), while it is $0.002879\ ^{+2.4\%}_{-1.8\%}(\text{QCD scale}) \pm 2.2\%(\text{PDF} + \alpha_s)$ pb for s-channel production of a top radiating off a Higgs and bottom quark (see fig. 2.6a) [22].

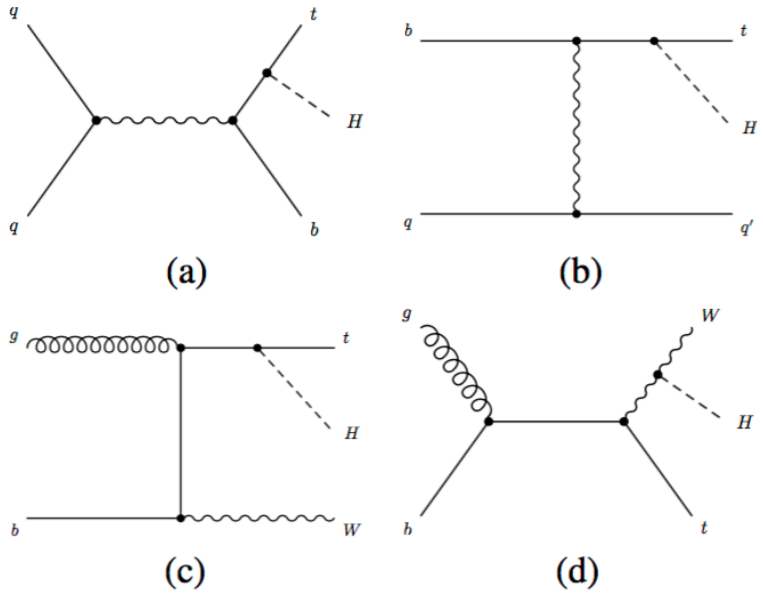


Figure 2.6.: Leading order Feynman diagrams for Higgs production radiated off a top quark and an additional bottom quark, originating from two quarks in s-channel (a) and t-channel producing a top quark radiating off a Higgs with an additional quark (b) and gluon-bottom quark fusion producing a top and W with either of them radiating off a Higgs (c) and (d). Figure taken from source [19].

Higgs Decay Channels

The Higgs particle can decay into several different pairs of particles whereof most of the decay channels provide a signal that can be used in searches and will be discussed here. Higgs decays happen either at tree level into heavy vector bosons and fermions or through loops to decay into photons.

The leading decay channel with a branching ratio of $0.577\ ^{+3.2\%}_{-3.3\%}$ [26] is the $H \to b\bar{b}$ decay, which has been discovered in 2018 [21][20]. Although this decay channel is the most abundant, it has proven hard to obtain a significant signal from it due to the QCD background. The production channel contributing the most to the discovery of the $H \to b\bar{b}$ decay has been the VH channel, which is an order of magnitude rarer than gluon fusion.

The Higgs to WW^* is the second leading decay channel with a branching ratio of $0.215\ ^{+4.3\%}_{-4.2\%}$ [26]. At least one of the W bosons is produced off-shell because the Higgs particle is lighter than $2M_W$. The observation of this decay channel has been reported in 2015 , with the leptonic decays of W bosons being most viable for the search. Measurements on this decay channel have also determined key properties of the Higgs particle like the spin nature of the Higgs particle (spin 0 or spin 2) [27].

The next decay in terms of branching ratio is the Higgs to $\tau^+\tau^-$ decay, which has been reported as an observation in 2016 [19]. The branching ratio of this channel is $0.0632 \pm 5.7\%$ [26]. Studies in this channel have given additional insight in the Higgs coupling to fermions.

Just like the Higgs can decay into WW^*, the Higgs can also decay into ZZ^*, which has been reported as observed in 2014 [28] and verified in 2015 [29]. The branching ratio of this channel is $0.0264\ ^{+4.3\%}_{-4.2\%}$ [26]. From this decay channel it is also possible to establish the spin properties of the Higgs particles with respect to whether the Higgs particle is a pure scalar or a pseudoscalar [30][31].

The final Higgs decay channel discussed here is the Higgs to two photons, which produces the clearest signal of all the decay channels in spite of a branching fraction of merely $2.28 \cdot 10^{-3}\ ^{+5.0\%}_{-4.9\%}$ [26]. This decay channel is the main contributor to the discovery of the Higgs particle in 2012 [5][4].

2.1.6. The Top Quark

As the heaviest particle in the SM, the top quark has the strongest coupling to the Higgs particle. It has some unique properties compared to the other quarks in the sense that it decays before it can hadronize, meaning that the top quark can be studied individually rather than as a component of a hadron.

The top can be produced in pairs as shown for ttH in fig. 2.5, just without the Higgs. At 13 TeV proton-proton collisions the cross-section of the $t\bar{t}$ production is predicted at NNLO to be $831.76\ ^{+19.77}_{-29.2}(\text{Scale})\ ^{+23.18}_{-22.45}(\text{Mass}) \pm 35.06(\text{PDF} + \alpha_s)$ pb, where mass refers to the uncertainty in the mass of the top quark which has been set to 1 GeV, under the assumption that the top mass is 172.5 GeV [32][33]. Just as the $t\bar{t}$ process occurs without Higgs radiation, the single top processes shown in fig. 2.6, can also occur without the Higgs. Other top production events that can occur are with four top quarks or a $t\bar{t}$ pair with emission of an additional vector boson in a similar fashion as ttH.

The top quark decays almost exclusively into a W boson and a bottom quark, with the W decaying either hadronically into quarks with a branching fraction of ~ 0.66 or leptonically into a lepton and neutrino with a branching fraction of ~ 0.33 [34]. The spin of the top quark can most effectively be measured from the down-type decay product of the W boson, which in the case of leptonic decay is the charged lepton, which will tend to travel in the direction of the top spin in the top rest frame. This feature makes it possible to determine spin correlations in $t\bar{t}$ events [35][36].

2.1.7. Dark Matter

One example of particles beyond the Standard Model (BSM) is Dark Matter (DM). Strong astrophysical evidence suggests that DM exists, for example the rotation speed of galaxies [37] and gravitational lensing of galaxies passing through each other [38]. The exact nature of DM is practically unknown, and DM hypotheses span over a range of 40 orders of magnitude. Within this vast scope of parameter space lie several hypotheses with a variety mechanisms for DM to couple to SM particles. From these hypotheses, one can conduct experiments to prove the hypotheses.

One of the DM hypotheses of particular interest is that of DM being a weakly

interacting massive particle (WIMP). The main assumptions of this hypothesis are that DM has a mass at the electroweak scale and couples to the SM at the scale of the weak force. It is considered cold DM and is generated via thermal freeze-out, allowing DM to form with minimal disruption of the formation and evolution of the known universe. One major advantage of this DM hypothesis is that it is possible to test by means of our current technological capabilities. Dedicated experiments around the world are built for the purpose of detecting WIMPs, mainly through elastic scattering with heavy atomic nuclei such as Xenon [39]. This approach to searching for DM is often referred to as direct detection. In addition, searches for WIMPs are also performed in accelerator experiments, where one hopes to produce DM that will escape undetected, but leave a recoil of missing energy [40]. Astronomers are searching space for cosmic radiation influenced by DM, for example a resonance of gamma radiation originating from DM annihilation [41]. These three types of searches complement each other and ideally it is hoped that a discovery in one type of search can be verified in a search of another type in the cases where the regions of parameter space in the given searches overlap [42]. The current experimental reach of direct detection experiments and the CMS experiment under the Higgs portal hypothesis in terms of the spin-independent cross-section and DM mass are shown in fig. 2.7.

The various DM searches have so far failed to produce consistent evidence for any WIMP hypothesis and have thus excluded much of the parameter space where WIMPs can exist (see fig. 2.7). To justify continued searches, it is necessary to identify mechanisms that can explain why the WIMP has not been found yet. To make matters worse, much of the remaining parameter space for searches lies in regions where one expects DM overproduction according to the most commonly tested WIMP hypotheses [43].

2.2. The Dark Higgs Model

The Dark Higgs Model extends the SM with three new particles, one of which is a DM candidate. The advantage of adding three new particles to SM physics is that one can add a new $U(1)'$ gauge group which adds flexibility to the search region where DM can be probed, relaxing constraints on DM overproduction and opening new DM search regions.

The three particles added to the SM are as follows: A Majorana fermionic DM particle (χ), a spin 1 vector boson (Z') and a scalar boson (s). These new particles

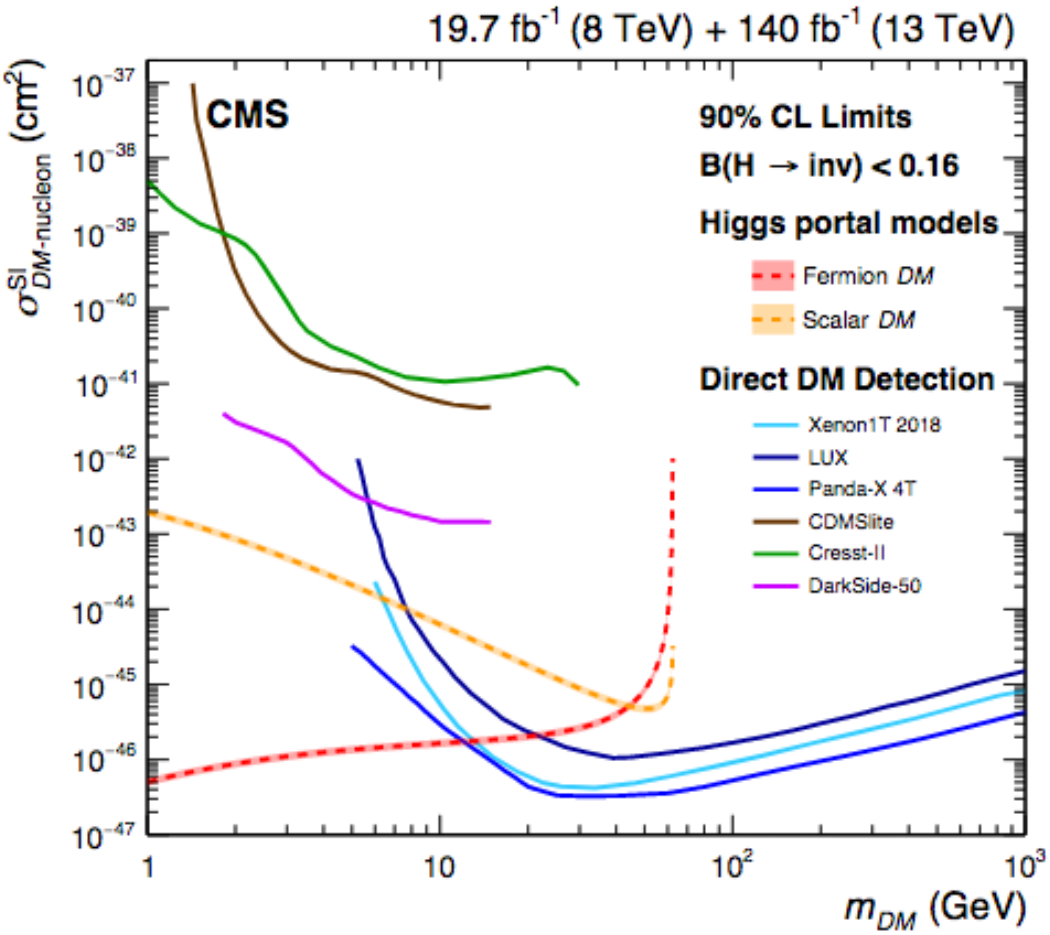


Figure 2.7.: Summary of the results from the leading direct detection experiments (cyan, blue, cobalt, brown, green and magenta lines) and the result from the CMS experiment under the Higgs portal DM hypothesis for scalar DM (orange line) and fermionic DM (red line). Figure taken from source [40].

form a new U(1)' gauge group, where spontaneous symmetry breaking gives rise to the following terms:

$$\mathcal{L}_\chi = -\frac{1}{2} g_\chi Z'^\mu \bar{\chi}\gamma^5\gamma_\mu\chi - g_\chi \frac{m_\chi}{m_{Z'}} s\bar{\chi}\chi + 2g_\chi Z'^\mu Z'_\mu (g_\chi s^2 + m_{Z'} s), \tag{2.8}$$

where g_χ is the coupling between Z' and DM, $m_{Z'}$ and m_χ are the masses for Z' and DM respectively.

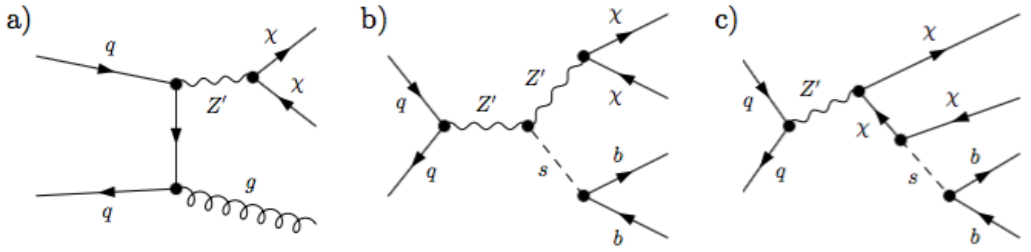


Figure 2.8.: a. typical mono-jet process with DM (χ) production, b. and c. possible processes for Dark Higgs (s) production, figure taken from the source [44].

In addition, the Z' is coupled to the SM quarks, giving rise to the following term:

$$\mathcal{L}_q = -g_q Z'^{\mu} \bar{q} \gamma_{\mu} q \tag{2.9}$$

The mass relations of the particles are as follows: $m_{Z'} > 2m_{\chi}$ and $m_s < m_{\chi}$. Requiring the Dark Higgs to be lighter than the DM particle opens up the annihilation channel, $\chi\chi \to ss$ followed by $s \to SM$. This annihilation channel reduces the parameter space for overproduction of DM in the early universe.

The Dark Higgs has a mixing angle with the SM Higgs, allowing it to quickly decay into SM particles, whereas the reverse process is suppressed to insignificance. Depending on the mass of the Dark Higgs, it will have different decay modes, where focus in this work is on the case where the Dark Higgs decays into a $b\bar{b}$ pair, which is the dominant decay mode when the Dark Higgs has a mass between 10 and 160 GeV.

With the assumption that the Dark Higgs has a mass between 10 and 160 GeV, we will have the dominant decay $s \to b + \bar{b}$. The dominant processes involving $s \to b + \bar{b}$ follow the Feynman diagrams in fig. 2.8b and fig. 2.8c. According to eq. 2.8, Z' decays with a Dark Higgs generated (fig. 2.8 b and c) are less likely than a decay into only DM (fig. 2.8a). The processes with Dark Higgs production will still add up to a branching fraction of more than 5%, given that $m_{Z'}, \gg m_{\chi}, m_s$. Although the process in fig. 2.8a is far more common than Dark Higgs production, it is also more difficult to separate from background. A search for the decay products of the Dark Higgs provides a complimentary DM search to mono-jet and dijet searches for DM as explained below.

The standard value for the coupling between quarks and a new heavy vector boson

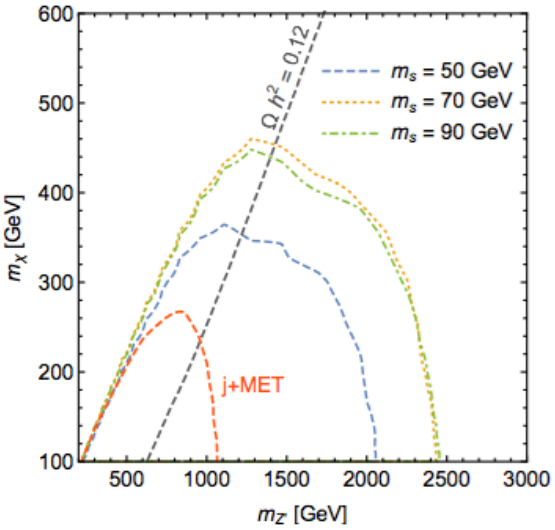


Figure 2.9.: Expected sensitivity for a Dark Higgs boson with $m_s = 90$ GeV (green dot-dashed line), $m_s = 70$ GeV (orange dotted line) and $m_s = 50$ GeV (blue dashed line) in the $m_{Z'} - m_\chi$ parameter space where $g_q = 0.25$ and $g_\chi = 1$. The expected sensitivity for a conventional monojet search (dark orange dashed line) is also shown for comparison along with the $m_{Z'} - m_\chi$ combinations resulting in the observed DM relic abundance (black dashed line). Figure taken from the source [44].

is $g_q = 0.25$, while the default coupling for DM to Z' is $g_\chi = 1$ [45].

When using the default couplings, the expected sensitivity for a Dark Higgs search can probe Z' masses up to 2.5 TeV for Dark Higgs hypotheses with the strongest signal (see fig. 2.9). This is significantly more than for the search result in the mono-jet channel for the DM model with a vector mediator where the mediator mass reaches 2 TeV [46][47]. In the dijet channel, the exclusions are much stronger, reaching Z' masses up to 4 TeV [48]. One can however adjust the coupling g_χ to suppress the dijet exclusion by either reducing it and thereby suppressing the overall DM production, leaving a region of parameter space covered by only the mono-jet search. One can alternatively increase g_χ enhancing the coupling between Z' and DM while reducing the branching ratio for $Z' \to q\bar{q}$ significantly. One consequence of increasing g_χ and thereby the DM production is that the cosmological limit of $\Omega h^2 = 0.12$ [49] is exceeded in more of the parameter space than shown in fig. 2.9, even when the annihilation channel $\chi\chi \to ss$ is considered. A strategy

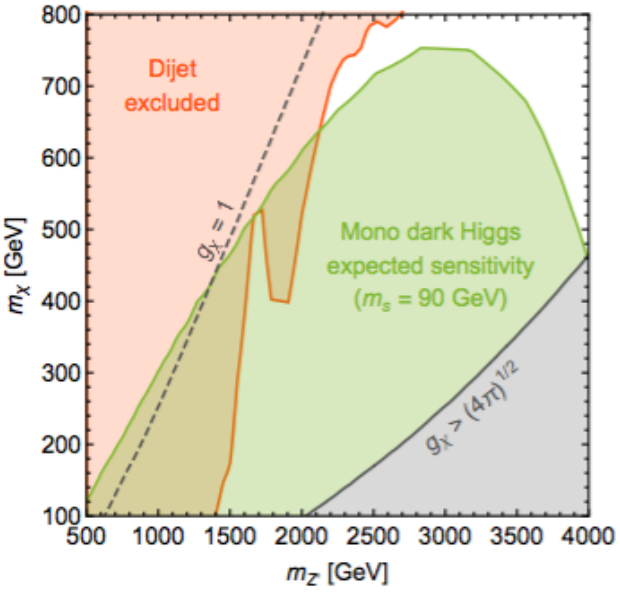


Figure 2.10.: Expected sensitivity for a Dark Higgs boson with $m_s = 90$ GeV in the $m_{Z'}, m_\chi$ parameter space at 40 fb^{-1}, where $g_q = 0.25$ and g_χ is tuned to match the observed DM relic abundance for each combination of $m_{Z'}$ and m_χ. Figure taken from the source [44].

suggested by the authors of the Dark Higgs phenomenology paper [44] is to make a running coupling $g_\chi(m_{Z'}, m_s)$ so that DM production in the early universe either matches or is slightly below the observed DM relic abundance. The consequences of applying such a running coupling is shown in fig. 2.10, which demonstrates the vast amount of parameter space that can be covered with a Dark Higgs search along with the effective suppression of the dijet exclusion. Searching for a Dark Higgs is clearly worthwhile in so far as that one can probe parameter space beyond the reach of existing mono-jet and dijet searches. There are also different implications in case of a discovery of either DM or a Dark Higgs, especially if one search is negative while the other search is positive. For convenience and easier comparison, the search results are reported with the conventional couplings, $g_\chi = 1$ and $g_q = 0.25$.

2.3. Heavy Higgs Models

Some theories require the existence of an additional Higgs particle, and the most general way to describe this scenario is called the two-Higgs-doublet model (2HDM). In the following, only the neutral Higgs models are considered for simplicity. The Higgs potential can be written as [50]:

$$\begin{aligned}\mathcal{V}_{gen} =& m_{11}^2\Phi_1^\dagger\Phi_1 + m_{22}^2\Phi_2^\dagger\Phi_2 - m_{12}^2[\Phi_1^\dagger\Phi_2 + h.c.] \\ &+ \frac{1}{2}\lambda_1(\Phi_1^\dagger\Phi)^2 + \frac{1}{2}\lambda_2(\Phi_2^\dagger\Phi_2)^2 + \lambda_3(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2) + \lambda_4(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) \\ &+ \left\{\frac{1}{2}\lambda_5(\Phi_1^\dagger\Phi_2)^2 + \left[\lambda_6(\Phi_1^\dagger\Phi_1) + \lambda_7(\Phi_2^\dagger\Phi_2)\right](\Phi_1^\dagger\Phi_2) + h.c\right\} \end{aligned} \tag{2.10}$$

With Φ_1 being the field of one of the two spin zero particles and Φ_2 being that of the other. They have the masses m_{11} and m_{22} respectively. There is also an off-diagonal mass term, m_{12}^2 that can be complex. The $\lambda_i, i \in [1, 2, 3, ..., 7]$ are coefficients for the four field vertices.

The couplings to quarks are given as [52]:

$$\begin{aligned}-\mathcal{L}_Y =& \frac{1}{\sqrt{2}}\bar{D}\left[\kappa^D \boldsymbol{s}_{\beta-\alpha} + (\rho^D P_R + \rho^{D\dagger} P_L)\boldsymbol{c}_{\beta-\alpha}\right] Dh + \frac{i}{\sqrt{2}}\bar{D}(\rho^D P_R - \rho^{D\dagger} P_L) DA \\ &+ \frac{1}{\sqrt{2}}\bar{D}\left[\kappa^D \boldsymbol{c}_{\beta-\alpha} - (\rho^D P_R + \rho^{D\dagger} P_L)\boldsymbol{s}_{\beta-\alpha}\right] DH \\ &+ \frac{1}{\sqrt{2}}\bar{U}\left[\kappa^U \boldsymbol{s}_{\beta-\alpha} + (\rho^U P_R + \rho^{U\dagger} P_L)\boldsymbol{c}_{\beta-\alpha}\right] Uh - \frac{i}{\sqrt{2}}\bar{U}(\rho^U P_R - \rho^{U\dagger} P_L) UA \\ &+ \frac{1}{\sqrt{2}}\bar{U}\left[\kappa^U \boldsymbol{c}_{\beta-\alpha} - (\rho^U P_R + \rho^{U\dagger} P_L)\boldsymbol{s}_{\beta-\alpha}\right] UH \end{aligned} \tag{2.11}$$

Where h is the lightest scalar particle and H is the heavy scalar particle, A is the pseudoscalar. D has the mass eigenstates for down-type quarks while U has the mass eigenstates for up-type quarks. The matrices $\kappa^U, \kappa^D, \rho^U, \rho^U$ are Yukawa coupling matrices where κ^U and κ^D are real and diagonal by construction and correspond to the mass matrices M_U and M_D by the relations $M_U = \frac{v}{\sqrt{2}}\kappa^U$ and $M_D = \frac{v}{\sqrt{2}}\kappa^D$, where v is the vacuum expectation value of the two Higgs fields. The relations between Yukawa matrices for up-type and down-type quarks respectively

depend on the type of Yukawa couplings selected. Furthermore it is noted that the coefficients $P_R = \frac{1}{2}(1+\gamma^5)$ and $P_L = \frac{1}{2}(1-\gamma^5)$. $\boldsymbol{s}_{\beta-\alpha}$ is short-hand for $\sin(\beta-\alpha)$ whereas $\boldsymbol{c}_{\beta-\alpha}$ means $\cos(\beta-\alpha)$. The value for α is selected so that the relation $-\frac{\pi}{2} \leq \beta-\alpha \leq \frac{\pi}{2}$ is fulfilled. If one wants the 2HDM to work with the minimal supersymmetric extension of the Standard Model (MSSM) [51], it is necessary to select a Yukawa coupling dubbed 'type 2', where $\rho^D = \kappa^D \tan\beta$ and $\rho^U = \kappa^U \cot\beta$. The variable $\tan\beta$ is a physical value for type 2 Yukawa couplings, but for other types of couplings, this is not always the case. In this work, however $\tan\beta$ will be treated as a physical parameter. Since the possibility of an additional (pseudo)scalar heavier than the SM Higgs is investigated in this paper, the mass of the lightest scalar is experimentally constrained. One can from the mass of the pseudoscalar m_A and $\tan\beta$ obtain the following relations [53]:

$$m_A^2 = \frac{m_{12}^2}{\boldsymbol{c}_\beta \boldsymbol{s}_\beta} - \frac{1}{2}v^2(2\lambda_5 - \lambda_6\cot\beta + \lambda_7\tan\beta) \tag{2.12}$$

$$\mathcal{M}^2 = m_A^2 \begin{pmatrix} \boldsymbol{s}_\beta^2 & -\boldsymbol{s}_\beta\boldsymbol{c}_\beta \\ -\boldsymbol{s}_\beta\boldsymbol{c}_\beta & \boldsymbol{c}_\beta^2 \end{pmatrix} + \mathcal{B}^2 \tag{2.13}$$

$$\mathcal{B}^2 = v^2 \begin{pmatrix} \lambda_1\boldsymbol{c}_\beta^2 + 2\lambda_6\boldsymbol{s}_\beta\boldsymbol{c}_\beta + \lambda_5\boldsymbol{s}_\beta^2 & (\lambda_3+\lambda_4)\boldsymbol{s}_\beta\boldsymbol{c}_\beta + \lambda_6\boldsymbol{c}_\beta^2 + \lambda_7\boldsymbol{c}_\beta^2 \\ (\lambda_3+\lambda_4)\boldsymbol{s}_\beta\boldsymbol{c}_\beta + \lambda_6\boldsymbol{c}_\beta^2 + \lambda_7\boldsymbol{c}_\beta^2 & \lambda_2\boldsymbol{s}_\beta^2 + \lambda_7\boldsymbol{s}_\beta\boldsymbol{c}_\beta + \lambda_5\boldsymbol{c}_\beta^2 \end{pmatrix} \tag{2.14}$$

By diagonalising $\mathcal{M}^2$ with the rotation matrix $\mathcal{R}(\alpha)$, the following relations are obtained [53]:

$$\mathcal{R}(\alpha) = \begin{pmatrix} \boldsymbol{c}_\alpha & -\boldsymbol{s}_\alpha \\ \boldsymbol{s}_\alpha & \boldsymbol{c}_\alpha \end{pmatrix} \tag{2.15}$$

$$\begin{pmatrix} m_H^2 & 0 \\ 0 & m_h^2 \end{pmatrix} = \mathcal{R}(\alpha)\mathcal{M}^2\mathcal{R}^T(\alpha) \tag{2.16}$$

$$m_{H,h}^2 = \frac{1}{2}\left[\mathcal{M}_{11}^2 \pm \sqrt{(\mathcal{M}_{11}^2 - \mathcal{M}_{22}^2)^2 - 4(\mathcal{M}_{12}^2)^2}\right] \tag{2.17}$$

$$\boldsymbol{s}_{2\alpha} = \frac{2\mathcal{M}_{12}^2}{\sqrt{(\mathcal{M}_{11}^2 - \mathcal{M}_{22}^2)^2 - 4(\mathcal{M}12^2)^2}},$$

$$\boldsymbol{c}_{2\alpha} = \frac{\mathcal{M}_{11}^2 - \mathcal{M}_{22}^2}{\sqrt{(\mathcal{M}_{11}^2 - \mathcal{M}_{22}^2)^2 - 4(\mathcal{M}12^2)^2}} \tag{2.18}$$

The mass relation from eq. (2.17) can be used to set the remaining parameters, given that m_h is experimentally known.

From eq. (2.11) one can obtain the coupling to top quarks and the decay width into top quarks. They are given as [52][50]:

$$g_{At\bar{t}} = -\,i\cot\beta\gamma_5 \tag{2.19}$$

$$g_{Ht\bar{t}} = \boldsymbol{c}_{\beta-\alpha} - \cot\beta\boldsymbol{s}_{\beta-\alpha} = \frac{\boldsymbol{s}_\alpha}{\boldsymbol{s}_\beta} \tag{2.20}$$

$$\Gamma_{A\to t\bar{t}} = \frac{3m_A m_t^2}{8\pi v^2}|g_{At\bar{t}}|^2\sqrt{1-4\frac{m_t^2}{m_A^2}} \tag{2.21}$$

$$\Gamma_{H\to t\bar{t}} = \frac{3m_H m_t^2}{8\pi v^2}\left[1-\left(\frac{m_t}{m_H}+\frac{m_{\bar{t}}}{m_H}\right)^2\right]|g_{Ht\bar{t}}|^2\sqrt{1-4\frac{m_t^2}{m_H^2}} \tag{2.22}$$

It is observed that the couplings to top quarks and the decay widths of the $\Phi \to t\bar{t}$ processes are directly dependent on $\tan\beta$. A QCD correction can also be applied to the result in eq. (2.21) and eq. (2.22) based on the running coupling ($\alpha_s(m_\Phi)$), where m_Φ is either m_A or m_H, that takes the form [50]:

$$\Gamma = \Gamma_0\left[1+5.67\frac{\alpha_s(m_\Phi)}{\pi}+(35.94-1.36n_f)\left(\frac{\alpha_s(m_\Phi)}{\pi}\right)^2\right] \tag{2.23}$$

The parameter n_f refers to the selected flavour scheme for quarks that treats the given number of lightest quarks as massless. The running coupling $\alpha_s(m_\Phi)$ goes to leading order as $\alpha_s(m_\Phi) \propto \frac{1}{\log(m_\Phi)}$ [54].

2.3.1. Heavy Higgs in $t\bar{t}$ Production

With the interplay of decay widths, couplings and masses in the 2HDM, we can move on to the topic of the impact of a Heavy Higgs on $t\bar{t}$ production in a collider experiment. The $t\bar{t}$ production process can originate directly from gluons or via a Heavy Higgs as shown in fig. 2.11. Interference also occurs, and the differential cross-sections at leading order are given as [55]:

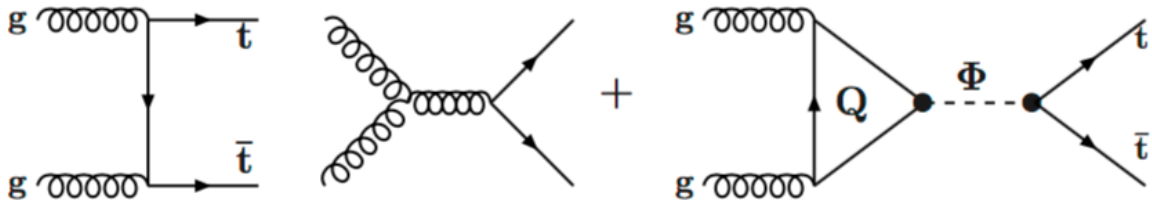


Figure 2.11.: Feynman diagrams at leading order for $t\bar{t}$ production in QCD processes (left) and the $gg \to \Phi \to t\bar{t}$ process (right), from the source [55].

$$\frac{\mathrm{d}\hat{\sigma}_B}{\mathrm{d}z} = \frac{\pi\alpha_s^2}{6\hat{s}}\hat{\beta}_t\left(\frac{1}{1-\hat{\beta}_t^2 z^2} - \frac{9}{16}\right)\left[3 + \hat{\beta}_t^2 z^2 - 2\hat{\beta}_t^2 - \frac{2(1-\hat{\beta}_t^2)^2}{1-\hat{\beta}_t^2 z^2}\right] \tag{2.24}$$

$$\frac{\mathrm{d}\hat{\sigma}_S}{\mathrm{d}z} = \frac{3\alpha_s^2 G_F^2 m_t^2}{8192\pi^3}\hat{s}^2 \sum_{\Phi} \frac{|\hat{\beta}_t^{p\Phi}\hat{g}_{\Phi t\bar{t}}\sum_Q \hat{g}_{\Phi Q\bar{Q}} A_{1/2}^{\Phi}(\hat{\tau}_Q)|^2}{(s-M_\Phi^2)^2 + \Gamma_\Phi^2 M_\Phi^2} \tag{2.25}$$

$$\frac{\mathrm{d}\hat{\sigma}_I}{\mathrm{d}z} = -\frac{\alpha_s^2 G_F m_t^2}{64\sqrt{2}\pi}\frac{1}{1-\hat{\beta}_t^2 z^2}\mathrm{Re}\left[\sum_{\Phi}\frac{\hat{\beta}_t^{p\Phi}\hat{g}_{\Phi t\bar{t}}\sum_Q \hat{g}_{\Phi Q\bar{Q}} A_{1/2}^{\Phi}(\hat{\tau}_Q)}{s - M_\Phi^2 + i\Gamma_\Phi M_\Phi}\right], \tag{2.26}$$

where $z = \cos\theta$ where θ is the scattering angle, $\beta_t = \sqrt{1 - \frac{m_t^2}{\hat{s}}}$, $p_\Phi = 1(3)$ for pseudoscalar(scalar) bosons and the couplings $\hat{g}_{\Phi t\bar{t}}$ and $\hat{g}_{\Phi QQ}$ are normalised to the SM Higgs coupling, $\hat{g}_{Hff}^{\mathrm{SM}} = m_f/v$. The differential cross-sections are given as $\frac{\mathrm{d}\hat{\sigma}_B}{\mathrm{d}z}$ for the background $t\bar{t}$ production, $\frac{\mathrm{d}\hat{\sigma}_S}{\mathrm{d}z}$ is the differential cross-section for the signal resonance and $\frac{\mathrm{d}\hat{\sigma}_I}{\mathrm{d}z}$ is the differential cross-section for the interference.

The form factor $A_{1/2}(\hat{\tau}_Q)$, is a complex parameter and a function of $\hat{\tau}_Q = \frac{m_\Phi^2}{4m_Q^2}$. The form factor is given by the following equations [55]:

$$A_{1/2}^H(\tau) = 2[\tau + (\tau-1)f(\tau)]\tau^{-2} \tag{2.27}$$

$$A_{1/2}^A(\tau) = 2\tau^{-1}f(\tau) \tag{2.28}$$

$$f(\tau) = \begin{cases} \arcsin^2\sqrt{\tau} & \text{for } \tau \leq 1 \\ -\frac{1}{4}\left[\log\frac{1+\sqrt{1-\tau^{-1}}}{1-\sqrt{1-\tau^{-1}}} - i\pi\right]^2 & \text{for } \tau > 1 \end{cases} \tag{2.29}$$

In the interference, the imaginary part of $A_{1/2}(\hat{\tau}_Q)$ causes destructive interference proportional to $\Gamma_\Phi M_\Phi$. This means that one can expect a relatively stronger interference for hypotheses with a high width for the (pseudo)scalar.

Bibliography

[1] A. Salam and J. C. Ward, *'Weak and electromagnetic interactions'*, Il Nuovo Cimento 11 (February, 1959) 568-577, doi:10.1007/BF02726525.

[2] S. Weinberg, *'A Model of Leptons'*, Phys. Rev. Lett. 19 (1967) 1264-1266, doi:10.1103/PhysRevLett.19.1264.

[3] J. C. Collinset al., *'Factorization of Hard Processes in QCD'*, Adv. Ser. Direct. High Energy Phys. 5 (1989) 1-91, doi:10.1142/9789814503266_0001, arXiv:hep-ph/0409313.

[4] **CMS Collaboration**, *'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC'*, Phys. Lett. B716 (2012) 30-61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[5] **ATLAS Collaboration**, *'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'*, Phys. Lett. B716 (2012) 1-29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[6] P. W. Higgs, *'Broken Symmetries and the Masses of Gauge Bosons'*, Phys. Rev. Lett. 13 (1964) 508-509, doi:10.1103/PhysRevLett.13.508.

[7] G. S. Guralnik et al., *'Global Conservation Laws and Massless Particles'*, Phys. Rev. Lett. 13 (1964) 585-587, doi:10.1103/PhysRevLett.13.585.

[8] F. Englert et al., *'Broken Symmetry and the Mass of Gauge Vector Mesons'*, Phys. Rev. Lett. 13 (1964) 321-323, doi:10.1103/PhysRevLett.13.321.

[9] H. Yukawa, *'On the Interaction of Elementary Particles I'*, Proc. Phys. Math. Soc. Jap. 17 (1935) 48-57, doi:10.1143/PTPS.1.1.

[10] M. Peskin, D. Schroeder, *'An Introduction to Quantum Field Theory'*, Westview Press, ISBN 978-0-201-50397-5.

[11] M. Kobayashi and T. Maskawa, *'CP Violation in the Renormalizable Theory of Weak Interaction'*, Prog. Theor. Phys. 49 (1973) 652-657, doi:10.1143/PTP.49.652.

[12] S. Kraml et al., E. Accomando et al., *'CP Studies and Non-Standard Higgs Physics'*, CERN-2006-009, arXiv:hep-ph/0608079.

[13] S. Weinberg, *'Implications of Dynamical Symmetry Breaking'*, Phys. Rev. D13 (1976) 974?996.

[14] L. Susskind, *'Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory'*, Phys. Rev. D20 (1979) 2619?2625.

[15] E. Gildener, *'Gauge Symmetry Hierarchies'*, Phys. Rev. D14 (1976) 1667.

[16] H. E. Haber, G. L. Kane, *'The Search for Supersymmetry: Probing Physics Beyond the Standard Model'*, Phys. Rept. 117 (1985) 75-263, doi: 10.1016/0370-1573(85)90051-1.

[17] J. M. Campbell et al., *'Hard Interactions of Quarks and Gluons: A Primer for LHC Physics'*, Rept. Prog. Phys. 70 (2007) 89, doi:10.1088/0034-4885/70/1/R02, arXiv:hep-ph/0611148.

[18] R. D. Ball et al., *'Parton distributions from high-precision collider data'*, Eur. Phys. J. C 77 (2017), p. 663. doi:10.1140/epjc/s10052-017-5199-5. arXiv:1706.00428 [hep-ph].

[19] **ATLAS, CMS Collaboration**, *'Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s} = 7$ *and 8 TeV'*, JHEP08(2016)045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266 [hep-ex].

[20] **CMS Collaboration**, *'Observation of Higgs boson decay to bottom quarks'*, CMS-HIG-18-016, Phys. Rev. Lett. 121, 121801 (2018), doi:10.1103/PhysRevLett.121.121801, arXiv:1808.08242 [hep-ex].

[21] **ATLAS Collaboration**, *'Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector'*, Phys. Lett. B 786 (2018) 59, doi:10.1016/j.physletb.2018.09.013, arXiv:1808.08238 [hep-ex].

[22] D. de Florian et al., *'Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector'*, CERN Yellow Reports: Monographs Volume 2/2017 (CERN–2017–002-M), doi:10.23731/CYRM-2017-002, arXiv:1610.07922 [hep-ph].

[23] C. Anastasiou et al., *'High precision determination of the gluon fusion Higgs boson cross-section at the LHC'*, JHEP 05 (2016) 058, arXiv:1602.00695 [hep-ph].

[24] **ATLAS Collaboration**, *'Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector'*, Phys. Lett. B 784 (2018) 173, doi:10.1016/j.physletb.2018.07.035, arXiv:1806.00425.

[25] **CMS Collaboration**, *'Observation of ttH production'*, Phys. Rev. Lett. 120 (2018) 231801, doi:10.1103/PhysRevLett.120.231801, arXiv:1804.02610.

[26] A. Denner et al., *'Standard Model Higgs-Boson Branching Ratios with Uncertainties'*, doi:10.1140/epjc/s10052-011-1753-8, arXiv:1107.5909 [hep-ph].

[27] **ATLAS Collaboration**, *'Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector'*, Phys. Rev. D 92 (2015) 012006, doi:10.1103/PhysRevD.92.012006, arXiv:1412.2641.

[28] **CMS Collaboration**, *'Measurement of the properties of a Higgs boson in the four-lepton final state'*, Phys. Rev. D 89 (2014) 092007, doi:10.1103/PhysRevD.89.092007, arXiv:1312.5353.

[29] **ATLAS Collaboration**, *'Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector'*,
Phys. Rev. D 91 (2015) 012006, doi:10.1103/PhysRevD.91.012006, arXiv:1408.5191.

[30] **CMS Collaboration**, *'Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs'*,
Phys. Rev. Lett. 110 (2013) 081803, doi:10.1103/PhysRevLett.110.081803, arXiv:1212.6639.

[31] **ATLAS Collaboration**, *'Evidence for the spin-0 nature of the Higgs boson using ATLAS data'*,
Phys. Lett. B 726 (2013) 120, doi:10.1016/j.physletb.2013.08.026, arXiv:1307.1432.

[32] M. Czarkon et al., *'Top++: a program for the calculation of the top-pair cross-section at hadron colliders'*,
Comput.Phys.Commun. 185 (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675v4.

[33] S. Dulat et al., *'New parton distribution functions from a global analysis of quantum chromodynamics'*, Phys. Rev. D 93 (2016) 033006,
doi:10.1103/PhysRevD.93.033006, arXiv:1506.07443.

[34] M. Tanabashi et al. (Particle Data Group), *'43. Monte Carlo Particle Numbering Scheme'*,
Phys. Rev. D 98, 030001 (2018) and 2019 update, pdg.lbl.gov/2019/reviews/rpp2019-rev-monte-carlo-numbering.pdf

[35] A. Czarnecki et al., *'Lepton spectra from decays of polarized top quarks'*,
Nuclear Physics B 351 (1991), no. 1, 70 ? 80, doi:10.1016/0550-3213(91)90082-9.

[36] M. Jezabek, *'Top quark physics'*, Nucl. Phys. Proc. Suppl. 37B (1994), no. 2, 197, doi:10.1016/0920-5632(94)90677-7, arXiv:hep-ph/9406411.

[37] G. Bertone et al., *'Particle Dark Matter: Evidence, Candidates and Constraints'*,

Phys.Rept.405:279-390,2005, doi:10.1016/j.physrep.2004.08.031, arXiv:hep-ph/0404175.

[38] D. Clowe et al., *'A direct empirical proof of the existence of dark matter'*, Astrophys.J.648:L109-L113,2006, doi:10.1086/508162, arXiv:astro-ph/0608407.

[39] **PandaX-4T Collaboration**, *'Dark matter search results from the PandaX-4Tcommissioning run'*, Phys. Rev. Lett.127(2021) 261802, doi:10.1103/PhysRevLett.127.261802, arXiv:2107.13438.

[40] **CMS Collaboration**, *'Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at $\sqrt{s}$ = 13 TeV'*, CMS-HIG-20-003, arXiv:2201.11585 [hep-ex].

[41] **Fermi-LATCollaboration**, M. Ackermann et al., *'Limits on Dark Matter Annihilation Signals from the Fermi LAT 4-year Measurement of the Isotropic Gamma-Ray Background'*, JCAP1509(2015), no. 09 008, [arXiv:1501.05464].

[42] L. Roszkowski et al., *'WIMP dark matter candidates and searches - current status and future prospects'*, Rept.Prog.Phys. 81 (2018) no.6, 066201, doi:10.1088/1361-6633/aab913, arXiv:1707.06277 [hep-ph].

[43] M. Duerr et al., *'How to save the WIMP: global analysis of a dark matter model with two s-channel mediators'*, JHEP 09 (2016) 042, [1606.07609].

[44] M. Duerr et al., *'Hunting the dark Higgs'*, JHEP 1704 (2017) 143, doi:10.1007/JHEP04(2017)143, arXiv:1701.08780v2.

[45] D. Abercrombie et al., *'Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum'*, Phys. Dark Univ. 26 (2019) 100371, doi:10.1016/j.dark.2019.100371, arXiv:1507.00966.

[46] **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), doi:10.1103/PhysRevD.97.092005, arXiv:1712.02345.

[47] **CMS Collaboration**, *'Search for dark matter in final states with an energetic jet, or a hadronically decaying W or Z boson using 12.9* fb^{-1} *of data at* $\sqrt{s} = 13$ *TeV'*,
JHEP 07 (2017) 014, doi:10.1007/JHEP07(2017)014, arXiv:1703.01651v2.

[48] M. Fairbairn et al., *'Constraints on Z models from LHC dijet searches and implications for dark matter'*, JHEP 09 (2016) 018, doi:10.1007/JHEP09(2016)018, arXiv:1605.07940v2.

[49] **Planck Collaboration**, P. A. R. Ade et al., *'Planck 2015 results. XIII. Cosmological parameters'*,
Astron. Astrophys. 594 (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589v3.

[50] D. Ericsson et al., *'2HDMC - Two-Higgs-Doublet Model Calculator'*,
Comput.Phys.Commun.181:189-205,2010, doi:10.1016/j.cpc.2009.09.011, arXiv:0902.0851v2.

[51] A. Djouadi et al., *'The Minimal Supersymmetric Standard Model: Group Summary Report'*,
Report of the MSSM working group for the Workshop 'GDR-Supersymmetrie', Dec 1998, arXiv:hep-ph/9901246.

[52] S. Davidson et al., *'Basis-independent methods for the two-Higgs-doublet model'*,
Phys.Rev.D72:035004,2005; Erratum-ibid.D72:099902,2005, doi:10.1103/PhysRevD.72.035004, doi:10.1103/PhysRevD.72.099902, arXiv:hep-ph/0504050.

[53] J. Gunion et al., *'The CP-conserving two-Higgs-doublet model: the approach to the decoupling limit'*,
Phys.Rev.D67:075019,2003, doi: 10.1103/PhysRevD.67.075019, arXiv:hep-ph/0207010.

[54] A. Deur et al., *'The QCD Running Coupling'*,

Prog. Part. Nuc. Phys. 90 1 (2016), doi:10.1016/j.ppnp.2016.04.003, arXiv:1604.08082v3.

[55] A. Djouadi et al., *'Interference Effects in the Decays of Spin-Zero Resonances into $\gamma\gamma$ and $t\bar{t}$'*,
JHEP 07 (2016) 105, doi:10.1007/JHEP07(2016)105, arXiv:1605.00542v2.

3. Experimental Setup

This work is based on analyses performed on data taken from the Compact Muon Solenoid (CMS) experiment situated at a collision point in the Large Hadron Collider (LHC) which is a hadron accelerator and collider with a circumference of 27 km, designed to produce hadron collisions at a centre of mass energy up to 14 TeV. The LHC is owned and operated by the European Council for Nuclear Research (CERN) as the final stage of a chain of accelerators with increasing energies as shown in fig. 3.1. The LHC has a total of four collision points, each fitted with a different experiment. Two of these experiments are general purpose detectors which primarily measure proton-proton scattering at a centre of mass energy of up to 14 TeV. They are called CMS and ATLAS. These two experiments operate independently of each other in order to be able to verify new discoveries or constraints on parameters for discovered particles that are not yet fully established.

3.1. Detector Elements of the Compact Muon Solenoid

The Compact Muon Solenoid is commonly referred to by its acronym CMS. As the name indicates, the detector has a solenoidal magnet and a muon spectrometer that by weight and volume are the largest components of the CMS detector. The Solenoidal magnet is designed to provide a continuous magnetic field of 4 T using superconductivity to generate the magnetic field. The CMS detector spans a length of 21.6 m and has a diameter of 14.6 m. The total weight of the CMS experiment is approximately 12500 metric tonnes [2]. Given the dimensions of the detector, one may wonder what is compact by CMS, but it has to do with the detector to be compact compared to other designs for general purpose detectors for hadron collisions at energies up to 14 TeV. The CMS detector is divided into three sections along the beam with the central region called the barrel region, followed by the endcap and ultimately the Very-forward Calorimeter (HF) to catch

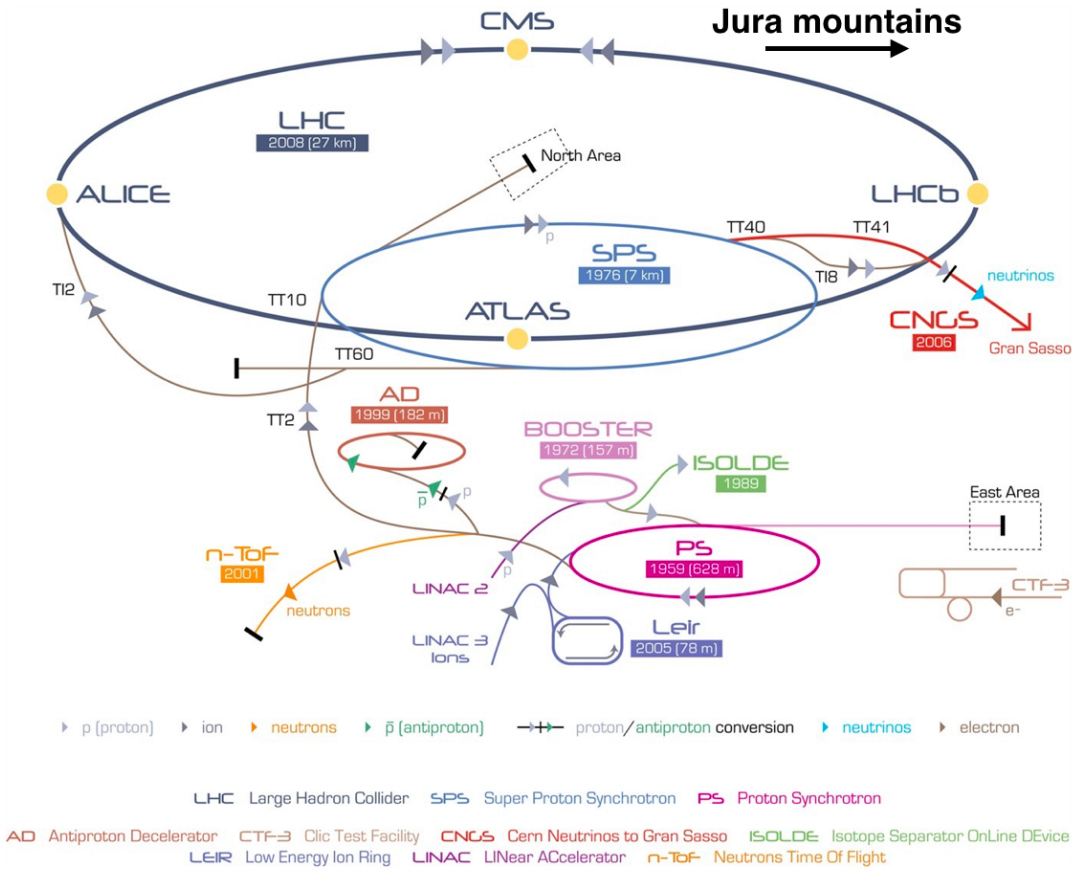


Figure 3.1.: Overview the CERN accelerator complex with the chain of accelerators leading up to the LHC. Taken from the source [1].

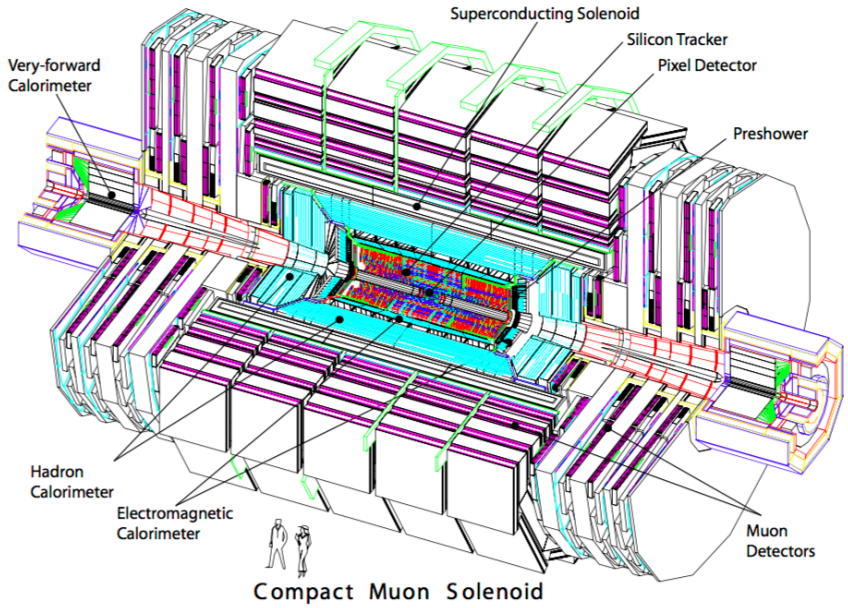


Figure 3.2.: Overview of the detector segments of CMS, taken from source [2].

the particles with the closest alignment to the beam as possible. CMS operates with the coordinates z, η and ϕ, where z is the beam axis with $z = 0$ at the collision point, pointing towards the Jura mountains from the location of the CMS detector (see fig. 3.1). The coordinate η is the pseudorapidity defined as $\eta = -\log[\tan(\theta/2)]$, where θ is the angle between the line from the interaction point of the beam to a given part of the detector and the beam axis where $z > 0$. ϕ is the azimuthal angle around the beam axis with $\phi = 0$ being the horizontal part of the detector pointing towards the centre of the LHC.

3.1.1. Tracker

The tracker system is the innermost layer of the CMS detector. It is divided in two segments called the pixel detector and the silicon strip tracker. A layout of the tracker is seen in figure 3.3. The silicon strip tracker is made entirely of silicon strip modules with a pitch ranging from 80 μm to 184 μm leading to a resolution in the ϕ direction ranging from 23 μm to 53 μm. Some of the modules are double and back to back in stereo angle to obtain a single point resolution of 230 μm and 530 μm along the z-axis for modules in the barrel and on the radial axis for

modules at the endcaps and the inner discs. The pixel detector contains pixels of

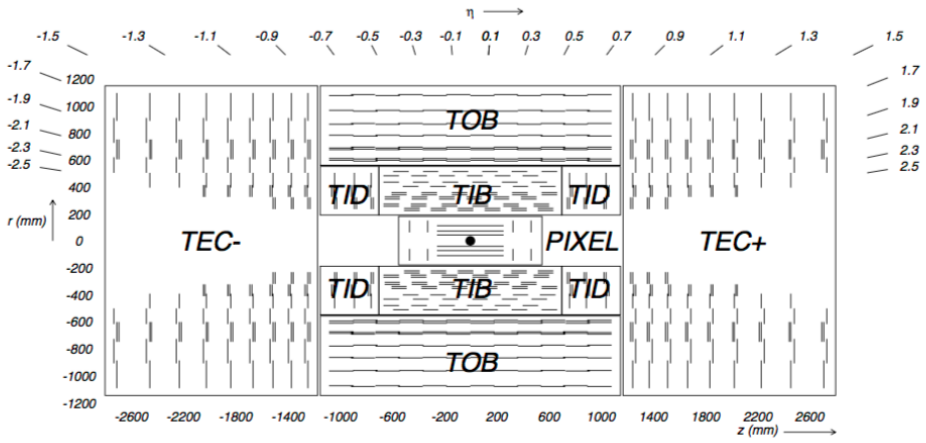


Figure 3.3.: Layout of the tracker divided into the segments: Tracker endcap (TEC), tracker outer barrel (TOB), tracker inner barrel (TIB), the tracker inner discs (TID) and the pixel detector (PIXEL). Taken from the source [3].

the size $100 \times 150 \mu m^2$ set up to cover a range in pseudorapidity up to $|\eta| = 2.5$. In the extended technical stop between 2016 and 2017, the pixel detector has been replaced and the modifications are seen in fig. 3.4. The number of layers of pixel detectors in the barrel region have increased from three to four while the forward layers have increased from two to three.

3.1.2. Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) consists of blocks of lead tungstate ($PbWO_4$) crystals that have a high density leading to a radiation length $X_0 = 0.89$ cm and are scintillating. They are arranged in the detector as shown in fig. 3.5 covering a range in pseudorapidity up to $|\eta| = 3.0$. In the barrel region, each crystal has the dimensions, $21.8 \times 21.8 \times 230$mm, whereas in the endcaps the dimensions are slightly different measuring $28.62 \times 28.62 \times 220$mm. The crystals in the barrel are arranged in supermodules with 85×20 crystals in each. The supermodules are subdivided into five modules to facilitate assembly. The crystals in the endcap are arranged in clusters of 5×5 crystals called supercrystals. The resolution of the ECAL in the barrel in terms η and ϕ is 0.0175×0.0175, whereas in the endcap

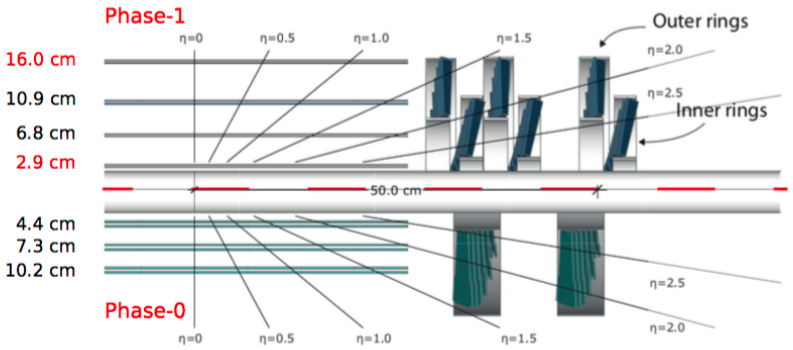


Figure 3.4.: Comparison between the pixel detector before replacement (Phase-0) and after replacement (Phase-1), taken from the sources [4][5].

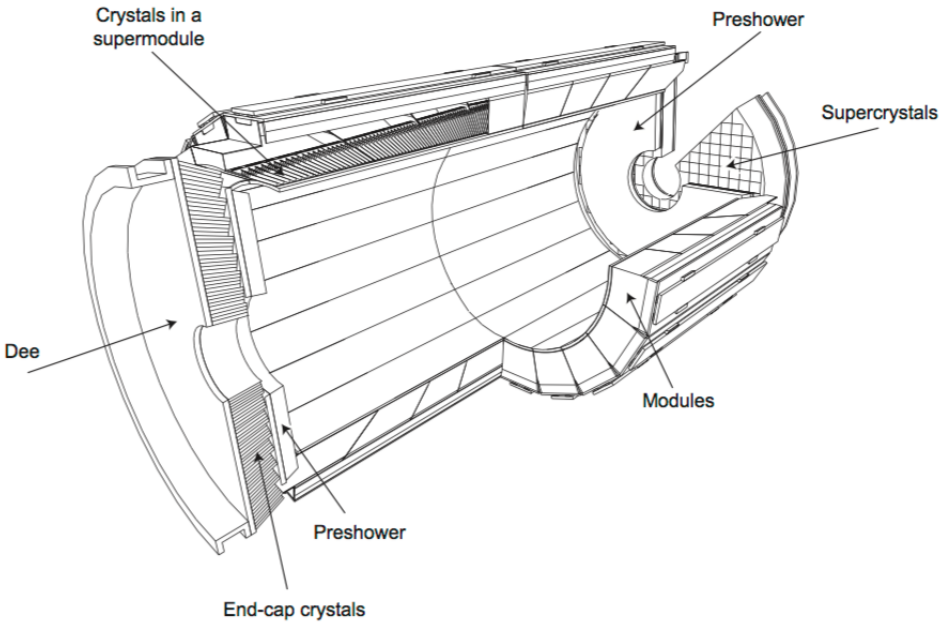


Figure 3.5.: Overview of the CMS electromagnetic calorimeter, taken from the source [2].

the resolution gradually decreases to 0.05×0.05. Pre-showering modules are also in place covering the parts of the endcap within the range of $1.653 < |\eta| < 2.61$ of the ECAL as seen in fig. 3.5. At energies below 500 GeV, the energy resolution in the ECAL follows the following equation:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{3.1}$$

where E is the shower energy, σ is the standard error, S is the stochastic term, N is the noise term and C is the constant term. The stochastic term has to do with fluctuations in shower developments and has a typical value of 2.8%. The noise term has to do with noise from pileup and electronics and the like and has a typical value of 0.12%. The constant noise term has a typical value of 0.30% and covers non-uniformities, calibration errors and energy leakage.

3.1.3. Hadronic Calorimeter

The hadronic calorimeter (HCAL) consists of the following modules: The barrel HCAL (HB) covering the range in pseudorapidity up to $|\eta| = 1.3$, the endcap HCAL (HE), the outer HCAL (HO) and the forward calorimeter (HF). The HB is made of brass layers with tiles of plastic scintillator in between. For structural support, the innermost layer as well as the outermost layer are made of steel. The HE is made entirely from brass and plastic scintillator in interchanging layers. The HO uses the solenoid magnet as absorber and consists of tiles of plastic scintillator placed at a radial distance from the beam of 4.07 m. In the region around $z = 0$, where the effective material budget is lowest, an additional piece of iron and an additional scintillator tile underneath are put between the scintillator tile and the solenoid magnet coil. The setup of the HB, HE and HO is shown in fig. 3.6. The granularity of the HB and the HE for $|\eta| < 1.6$ in terms of $\eta \times \phi$ is 0.087×0.087, whereas in the HE the granularity at $|\eta| \geq 1.6$ is 0.17×0.17.

The HF uses steel as absorber and quartz fibres as active material, and enables a range in pseudorapidity from $3 < |\eta| < 5.2$. The HF is designed for radiation hardness and uses Cherenkov based signal detection.

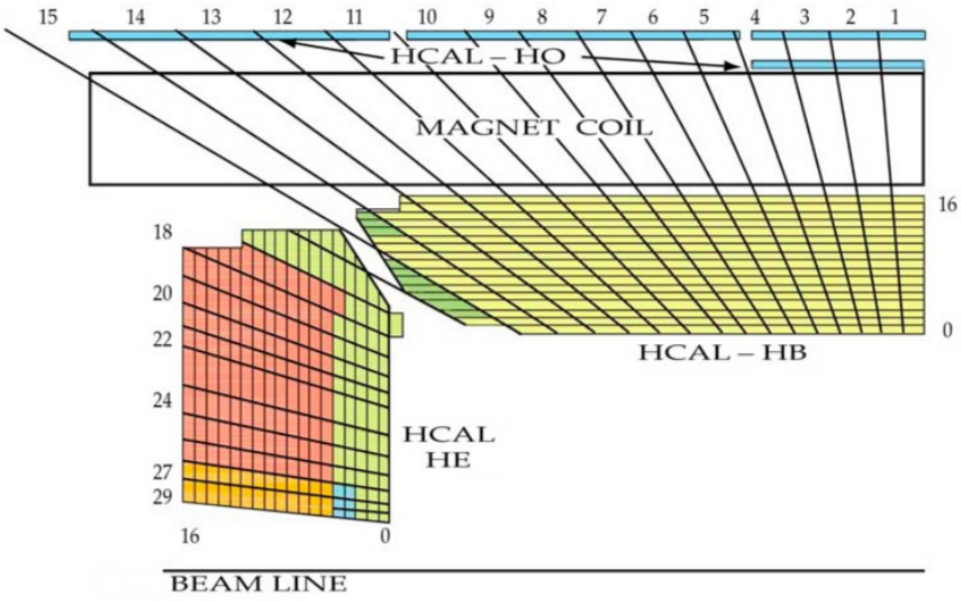


Figure 3.6.: Radial slice of the CMS hadronic calorimeters: HB, HE and HO. Figure taken from the source [2].

3.1.4. Muon Spectrometer

The iron return yoke serves as absorptive material for the muon spectrometer in the barrel region. To detect the muon tracks in the barrel region, drift tube modules (DT) have been installed between the gaps of the return yoke. Resistive plate chambers (RPC) are mounted on the DTs to provide adequate time resolution for the triggers. For the endcap, covering the range $0.9 < |\eta| < 2.4$, the tracking is done mainly with cathode strip chambers (CSC). RPCs are placed in the endcap as well covering a range in pseudorapidity up to $|\eta| = 1.6$. A schematic over the muon spectrometer is shown in fig. 3.7.

3.2. CMS Data Collection

The LHC beams consist of proton bunches, each 25 ns apart. Where the two beams intersect, collisions occur when two bunches cross each other. The rate of collisions is measured as instantaneous luminosity and is given in units of inverted cross-section per time unit. In 2017, the CMS has measured a peak instantaneous

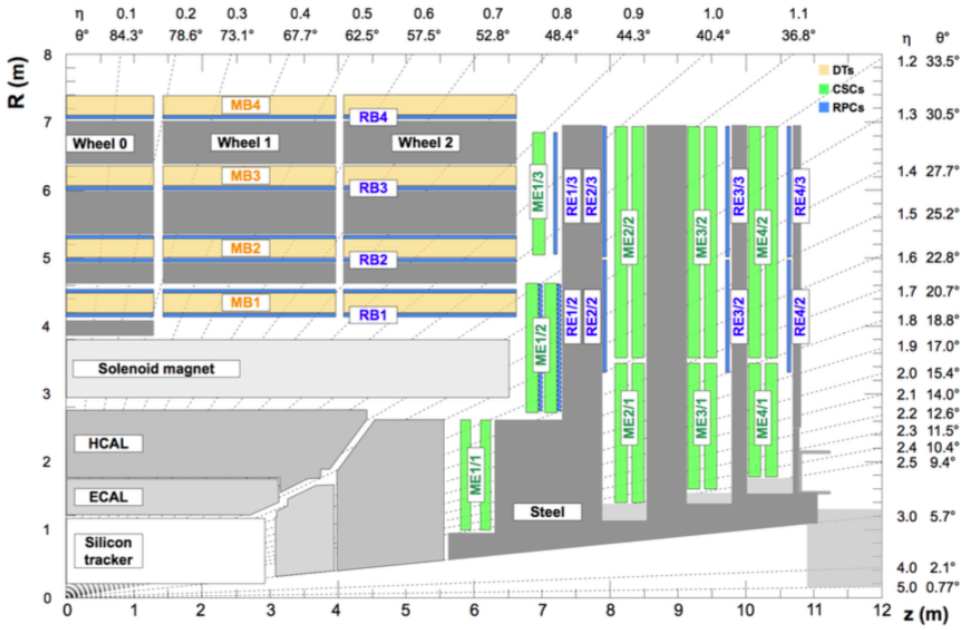


Figure 3.7.: Radial slice of the CMS showing the muon spectrometer with the drift tube modules (MB) and the endcap CSC modules (ME) along with the RPCs (RB in the barrel and RE in the endcap). Figure taken from the source [6].

luminosity of 20.88 nb^{-1}/s [7]. For each bunch crossing, the aim is to have a collision with as high momentum transfer as possible. To increase the rate of collisions with high momentum transfer, several collisions are produced per bunch crossing. The collision with the highest momentum transfer is called the hard collision while the remaining collisions in the given bunch crossing are called pileup. This pileup can potentially interfere with the measurement of the products of the hard collision, since they happen simultaneously. The trick is therefore to have as effective methods as possible to remove interference from pileup while preserving the data from the hard collision. The mean value for pileup in 2016 is 23 while for 2017 and 2018 it is 32 [8].

The rate of data collection of the CMS detector can reach up to 40 MHZ, generating an impractical rate and amount of data if one decides to store all detected events. Therefore a set of triggers are in place to reduce the event rate while minimising the amount of data lost on rare processes. This is done in two trigger levels, the level 1 trigger and the high level trigger (HLT).

3.2.1. Level 1 Trigger

The purpose of the level 1 trigger (L1) is to reduce the event rate from 40 MHz to $\sim$ 100 kHz. With each event containing $\sim$ 1 MB of data, the outgoing data rate will still be around 100 GB/s. With the incoming rate being very high, the real time computations of the L1 trigger have to be as simple and fast as possible. Instead of reading the full detector output, the L1 trigger only reads the output from the calorimeters and the muon stations [9]. Since the transverse momentum (p_{T}) of jets is a key parameter in determining which events are interesting and which are not, it is important to have this information evaluated as accurately as possible. For this purpose, corrections are applied by the use of lookup tables to apply the corrections within the time window provided. To produce this lookup table, the uncorrected L1 jets are matched with fully corrected reference jets. As a service task for the CMS experiment, I have performed such jet energy calibrations of the L1 trigger. This service task is presented in chapter 5.

3.2.2. High Level Trigger

The high level trigger (HLT) further reduces the event rate from $\sim$ 100 kHz down to a rate of $\sim$ 1 kHz. The way the HLT reduces the event count is by reconstructing

tracks in the regions of interest indicated by the L1 trigger [10] and using the track information to determine whether the L1 trigger conditions are met based on the additional information. Information from the tracks is also used to identify rare particles like b-quarks and tauons that indicate events of particular interest and are therefore much more likely to pass a given set of the HLT requirements. The conditions of firing the HLT can be adjusted and there are around 100 such conditions of which at least one has to be fulfilled for the event to be written on disc. Sometimes it is desirable to have loose requirements in the HLT, where the event rate is kept within the limit by the means of prescaling instead. A higher prescale number will reduce the rate at which events are stored on disc when the HLT is fired, with a prescale setting of one storing all events that fire the HLT. A prescale setting of zero will reject all events. The trigger conditions are then grouped in different streams and ultimately primary datasets. In the case of multiple trigger conditions being fulfilled, the event will occur in the primary datasets where the trigger conditions are allocated. Care needs to be taken in order to avoid double counting when one uses multiple primary datasets that are not mutually exclusive. The computation time per event to decide whether to accept or reject the event is of the order of 100 ms.

3.2.3. Offline Data Treatment

The data from the HLT is stored for further processing, where analysis is performed to make the detector output into analysis objects. The software used for reconstruction and analysis is called CMS software (CMSSW). The reconstructed data is available in different formats depending on the needs of the analyser with for example the RAW format which contains detector output and decisions from the L1 trigger and HLT. The analyses in this work have mainly used the MINIAOD format, where AOD is an abbreviation for Analysis Object Data. The object reconstructions for the objects used in analysis in this work are discussed in Chapter 4.

Bibliography

[1] **CERN**, *'The particle suppliers, Les fournisseurs de particules'*, CERN Bulletin (2010), BUL-NA-2010-077.

[2] **CMS Collaboration**, *'The CMS experiment at the CERN LHC'*, 2008 JINST 3 S08004.

[3] **CMS Collaboration**, *'CMS Tracker Technical Design Report'*, CMS-TDR-5.

[4] **CMS Collaboration**, *'CMS Technical Design Report for the Pixel Detector Upgrade'*, CMS-TDR-11.

[5] M. Lipinsky, **CMS Collaboration**, *'The Phase-1 Upgrade of the CMS Pixel Detector'*, CMS CR-2017/135.

[6] **CMS Collaboration**, *'Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at* $\sqrt{s} = 13$ *TeV'*, JINST 13 (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[7] **CMS Collaboration**, *'CMS Luminosity - Public Results'*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

[8] **CMS Collaboration**, *'Pileup mitigation at CMS in 13 TeV data'*, CMS PAS JME-18-001, 23/07/2019

[9] **CMS Collaboration**, *'CMS Technical Design Report for the Level-1 Trigger Upgrade'*, Technical Report CERN-LHCC-2013-011. CMS-TDR-12, Jun, 2013.

[10] **CMS Collaboration**, P. Sphicas, ed., *'CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger'*, 2002.

4. Object Reconstruction

4.1. Particle Flow Objects

The first step of reconstructing particles is to generate particle flow (PF) objects where hits in the different parts of the detector are located and fitted into tracks in the pixel detector, tracker and muon chambers or clusters in the calorimeters [1]. These objects are then linked together to match criteria set for particle candidates of different types.

4.1.1. Track Fitting

Charged particles leave hits in the tracker that are fitted together with a Kalman Filter in three stages: Initial seed generation where the first few hits are found to be compatible with the trajectory of a charged particle. The following stage attempts to build the full trajectory from the hits in the tracker. This process is called pattern recognition. The final stage is fitting to determine the properties of the charged particles. A global combinatorial track finder requires at least two consecutive hits in the pixel detector and a total of eight hits with up to one missing hit along the trajectory. Furthermore, the track has to originate within a few mm of the beam axis and have $p_{\mathrm{T}} > 0.9$ GeV. To enhance tracking efficiency, an additional set of tracking iterations are available that still keep the misreconstruction rate at a comparable level. These additional iterations require at least three hits in the tracker and $p_{\mathrm{T}} > 200$ MeV along with quality criteria to maintain the low misreconstruction rate.

4.1.2. Vertex Reconstruction

In a bunch crossing event, several proton-proton collisions are expected, but the focus of attention is on the collision with the highest energy transfer, the rest of the collision events in a single bunch crossing are called pileup. To reconstruct the proton interaction points in the bunch crossing of the beams, a process called vertex reconstruction is performed [2]. These candidates for proton interaction points are called primary vertices to separate them from vertices originating from particle decays outside the beam region. The reconstructed tracks are extrapolated and the point of closest approach to the beam is established. To establish when to group several tracks in one vertex, and when there is a case of several vertices bunched close together, an algorithm called deterministic annealing [3] is used. The resulting vertices are required to have at least two tracks associated with them in order to be accepted.

4.1.3. Calorimeter Clustering

The clustering for the calorimeters is done separately in each subdetector to account for the differences in the setup of these subdetectors. The general principle for clustering remains the same. First, cluster seeds are identified that have energy depositions in a single cell beyond a given threshold. The energy deposition in the four (for HCAL) or eight (for ECAL) neighbour cells are then evaluated, and if any neighbouring cells have an energy deposition above a threshold value, their neighbour cells are also included in the cluster.

4.2. Generator Objects

For the Monte Carlo simulations of signal events and background, the generator information is also stored in addition to the reconstructed objects from the detector responses. Generator information concerns the properties of generated particles and their subsequent decay products without considering interactions with any detector material. In MiniAOD format [4], a reduced amount of generator information is available for the purpose of testing reconstruction efficiency and accuracy of analyses. This includes the full set of information on particles originating from the hard process, called prompt particles, at the final vertex. The prompt particles include the initially generated tops or W/Z bosons along with a subsequent

chain of decays of short lived particles prior to hadronization. The decay products of τ leptons are for example not considered prompt particles. In addition to the prompt particles, the four momenta and particle types for final state particles are stored separately from the prompt particles. This enables running any type of jet reconstruction, but the necessary information on the generated jets with the default reconstruction algorithm and radius parameter is also available.

4.3. Muons

Muons have a relatively high mass and an electric charge, but are unaffected by the strong force. This combined with a relatively long lifetime enables the muons to travel further through the detector than hadrons or any other known detectable particles. Their electric charge makes them activate the tracker and deposit energy in both the ECAL and the HCAL, and if their momentum is high enough, they will even pass through the return yoke of the magnet, activating the muon spectrometer.

Muon candidates are divided into three categories depending on the extent to which the muon reconstruction is achieved. The category with reconstructed tracks from the muon spectrometer is called standalone muon. The tracks are formed from clustered hits from ether a DT or a CSC called segments that serve as seeds for a full pattern recognition from the DT, CSC and RPC hits along the estimated trajectory of a muon.

The tracks from the inner tracker with $p_{\mathrm{T}} > 0.5$ GeV with a total momentum of at least 2.5 GeV that can be matched with at least one segment from the muon spectrometer are called tracker muons.

Standalone muons are each matched with a track from the inner tracker, and sets that are compatible in their parameters are then passed as global muons.

The muon reconstruction rate within the acceptance of the muon spectrometer is as high as 99% with some contamination from backgrounds such as high energy charged hadrons punching through the HCAL and cosmic muons. To effectively reduce these backgrounds, and to accommodate varying needs in different analyses for specific types of muons, the selection is split into two parts; identification (ID) and isolation (ISO). The muon identification is categorised into working points, which evaluate key parameters such as the closest distance from the muon track

to the primary vertex (impact parameter), the displacement of the track from the muon chambers from the prediction based on the track from the tracker and the number of hits in the tracker and/or muon chambers associated with the muon candidate. The tracks are also checked for splitting with a kink finding algorithm, and the PF algorithm evaluates the deposits in the calorimeters for compatibility with the muon hypothesis in addition to the parameters derived from the trackers and muon stations [5].

The tight muon ID working point is designed to select prompt muons while the selection of muons from decays of hadrons and tauons is strongly suppressed along with that of the backgrounds. The efficiency of selecting prompt muons that have passed reconstruction with $p_{\mathrm{T}} > 20$ GeV is about 96%. The tight selection makes use of the impact parameter to ensure that the closest distance from the track to the primary vertex lies within a few millimetres. In addition the muon must be considered both a tracker muon and a global muon, with the tracker muon being matched to at least two muon stations and the global muon fit needs to have a χ^2/N_{dof} score below 10. It is further required that the muon leaves hits in at least six layers of the inner tracker and at least one hit in the pixel detector.
The medium muon ID working point is designed to have a high selection efficiency for prompt muons and muons from heavy flavour decays, while suppressing selection of backgrounds and muons originating from light flavoured decays. The selection requirement includes leaving hits in at least 80% of the layers of the inner tracker that the track traverses, and a satisfactory score from a boosted decision tree (BDT) based on the number of hits in both the tracker and the muon stations and the spatial consistency between these. The selected muons can either be tracker muons only or both tracker muons and global muons. The target efficiency for prompt muons for the medium working point is 99.5%, while the average efficiency is slightly above 98% for prompt muons with $p_{\mathrm{T}} > 20$ GeV.
The loose muon ID is designed to select muons that are either prompt or from heavy or light flavoured decays while still suppressing selection of background particles. It selects tracker muons or global muons that pass the muon identification in the PF algorithm, which has a variety of paths to selecting muon candidates, each with specific criteria such as isolated muons, tight muon ID and calorimeter evaluation if the tracker muon only activates one muon station. The average selection efficiency for prompt muons with $p_{\mathrm{T}} > 20$ GeV with the loose ID is 99.7%.
For the muon isolation, two working points for selection have been established. The muon isolation checks for tracks and neutral particles within a $\Delta R < 0.4$ cone ($\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$), adding up the p_{T} of the tracks and the energy of the neutral particles. This sum of p_{T} and energy is then compared to the muon p_{T}, where the fraction has to be below a certain threshold for the muon to pass either

the loose ISO working point (0.25) or the tight ISO (0.15). The efficiencies for the ISO working points are evaluated on muons passing the tight ID to be around 98% for the loose ISO and 95% for the tight ISO[1][5].

To facilitate analysis on muon events, a muon trigger has been set up with a set of selection requirements in the L1 trigger and the HLT. Due to limitations in terms of computation time and the absence of tracker input, the L1 trigger limits the possible efficiency of the muon trigger to be about 90% [5].

4.4. Electrons

Electrons have several properties that give them a unique signature in the detector. They have a very low mass and are electrically charged, making them very susceptible to bending in magnetic fields. Their charge makes them activate the trackers they pass through, and in combination with their low mass, they are expected to deposit most of their energy in the ECAL with the possibility of some energy leaking through to the HCAL. The high energy electrons emit significant amounts of radiation in the form of bremsstrahlung, due to electron interactions with the tracker material, and synchrotron radiation from the bending of the electron trajectory. In the ECAL, the electron will form a shower consisting of photons and electrons originating from bremsstrahlung and pair production respectively.

Electrons are either identified as clusters in the ECAL with an energy greater than 4 GeV or reconstructed from the track and matched with an ECAL cluster. To account for bremsstrahlung, track bending and showering before the calorimeter, additional ECAL clusters are included within a narrow window in η and a wide window in ϕ which is the direction the electron trajectory is bent due to the magnetic field. This collection of clusters is called a supercluster. From an ECAL supercluster, a track seed consisting of two hits in different layers of the pixel detector is matched to the cluster under the assumption of a positron or an electron [6].

To reconstruct tracks for electron identification, selected tracks are refitted with a Gaussian-sum filter (GSF) to the extent of 5 hits in the tracker. The track is then extrapolated to the ECAL surface and associated with the nearest cluster in the ECAL. A BDT analysis on the track is applied and the score is evaluated as a final quality cut. The tracker based electron reconstruction extends the electron reconstruction range down to $p_{\mathrm{T}} = 2$ GeV and increases the electron reconstruction

efficiency compared to only using ECAL based reconstruction.

From the track seeds, from either the ECAL based approach or the track based, the full tracks are fitted using the GSF approach across the full range of the tracker or until no further hits can be found therein in two consecutive layers. The reconstruction rate for electrons with $p_T > 10$ GeV in the ECAL supercluster is greater than 85% with increasing efficiency for electrons at higher transverse momentum up to 93% as shown in fig. 4.1. The electron reconstruction rate varies

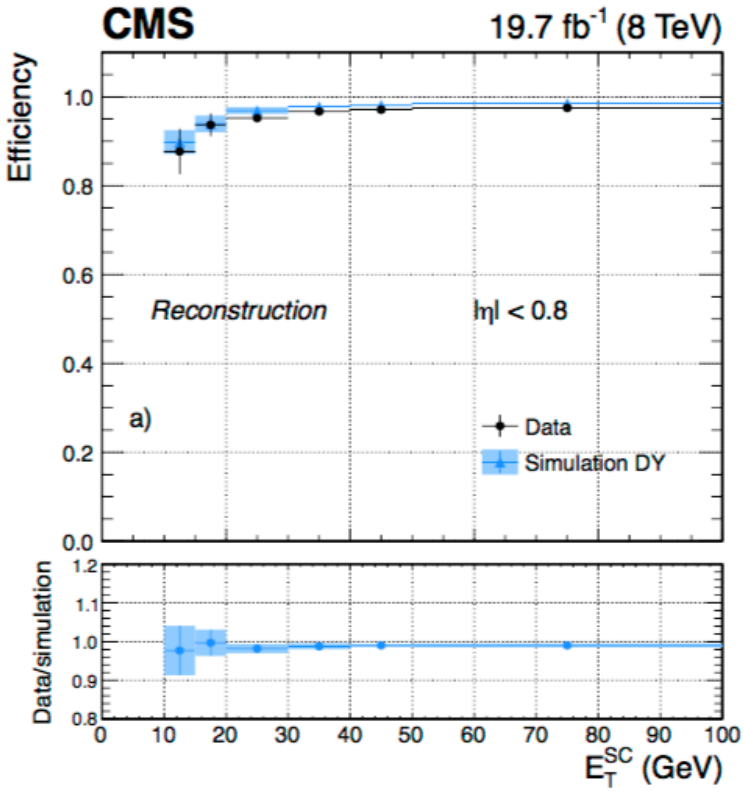


Figure 4.1.: Performance of the electron reconstruction in the barrel region ($|\eta| < 0.8$) for dielectron events in data (dots) and MC simulation of Drell-Yan events (triangles). The bottom part of the plot shows the ratio between data and MC simulation, which is used to determine the reconstruction scale factor. Figure taken from source [7].

according to the year of data taking due to detector upgrades, radiation damage and improved reconstruction methods, so the numbers presented here need to be taken with a grain of salt. One important feature pointed out in fig. 4.1 is the slight difference in the reconstruction rate in data and MC simulation, which is

accounted for with a scale factor.

The electron reconstruction is optimised for efficiency rather than purity and many of the reconstructed electron candidates do not originate from actual electrons. They originate mainly from some of the charged hadrons from jets that deposit most of their energy in the ECAL rather than the HCAL. To reduce the rate of misidentification, a variety of parameters are evaluated that have some discriminating power between actual electrons and misidentified electrons. Electrons tend to have a greater portion of their energy deposited in the ECAL than the misidentified electrons, which makes the ratio of energy deposition in the HCAL cells close to the ECAL supercluster to the energy deposited in the ECAL supercluster one of the parameters used to obtain higher electron purity.
Another set of variables used have to do with the distance from the ECAL supercluster to the point at the ECAL interface, extrapolated from the innermost hit from the track seed under the assumption of an ideal helix trajectory, based on the parameters at the interaction vertex for the track. This distance is given both in terms of the azimuthal angle ($\Delta\phi$) and pseudorapidity ($\Delta\eta$) as separate variables with the real electrons being more likely to have distance parameters close to zero.
The showers in the ECAL tend to be narrower for electrons than for charged hadrons, which gives rise to another set of discriminating variables ($\sigma_{\eta\eta}$ and $\sigma_{\phi\phi}$), that add up the distance from the seed crystal to each additional crystal in the cluster, weighted by the logarithm of the energy deposited in the given crystal, normalised to the sum of weights in the cluster.
The ratio between the ECAL energy and the track momentum tends to be closer to unity for actual electrons than for misidentified electrons, which leads to a set of parameters that can be used to further enhance the purity of the electron identification. One parameter is the ratio between the energy deposited in the ECAL seed cluster and the track momentum at its latest stage. Similarly the track momentum at the initial stage is compared to the energy of the ECAL supercluster, which has been corrected for energy depositions in non-ECAL material (HCAL, tracker, gaps), energy deposits outside the supercluster and contamination from pileup [7].
A final discriminating feature between electrons and charged hadrons is that electrons tend to be isolated with only minor activity around them, whereas charged hadrons are typically components of a larger ensemble of particles forming a jet. There are two variables of interest for establishing the isolation requirement; one is the radius of the isolation cone in terms of (η, ϕ) and the other is the energy within this cone radius which is not assigned to the electron relative to the transverse momentum of the electron.
Apart from the charged hadrons, there are also photons that can be classified as

electrons in the PF algorithm if the photon has formed an electron-positron pair in the tracker. To reduce this occurrence, the pixel detector is checked for missing hits and the tracks are checked for vertex displacement.

The above mentioned parameters are used to obtain different levels of purity depending on the needs of the given analysis called working points. Two methods are used to establish these working points; the sequential method and an approach based on multivariate analysis (MVA) where the variables have been combined using BDT. The sequential method applies cuts on the discriminating variables to produce a set of four working points (veto, loose, medium and tight) with the veto being the loosest one with the highest selection efficiency for the reconstructed electrons. The purpose of the veto working point is to reject events where electrons appear, making efficiency a much higher priority than purity. When one on the other hand needs a given number of electrons present, one usually requires a certain level of confidence that the vast majority of the selected electrons are actual electrons. The target electron selection efficiencies are 90% for the loose working point, 80% for medium and 70% for the tight working point.
Wth the MVA based approach, two working points are selected, one with a target efficiency of 90% and one with 84%. The MVA based approach outperforms the sequential method significantly as shown in fig. 4.2, where the selection efficiency on electrons with $p_{\mathrm{T}} > 20$ GeV is demonstrated compared to the efficiency of misidentified electrons.

4.5. Jets

When partons are produced, the QCD processes of hadronization, decay and gluon radiation produce a shower of energetic particles. This shower is called a jet, which is an essential object for the analyses in this work. The jets are reconstructed with an iterative anti-k_{t} algorithm with a jet cone radius, R of 0.4 in (y, ϕ) space where y is the rapidity and ϕ is the azimuthal angle. The rapidity is defined as follows:

$$y = \frac{1}{2}\log\left(\frac{E + p_{\mathrm{z}}}{E - p_{\mathrm{z}}}\right), \tag{4.1}$$

where p_z is the projection of the particle momentum on the z-axis and E is the particle energy. The radius parameter is thus given as $R = \sqrt{y^2 + \phi^2}$.

The anti-k_{t} algorithm takes the hardest hits in the detector and clusters the hits within the radius parameter to form a jet. The anti-k_{t} algorithm keeps the area

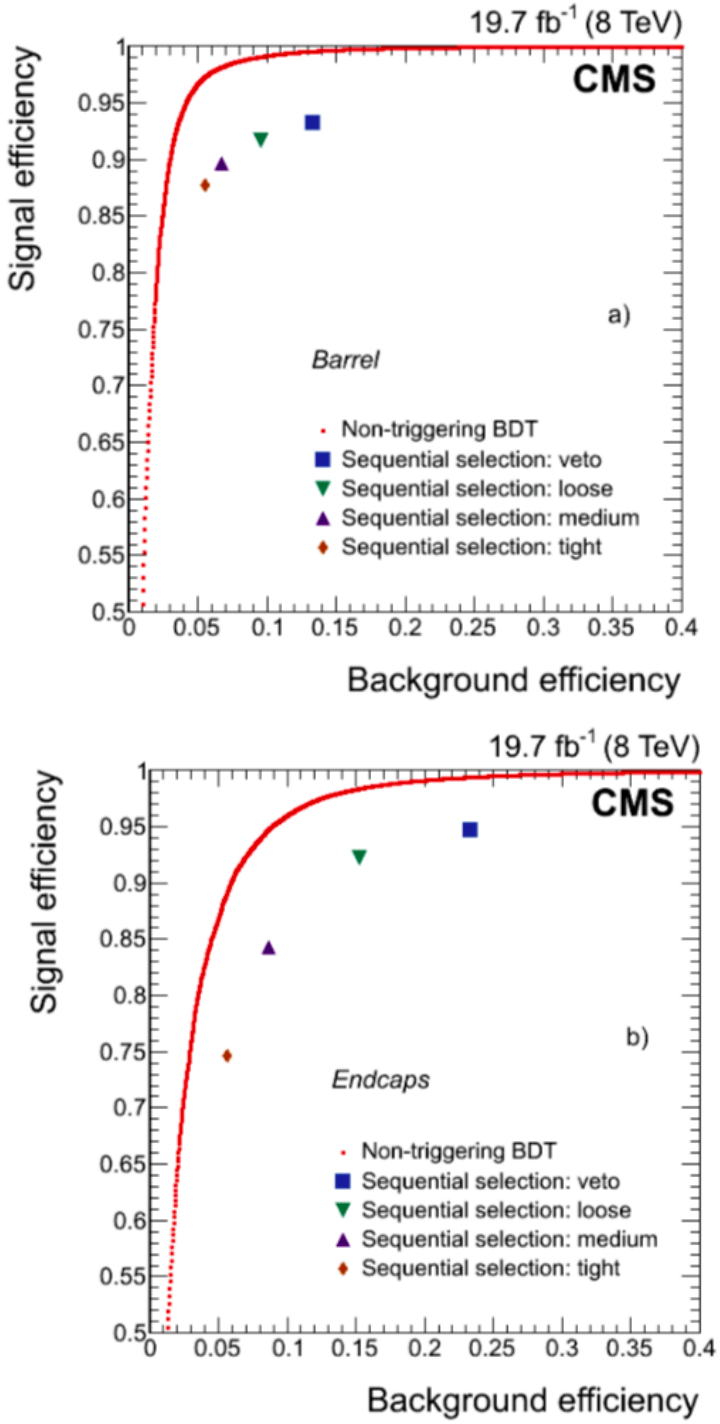


Figure 4.2.: Performance of the MVA based method (red dots) compared to the working points from the sequential method. Figure taken from source [7].

of the jet cone constant, but the shape can vary when two jets overlap. Further details on the anti-k_t algorithm are found in [8]. Jets reconstructed with the anti-k_t algorithm and the radius parameter of $R = 0.4$ are referred to as AK4 jets.

Another type of jets called fat jets are also reconstructed with the anti-k_t algorithm but with a radius parameter $R = 1.0$. Within the cone of the fat jets, sub-jets with $R = 0.2$ are also reconstructed with the anti-k_t algorithm. Fat jets with $p_T > 100$ GeV are stored.

Three types of jets are used for different purposes in analysis: Calo-jets which are established only from the calorimeter clusters within the jet cone, PF-jets that use all the reconstructed particles from the PF algorithm within the jet cone. REF-jets are made from the stable particles from the event generator, within the jet cone, except neutrinos. PF-jets are available down to $p_T = 10$ GeV, whereas calo-jets only are reliably reconstructed down to $p_T = 20$ GeV.

4.5.1. Jet Energy Corrections

The energy of jets is attributed mainly to charged hadrons which on average have 65% of the total jet energy, whereas 25% of the jet energy is associated with photons and 10% with neutral hadrons. Due to the high rate of pileup, it is necessary to subtract the estimated energy depositions from charged hadrons originating from a primary vertex other than that of the jet. This procedure is called charged hadron subtraction (CHS). To correct for energy lost in the tracker and for remaining energy added to the jet from pileup events and for other effects, jet energy corrections are applied. These corrections are applied in three stages, each stage requiring the application of the preceding stage. The full procedures for jet energy corrections at CMS are described in [9] and outlined below:

The first stage is called pileup offset removal, which corrects jet energies in a pileup rich environment to match jet energies for jets in a pileup free environment. This is done using jets from MC simulations of multijet QCD events where the simulated sample of jets without pileup are matched with jets from the same simulation with a distribution of pileup events added to match the pileup environment of the data. The response is then taken between the energies of the matched jet pair to derive the correction. A scale factor for this correction is derived by using zero bias data and simulation together with the random cone method (RC) [10].

The second stage corrects the discrepancy between the particle energy and the

detector response in MC simulations of QCD events.

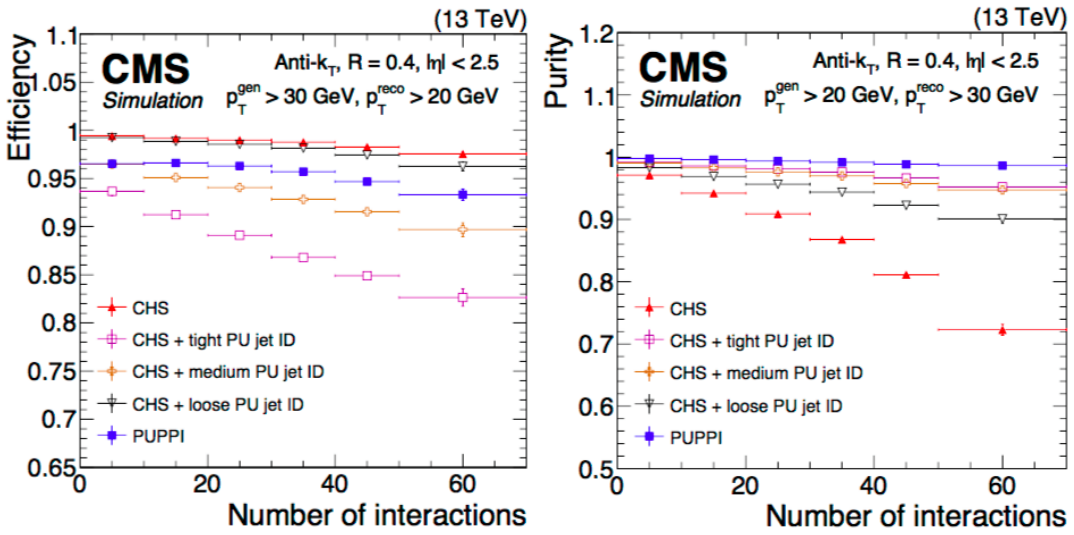
The third stage corrects the discrepancy between the simulated detector response and the actual detector response. Two main categories of corrections are applied at this stage called relative and absolute. In both cases, the p_{T} balance and the missing transverse momentum projection fraction (MPF) balance are used to establish the remaining discrepancies in jet energy to be corrected. Corrections in the relative category uses samples of dijet events where the response of the jet within $|\eta| < 1.3$ is compared to that of the remaining jet which is unconstrained in pseudorapidity. Among the relative jet energy corrections is the jet energy resolution (JER) with scale factors that evolve over $|\eta|$ which are applied to the JER in MC simulations. To scale the width of the jet energy distribution, the energy of each jet is smeared with a Gaussian, with a mean of one and a standard deviation of $\sqrt{k^2 - 1}\sigma_{MC}$, where k is the scale factor and σ_{MC} is the JER in MC simulations.
The basis for the absolute corrections are $Z \to ee + \text{jets}$, $Z \to \mu\mu + \text{jets}$, $\gamma + \text{jets}$ and multijet events in data and MC. For these corrections, the p_{T} of jets with 30 GeV $< p_{\mathrm{T}} <$ 800 GeV is compared to that of the vector boson Z or γ, the given jet recoils against. Jets with $p_{\mathrm{T}} > 800$ GeV are evaluated in multijet events where the leading jet recoils against a system of two or more jets.

4.5.2. Jet Selection and Pileup Jet Rejection

Since jets are not just any bunch of particles one can fit in a cone, it is necessary to impose selection criteria to remove jets originating from either detector noise, non-QCD processes or pileup. To effectively suppress jets reconstructed from detector noise and non-QCD processes, a set of jet identification (jet ID) working points have been established. In both analyses in this work, only jets with $|\eta| < 2.4$ have been selected for study, so the jet ID working points of interest are only described within this range. The charged hadron fraction is required to be above 0.0 and the charged particle multiplicity is required to be above 0 for all working points. The loose and tight jet ID working points require a charged electromagnetic energy fraction of less than 0.99, whereas the tight lepton veto jet ID requires a charged electromagnetic energy fraction of less than 0.90 along with a muon energy fraction of less than 0.80 to prevent misidentification of isolated leptons as jets. The neutral hadron and electromagnetic energy fractions are each required to be less than 0.99 for the loose jet ID and less than 0.90 for the tight and tight lepton veto IDs.
The selection efficiency for the tight jet ID is greater than 98% for jets with

30; GeV $< p_{\mathrm{T}} <$ 100 GeV and $|\eta| < 2.4$, whereas it is reported to be 99.8% for $|\eta| < 0.5$ and $p_{\mathrm{T}} > 30$ GeV. The tight jet ID working point achieves a background rejection of 99.999% [11].

With an increasing rate of pileup interactions per bunch crossing resulting from higher instantaneous luminosities, the CHS method alone proves insufficient to prevent selection of jets originating from pileup events. This is especially the case for jets with $p_{\mathrm{T}} < 50$ GeV, on which a set of three PU jet ID working points have been established based on 15 variables in a BDT trained on $Z +$ jets events. The PU jet ID working points are established based on the selection efficiency of quark based jets with the tight working point having an 80% efficiency on such, the medium working point has a 90% efficiency and the loose working point has a 99% efficiency within $|\eta| < 2.5$. An alternative to the PU jet ID is to make use of jets that have undergone an extended version of CHS called PUPPI designed for a high pileup environment [12].



(a) (b)

Figure 4.3.: Efficiency for jets originating from the hard collision at the different PU jet ID working points (a) and the purity of the selection (b) as a function of the number of pileup interactions. Figure taken from source [12].

As seen in fig. 4.3, the purity of jets only with CHS decreases rapidly as the number of pileup interactions per event increases. The efficiency also decreases with an increasing number of pileup interactions, in particular for the medium and tight PU jet ID working points. Since the average number of pileup interactions have

increased from 23 in 2016 to 32 in 2017 as stated in section 3.2, it has become necessary to use the PU jet ID in analysis. At even higher pileup rates, the advantages from switching to PUPPI jets are impossible to ignore, since they actually outperform even the tight PU jet ID working point in purity, while maintaining an efficiency above 90% at the highest pileup rate presented. The efficiency and purity is here presented for jets within $|\eta| < 2.5$, since both analyses in this work only select jets within this range.

4.5.3. Heavy Flavour Jet Tagging

Jets that originate from a heavy flavour quark are tagged when possible using machine learning algorithms. Especially the tagging of jets originating from b-quarks, called b-tagging is of interest in the analyses in this work. The b-tagging algorithm used on 2016 data is called CSVv2 [13], whereas an improved algorithm based on deep neural networks (DeepCSV) is recommended for analysis on 2017 data [14]. For analysis, there are three working points for the b-tagging algorithms based on the false positive rate for light flavour jets ($udsg$) that are set at 0.1% for the tight working point, 1% for the medium working point and 10% for the loose working point. The false positive rate for c flavoured jets is typically around 12% for the medium working point.

4.6. Missing Transverse Energy

One key parameter of interest in searches for new physics is missing transverse energy. Missing energy is expected from neutrinos and perhaps other particles that do not interact with any part of the detector. In reconstructing the missing transverse energy, one takes the negative vector sum of the transverse momentum of all particles in the event as follows:

$$E_{\mathrm{T}}^{miss} = -\sum_{i=1}^{N_{particles}} \vec{p}_{\mathrm{T},i}, \tag{4.2}$$

where E_{T}^{miss} is the missing transverse energy and $\vec{p}_{\mathrm{T},i}$ is the transverse momentum vector for a given particle indexed with i.

Bibliography

[1] **CMS Collaboration**, *'Particle-flow reconstruction and global event description with the CMS detector'*,
JINST 12 (2017) no. 10, P10003, doi:10.1088/1748-0221/12/10/p10003, arXiv:1706.04965v2.

[2] **CMS Collaboration**, *'Description and performance of track and primary-vertex reconstruction with the CMS tracker'*,
JINST 9 (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569v2.

[3] K.Rose, *'Deterministic Annealing for Clustering, Compression, Classification, Regression and related Optimisation Problems'*,
Proceedings of the IEEE 86 (1998), doi:10.1109/5.726788.

[4] G. Petrucciani et al. (for the CMS Collaboration), *'Mini-AOD: A New Analysis Data Format for CMS'*,
J.Phys.:Conf. Ser. 664 (2015) 7 072052, doi:10.1088/1742-6596/664/7/072052, arXiv:1702.04685v2.

[5] A. M. Sirunyan et al., *'Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at* $\sqrt{s} = 13\ TeV$*'*,
2018 JINST 13 P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528 [physics.ins-det].

[6] S. Baffioni et al., *'Electron reconstruction in CMS'*,
Eur. Phys. J. C 49, 1099?1116 (2007), doi:10.1140/epjc/s10052-006-0175-5.

[7] **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, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[8] M. Cacciari et al., *'The anti-k_t jet clustering algorithm'*, JHEP 0804:063,2008, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189v2.

[9] **CMS Collaboration**, *'Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV'*, JINST 12 (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[10] **CDF Collaboration**, *'Determination of the jet energy scale at the Collider Detector at Fermilab'*, Nucl. Instrum. Meth. A 566 (2006) 375, doi:10.1016/j.nima.2006.05.269, arXiv:hep-ex/0510047.

[11] **CMS Collaboration**, *'Jet algorithms performance in 13 TeV data'*, CMS PAS JME-16-003 (24/3/2017).

[12] **CMS Collaboration**, *'Pileup mitigation at CMS in 13 TeV data'*, 2020 JINST 15 P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503 [hep-ex].

[13] **CMS Collaboration**, *'Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV'*, JINST 12 (2017) no. 10, P10003, doi:10.1088/1748-0221/13/05/p05011, arXiv:1712.07158v2.

[14] M. Stoye, **CMS Collaboration**, *'Deep Learning in jet reconstruction at CMS'*, J. Phys.:Conf. Ser. 1085 042029, doi:10.1088/1742-6596/1085/4/042029.

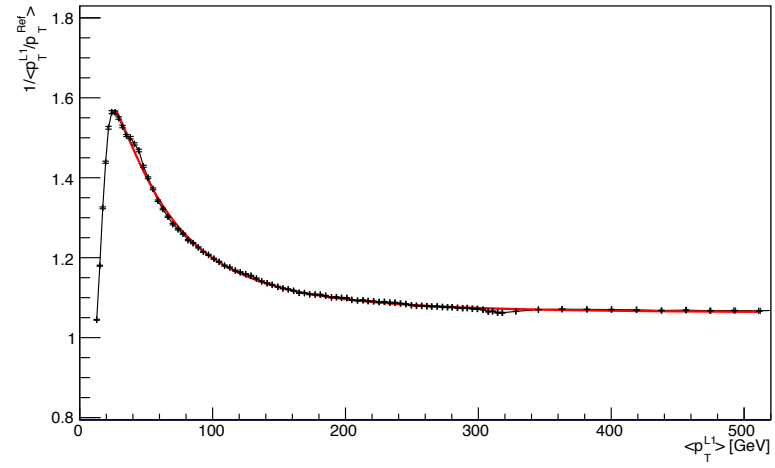
5. Level 1 Trigger Jet Energy Calibration

As a service task, I have investigated the possibility of using PF-jets instead of the MC generated REF-jets normally used for calibration of jet energies in the L1 trigger. The REF-jets, PF-jets and calo-jets are explained in chapter 4. The motivation for this investigation is the underperformance of the calibrations based on REF-jets in the HF region of the detector compared with using the uncorrected jet energies from the calorimeter fed directly to the L1 trigger. This issue has occurred for the first time in 2017 data, and in my service work I have used a subset of the data from 2018 for calibrations.

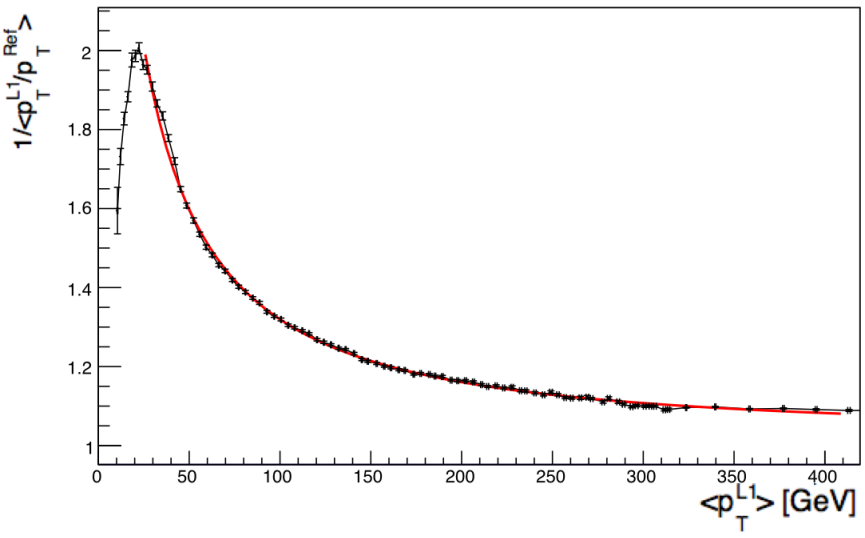
Events for calibration are selected which have pileup rates in the high end of the peak value ranging from 40 to 50 primary vertices. The correspondence between cluster size in the calorimeters and jet energy varies with pseudorapidity. This is due to differences in calorimeter composition and in energy loss when particles pass through detector material from the interaction point to the given segment of the calorimeters. Since the detector layout is the same in the positive and negative direction along the z-axis, the differences between positive and negative pseudorapidities are expected to be small enough to work with segments in absolute pseudorapidity. The selected bins for the corrections are given with the bin edges for $|\eta|$: 0, 0.435, 0.783, 1.131, 1.305, 1.497, 1.653, 1.83, 1.93, 2.043, 2.172, 2.322, 2.5, 2.964, 3.489, 4.191, 5.191.

For the selected events in a given bin of $|\eta|$, the momentum response in a given p_{T} bin of the L1 jet is evaluated with a Gaussian fit of the p_{T} distribution of the reference jets. The mean of the Gaussian fit then provides a data point for one of the plots of response vs p_{T}^{L1} seen in figg. 5.1, 5.2 and 5.3.

These data points are then fitted to cover as wide a range as possible in p_{T}^{L1}. The fit is extended with horizontal lines from the end points of the fit to cover the full p_{T}^{L1} range in the lookup table. The process is repeated for all bins in $|\eta|$, wherefrom



(a)



(b)

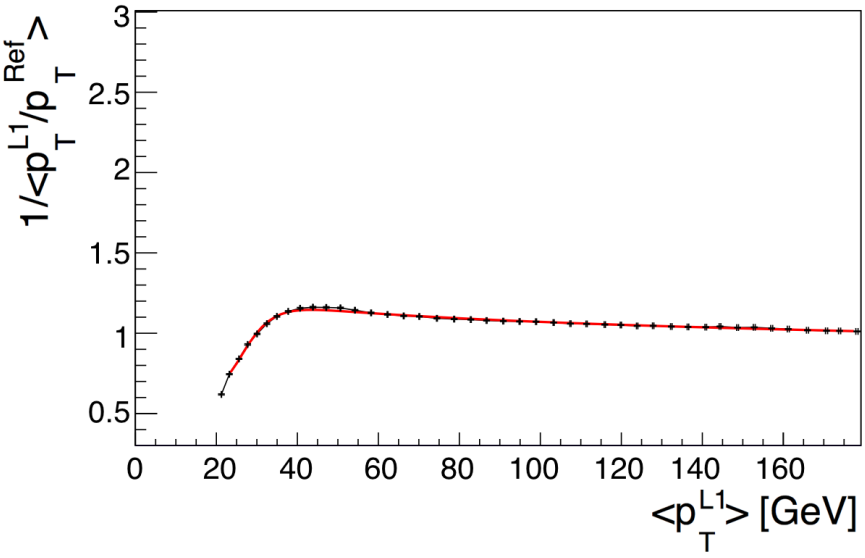
Figure 5.1.: Fitted response of the L1 jet energy (the mean ratio of PF-jet p_{T} over L1 jet p_{T}) as a function of the L1 jet p_{T} (black dots), where vertical error bars refer to the width of the Gaussian fit of the response within a given bin of L1 jet p_{T}, indicated as horizontal error bars. A fitted function to the data (red line) is then used at the later stages of the L1 jet energy corrections. These plots cover the regions of (a) $0 \leq |\eta| < 0.435$ and (b) $1.83 \leq |\eta| < 1.93$ with the number of primary vertices set to be between 40 and 50 on a subset of 2018 data.

the fitted functions form the basis for the lookup table.

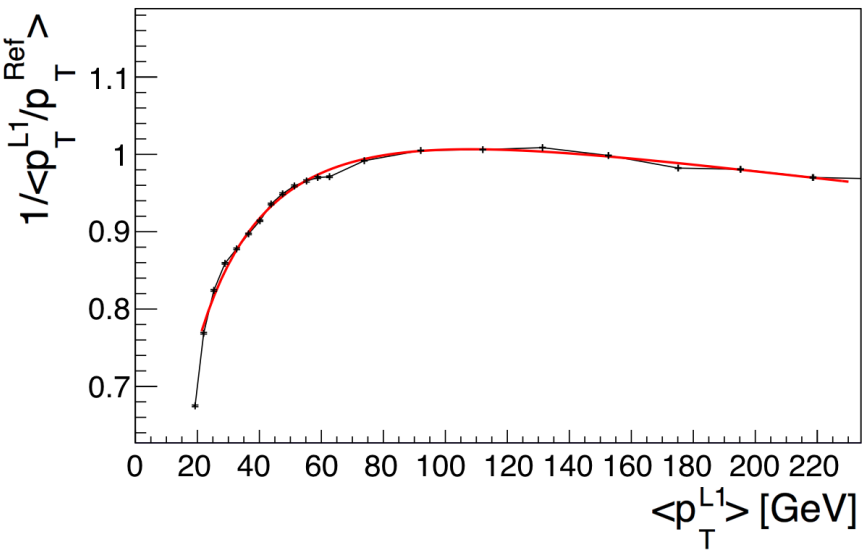
In the barrel-endcap regions of the detector up to $|\eta| = 2.5$, the fits have similar shapes with a sharp peak at low L1 jet p_T and converging towards a plateau close to unity as the L1 jet p_T increases. These features are shown in fig. 5.1 for two η regions, one in the barrel region and one in the endcap region. In the η region transitioning between the endcap and the HF calorimeter, the low p_T peak almost vanishes due to the lack of tracking in the endcap at $|\eta| \geq 2.5$ as shown in fig. 5.2a. The fits for the HF regions have broad peaks at different values of L1 jet p_T. These values are: $p_T \approx 100$ GeV for $2.964 \leq |\eta| < 3.489$ as seen in fig. 5.2b, $p_T \approx 80$ GeV for $3.489 \leq |\eta| < 4.191$ (see fig. 5.3a) and $p_T \approx 60$ GeV for $4.191 \leq |\eta| < 5.191$ as seen in fig. 5.3b.

The information from the fits has been transferred to lookup tables so that they can be installed in the L1 trigger. At first, a test is performed on the selection efficiency of the L1 trigger with the new PF-jet based corrections (green, orange and red triangles in fig. 5.4) at various cuts on jet p_T. For comparison, the same test is performed on the existing corrections (variety of blue dots in fig. 5.4) of L1 jet energy from MC generated REF-jets in the barrel-endcap region. In the HF region, the existing correction is the raw measured calo-jet energy, due to unsuccessful correction attempts. A more accurate correction on jet p_T leads to a sharper transition from zero efficiency to full efficiency.

As seen in fig. 5.4, the performance of the PF-jet based corrections is comparable with the performance of corrections based on MC generated REF-jets in the barrel-endcap region. There are even hints of a slight improvement in accuracy at low jet p_T values when using the new PF-jet based corrections. In the HF region, the PF-jet based corrections perform as well as uncorrected calo-jets. Since these corrections deviate significantly from unity without leading to a significant improvement in accuracy, there must be something wrong with estimating the p_T of the PF-jets in the HF region. The same issue has been observed for MC generated REF-jets, which has motivated this study. The results from this study have been reported to the CMS trigger group for further study into the issues in the HF region, so that the CMS trigger can run optimally.

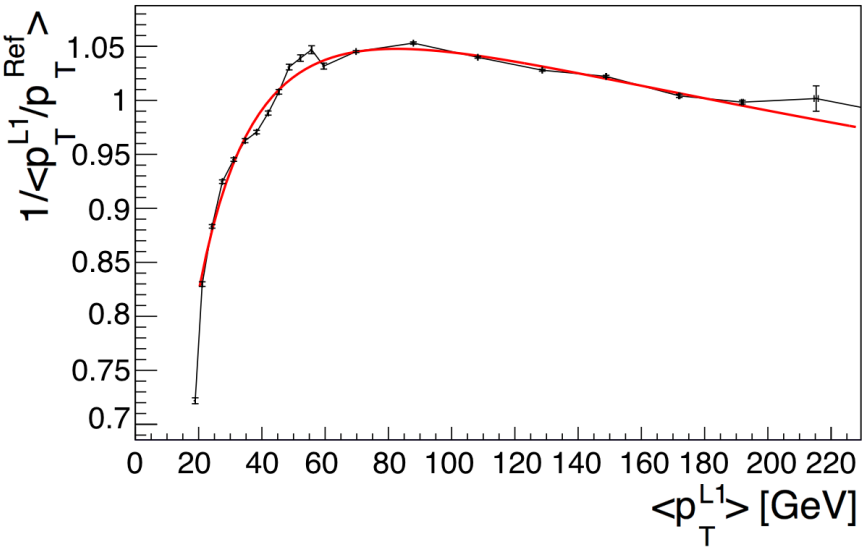


(a)

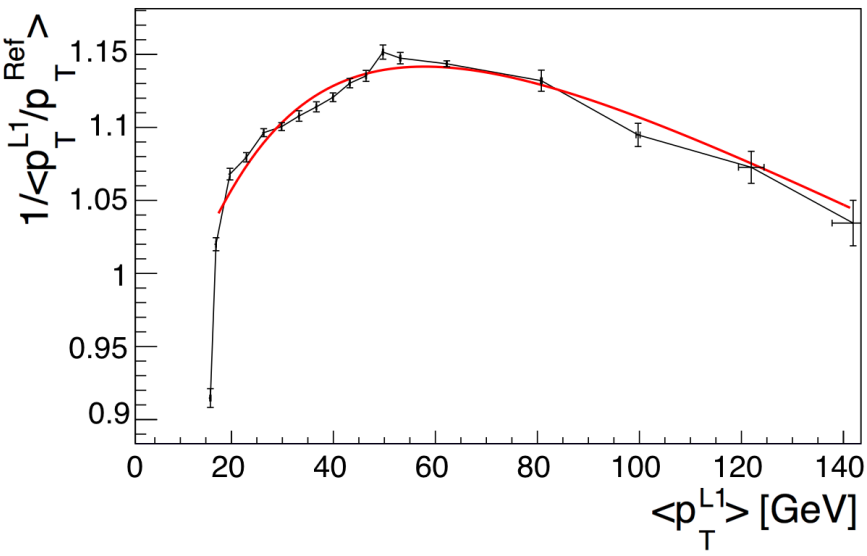


(b)

Figure 5.2.: Fitted response of the L1 jet energy (the mean ratio of PF-jet p_{T} over L1 jet p_{T}) as a function of the L1 jet p_{T} (black dots), where vertical error bars refer to the width of the Gaussian fit of the response within a given bin of L1 jet p_{T}, indicated as horizontal error bars. A fitted function to the data (red line) is then used at the later stages of the L1 jet energy corrections. These plots cover the regions of (a) $2.5 \leq |\eta| < 2.964$ and (b) $2.964 \leq |\eta| < 3.489$ with the number of primary vertices set to be between 40 and 50 on a subset of 2018 data.

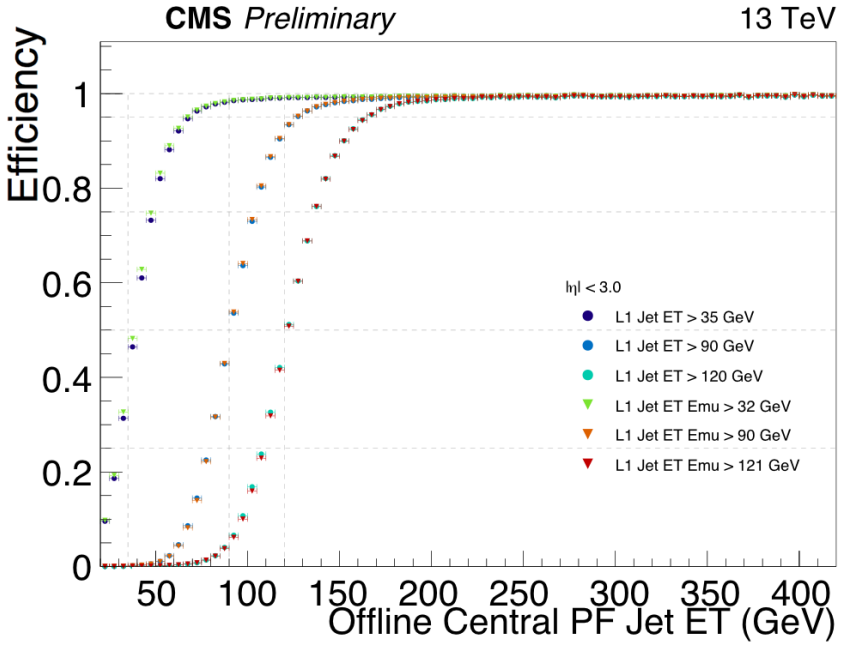


(a)

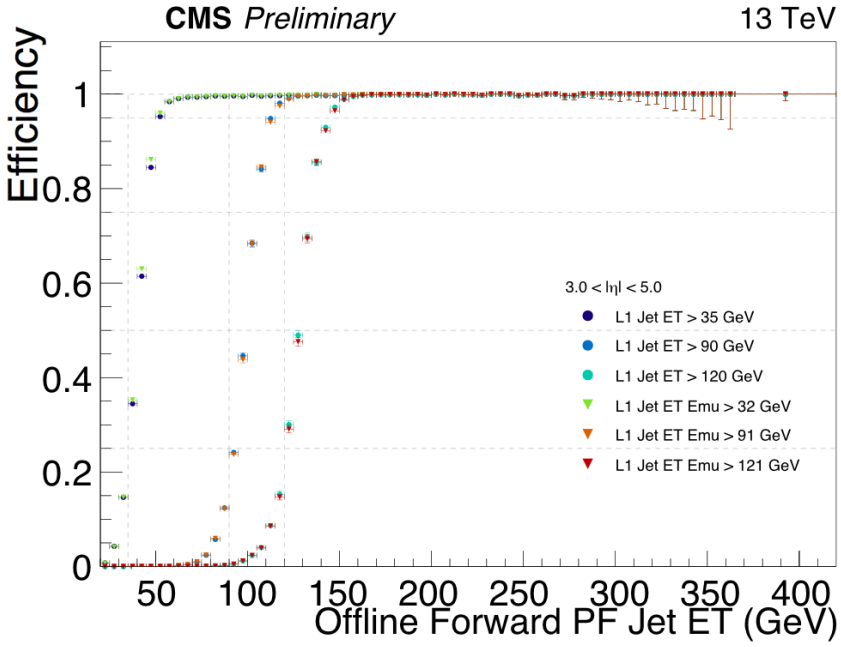


(b)

Figure 5.3.: Fitted response of the L1 jet energy (the mean ratio of PF-jet p_{T} over L1 jet p_{T}) as a function of the L1 jet p_{T} (black dots), where vertical error bars refer to the width of the Gaussian fit of the response within a given bin of L1 jet p_{T}, indicated as horizontal error bars. A fitted function to the data (red line) is then used at the later stages of the L1 jet energy corrections. These plots cover the regions of (a) $3.489 \leq |\eta| < 4.191$ and (b) $4.191 \leq |\eta| < 5.191$ with the number of primary vertices set to be between 40 and 50 on a subset of 2018 data.



(a)



(b)

Figure 5.4.: Jet selection efficiency of the L1 trigger as a function of PF-jet transverse energy in (a) the barrel-endcap region and in (b) the HF region. The dots (dark blue, azure, teal) indicate the existing jet energy corrections based on MC generated REF-jets in the barrel-endcap region and jets without corrections in the HF region, the triangles (green, orange, red) indicate the new jet energy corrections using PF-jets instead.

6. Validation and Production of Signal Samples

6.1. Monte Carlo Simulation

Monte Carlo simulations (MC) are performed to generate a large number of simulated collision events, such that the probability to produce an event is proportional to the probability that the corresponding actual event is produced in the real world [1]. The hard subprocess is simulated where constituents of the colliding protons, the partons, whose momentum fractions are given according to parton distribution functions (PDFs), interact at a high momentum scale to produce outgoing fundamental partons. All initial partons involved, as well as any new particles with colour, radiate virtual gluons, which can themselves emit further gluons or produce quark-antiquark pairs, leading to the formation of parton showers. During parton showering the interaction scale falls until a cut-off scale. Below that scale, the process of hadronization transforms the partons into colourless hadrons. The other remnant partons of the incoming protons undergo multiple interactions which produce the underlying event. Since many of the produced hadrons are unstable, a simulation of hadron decays follows. In a final step, to be able to compare the predictions directly to experimental data, the interactions of the final state particles with the various detector components are simulated.

6.1.1. Generating the Hard Subprocess

The computation of the hard subprocess of an event is performed by calulating the quantum field theory matrix element (ME) utilizing Feynman diagrams. It describes the proton–proton collision process from the initial interaction between the two partons through to final state particles.

As input to the MEs, partons are taken from the protons according to their PDF. The PDF is reweighted on an event-by-event basis for the variations representing the uncertainty of the choice of function for fitting the data for the PDF. The same reweighting is generally performed on alternative PDF sets in case one wants to study the process based on a different PDF set than the one chosen for the event generation.

Modern MC event generators usually contain both MEs at leading order QCD (LO QCD), and MEs at Next-to-leading order QCD (NLO QCD) including virtual and real corrections. Adding such MEs in higher-order α_s improves the accuracy of the cross-sections and the kinematic distributions of the final state particles. In addition, tree-level MEs at higher order in α_s producting larger numbers of final state partons are typically included to improve the accuracy of the predictions further, in particular for events with high parton multiplicities.

Beyond that, additional contributions due to higher orders in QCD involving leading logarithms (LL) are generated through the so-called parton shower simulation. Here partons at the initial state at high energy radiate off some of their energy in the form of quarks or gluons, which is called initial state radiation (ISR). Similarly, final state partons also radiate quarks and gluons, called final state radiation (FSR) [2]. During parton showering the interaction scale falls until it reaches a cut-off scale, where the process of parton radiation stops.

Since additional partons can therefore be produced both by MEs and by partons showers, it is an important issue to avoid double counting between them. This is what is usually referred to as matching. There are different approaches to matching. One can either apply weights to suppress events that occur from both radiative corrections at ME level and from the parton shower, or simply reject parton shower events that have an equivalent at ME level [3]. The latter approach is taken in MLM matching, which has the advantage of being independent of jet definitions [4]. The MLM matching is therefore widely used in this work. There is an energy threshold below which parton showers are not included in the consideration whether or not to keep the event. This energy threshold is slightly higher than the lower bound for parton radiation energy at ME level. If this energy threshold is set too high, there will be a frequent occurrence of events being rejected due to a lack of available parton clusters to match with the outgoing partons from the hard subprocess. On the other hand, a low threshold will result in many events being rejected due to an overabundance of parton clusters compared to the number of outgoing partons.

To initiate the event generation process practically, one has to select a nominal PDF to determine the likelihood for a sufficient amount of momentum transfer to generate a given event. A random seed is set that will initiate a sequence of randomised selections of momentum transfer from the PDFs along with the specific Feynman diagram to generate events. There are intermediate heavy particles being produced that decay practically instantly, such as the W, Z and Higgs bosons as well as the top quark and additional new particles with similar qualities. The final state particles are either leptons, photons or partons that live long enough to hadronize. Subsequent decays or hadronizations from these final state particles are not included at this stage. Any new physics one wishes to include in these Feynman diagrams need to be accessed from a so-called UFO file, with UFO being short for 'the Universal Feynrules Output' [5].

The product is a sequence of generated events with all the relevant information on the particles produced in the hard subprocess in a format called Les Houches event (LHE) [6], which lists the particle ID, origin, momentum in the x, y, z directions, the mass, spin, colour etc. in a conventional order. The simulated events are unweighted, such that events with a higher probability to occur are produced more often compared to events with a lower corresponding probability. When a sufficiently large sample of events is generated, the sum of all events corresponds to the theoretical cross-section.

For convenience and consistency in generating additional LHE events with specific settings, a gridpack is generated which contains the full set of Feynman diagrams describing the hard subprocess, up to the selected order of intermediate loops and parton radiation as well as the PDFs for calculation and reweighting. A gridpack also contains a pilot run of 2000 LHE events that are used to estimate the cross-section of the given process. The only adjustable options from the gridpacks once they have been made are the number of LHE events to generate and the random seed for the event generation.

6.1.2. Generation of Full Events

In addition to the hard subprocess from the proton-proton collision, there are also soft processes involved from various sources. Starting from the collision itself, there are interactions between the remaining partons that have been ignored in generating the hard process. These soft interactions between the partons from the remnants of the two colliding protons, called the underlying event, are simulated by the same MC generator that generates the parton showers [7, 8].

Furthermore, subsequent decays of unstable partons produced in the previous step of the generation of the hard subprocess are simulated, and all coloured partons are fragmented into colourless hadrons. The latter simulation step is called hadronization. Generators typically use either string or cluster fragmentation [7, 8]. At this stage, events that have been generated with both the hard and soft processes are called GEN level events.

Finally, each GEN level particle is simulated as it passes through the detector. The software used for this simulation is called GEANT4 [9], which takes into account bending from the magnetic field, momentum loss according to the material budget of the detector components and the electronic response of the detector. After the full GEN event is simulated, one can add pileup, defined in section 4.1.2, which is taken from a library of simulated minimal bias events with a slightly shifted vertex. This procedure mimics the noise generated from pileup in data.
The simulated GEN events (GENSIM) are then processed with regards to the simulated detector response in the same way as data with the L1 trigger, HLT and object reconstruction as described in chapter 4.

As a final result, simulated events can be directly compared to experimental data.

6.2. Dark Higgs Samples

Official Dark Higgs samples have been produced for 2016 analysis. An extensive validation study has been performed on the Monte Carlo (MC) generation for the Dark Higgs process, since it is the first time in CMS that MC generation for the Dark Higgs process has been produced. A privately produced set of Dark Higgs samples has been used to test the performance in analysis, to estimate an appropriate range of signal points to request. A list of the requested samples is given in the end of this section.

6.2.1. Dark Higgs Validation Study

The first step in validating the new Dark Higgs MC production is to check that the UFO file is set up correctly. Samples for validation are produced at GEN level, meaning that the hadronization and showering of the generated particles

are simulated. The GEN files are then analysed using the program Rivet [10]. A reference sample has also been produced, using the UFO file for the phenomenology study [11], after which there have been some modifications to extend the available Dark Higgs hypotheses.

The GEN samples contain $2.5 \cdot 10^5$ events before matching and are made with the masses: $m_s = 90$ GeV, $m_\chi = 200$ GeV and $m_{Z'} = 1200$ GeV. They are generated from gridpacks, produced with Madgraph5AMC@NLO [12] (Madgraph) version 2.3.3. The PDF set used is NNPDF30_lo_as_130_nf_4, and the showering is done with pythia8 using the CUETP8M1 tune with the CMS software (CMSSW). The jets are matched with MLM matching using the anti-k_t jet algorithm [13].

To ensure that the reference sample is consistent with the samples produced for the phenomenology study, a reference sample has been produced with the workflow for the samples used in the phenomenology study using standalone software.

The pseudorapidity (η) and transverse momentum (p_T) distributions for the Dark Higgs (s) and the vector mediator (Z') are evaluated at matrix element (ME) level in order to ensure consistency before any showering has been applied. At parton shower (PS) level, the η and p_T distributions are given for Dark Matter (DM) along with the Z' rapidity. Distributions for Z', derived from the sum of the four momenta of the final state DM pair that the Z' decays into, are also given for p_T, η and invariant mass.

Predictions for distributions at analysis level are also evaluated, first by comparing the fat jet multiplicity and p_T after a cut on the missing transverse energy (E_T^{miss}) of $E_\mathrm{T}^{miss} > 500$ GeV. Afterwards, the fat jets are evaluated after a full range of selection cuts, that have been made consistent with that of the phenomenology study, given in table 6.1.

Object	Selection
Track-jets	$\Delta R = 0.2$, $\lvert\eta\rvert < 2.5$, $p_\mathrm{T} > 10$ GeV, $n_\mathrm{tracks} \geq 2$
Fat Jets	$\Delta R = 1.0$, $\lvert\eta\rvert < 2$, $p_\mathrm{T}^\mathrm{trimmed} > 250$ GeV, $n_b \geq 2$
Fat Jet Veto	Veto event if $n_\mathrm{Fatjets} \neq 1$
Veto Electron	$\lvert\eta\rvert < 2.47$, $p_\mathrm{T} > 7$ GeV
Veto Muon	$\lvert\eta\rvert < 2.5$, $p_\mathrm{T} > 7$ GeV

Table 6.1.: List of selection cuts in rivet analysis for Dark Higgs

At ME level, the reference samples are consistent overall as seen in figure 6.1.

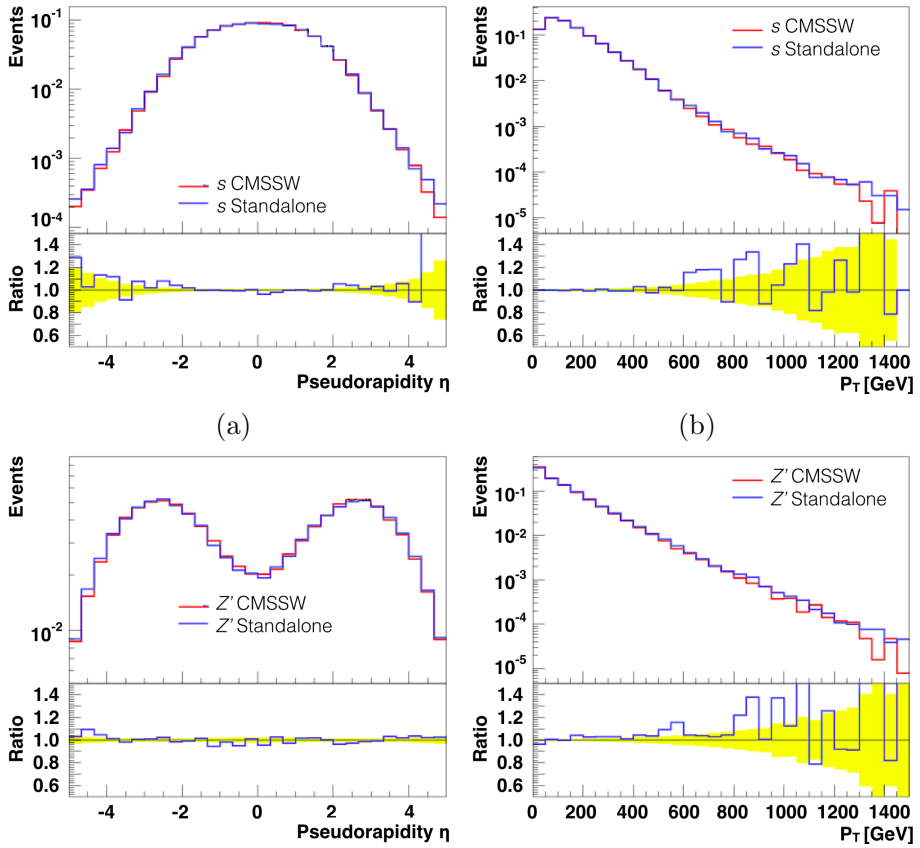


Figure 6.1.: Normalised distributions for pseudorapidity η (a) and transverse momentum p_{T} (b) for the Dark Higgs along with η (c) and p_{T} (d) for Z' at ME level compared between the early production (blue) and the reference sample made with Madgraph for the CMSSW build (red). The yellow shaded area is the MC statistical uncertainty.

There are however slight slopes in the p_{T} ratios, most clearly visible for the Z' (fig. 6.1d). It is noted that the pseudorapidity for Z' peaks around $|\eta| = 3$ instead of $\eta = 0$ which is the most common distribution for pseudorapidity. A deviation from a distribution in pseudorapidity peaking at zero is not unprecedented, for example the pseudorapidity distribution for the SM Z boson [14]. The rapidity distribution for Z' in fig. 6.2c is also slightly narrower in the reference sample from gridpacks with Madgraph than the standalone sample. The narrow rapidity distributions are what one expects when the momentum is low, which is typical for heavy particles like Z'. This is a special case where the rapidity is significantly different from the pseudorapidity. Since the PDF used in the phenomenology paper is different from the PDF used in the production, one can expect slight differences, which in this case are seen most clearly in the p_{T} distributions for the generated particles in the events.

The observables after showering are generally consistent within the statistical errors, that are rather large for events passing the full set of selection cuts (figures 6.3c, 6.3d, 6.3e and 6.3f). There is also consistency when the statistical error is more constrained in distributions where only the $E_{\mathrm{T}}^{miss} > 500$ GeV cut has been applied (figures 6.3a and 6.3b).

Having confirmed that the reference samples are consistent in key parameters at different simulation levels, one can proceed with the validation of the new Dark Higgs production.

In the distributions at ME level, it is observed that there is a deficit in Z' mediators that require investigation (see figures 6.4c and 6.4d). The Dark Higgs distributions are also significantly different as seen in figures 6.4a and 6.4b.
When the Z' parameters are derived from the final state DM pair however, the distributions are consistent with a slight increase of the stronger of the two Z' mass resonances in the new sample, and a slight decrease in the weaker Z' mass resonance as seen in figures 6.5c, 6.5d and 6.5e. The two peaks of the Z' mass derived from the two DM particles originate from the two different Feynman diagrams shown in section 2.2 where the Z' is either generated on-shell with one of the DM particles radiating off a Dark Higgs or off-shell radiating off a Dark Higgs, leaving an on-shell Z' that is fully reconstructed from the DM particles. The problem causing the severe inconsistencies in fig. 6.4 has been investigated with the following conclusion: When calling the particle ID, which follows the convention from the particle data group (PDG ID) [15] of the Z' mediator, an internal bug in Madgraph reveals itself that causes 97% of the Z' mediators to be labelled as either an SM Higgs or a Dark Higgs. This mislabelling also leads to significant

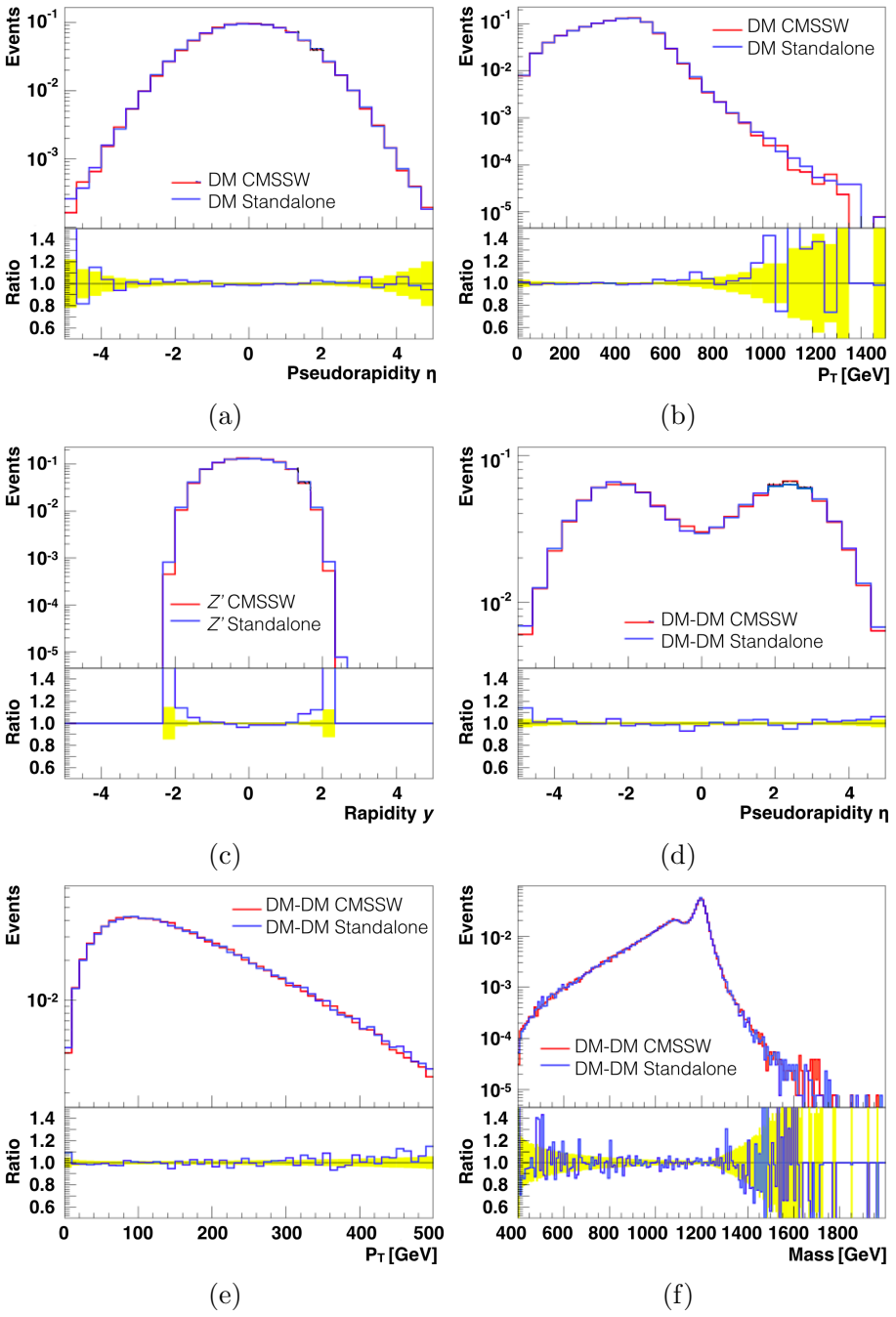


Figure 6.2.: Normalised distributions of pseudorapidity η (a) and transverse momentum p_{T} (b) for DM, the rapidity for Z' (c) and parameters for Z' derived from the four momenta of outgoing DM particles. In (d) the pseudorapidity is shown, (e) shows the transverse momentum from the same system. The invariant mass of the system is shown in (f) compared between the early production (blue) and the reference sample generated with CMSSW (red). The yellow shaded area is the MC statistical uncertainty.

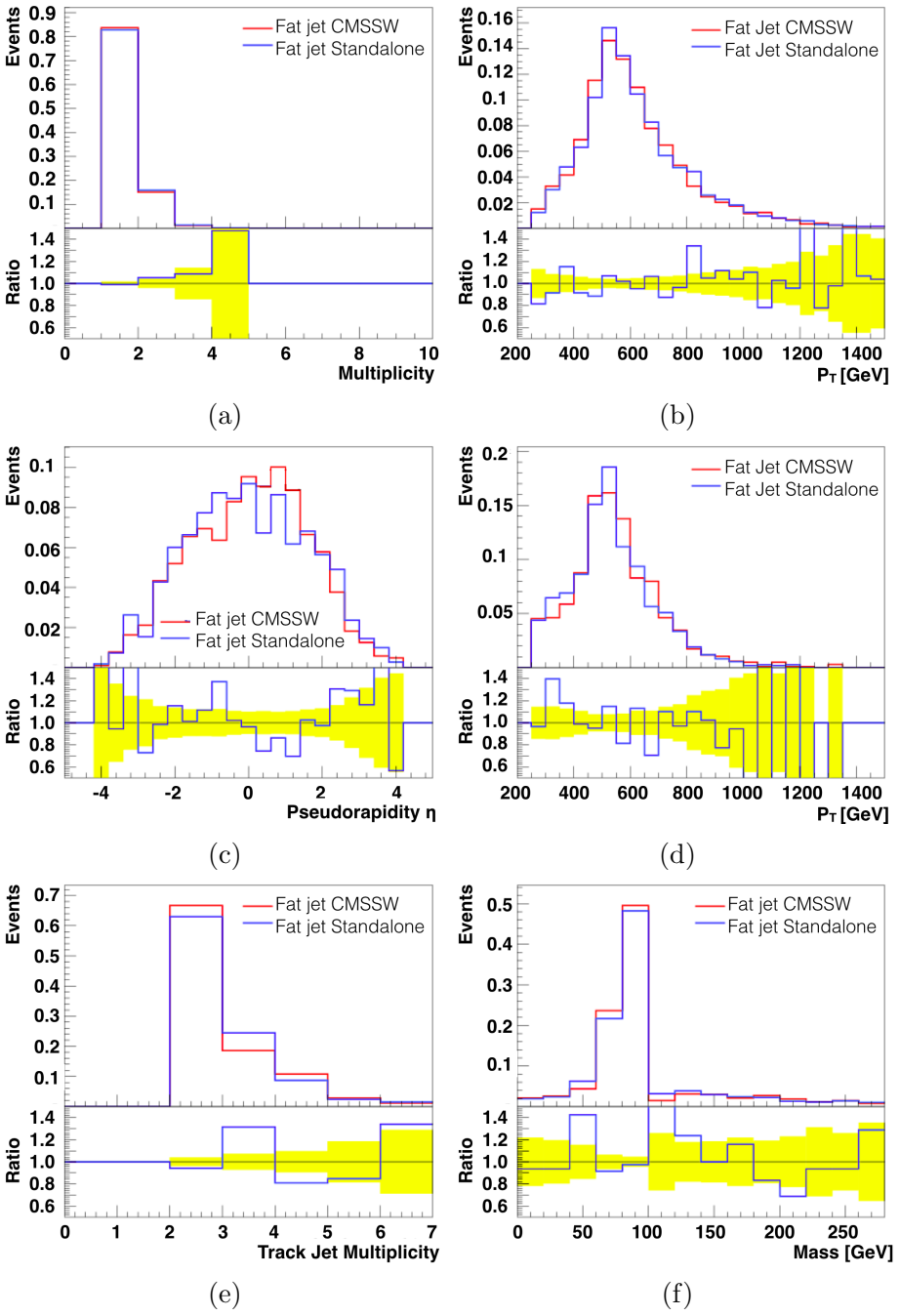


Figure 6.3.: Normalised distributions for the fat jet multiplicity (a) and p_{T} after an $E_{\mathrm{T}}^{miss} > 500$ GeV cut. After the full set of cuts, the η (c) and p_{T} (d) distributions are given along with the number of track-jets in the trimmed fat jet (e) and the mass of the fat jet (f) compared between the early production (blue) and the reference sample made with Madgraph for the CMSSW build (red). The yellow shaded area is the MC statistical uncertainty.

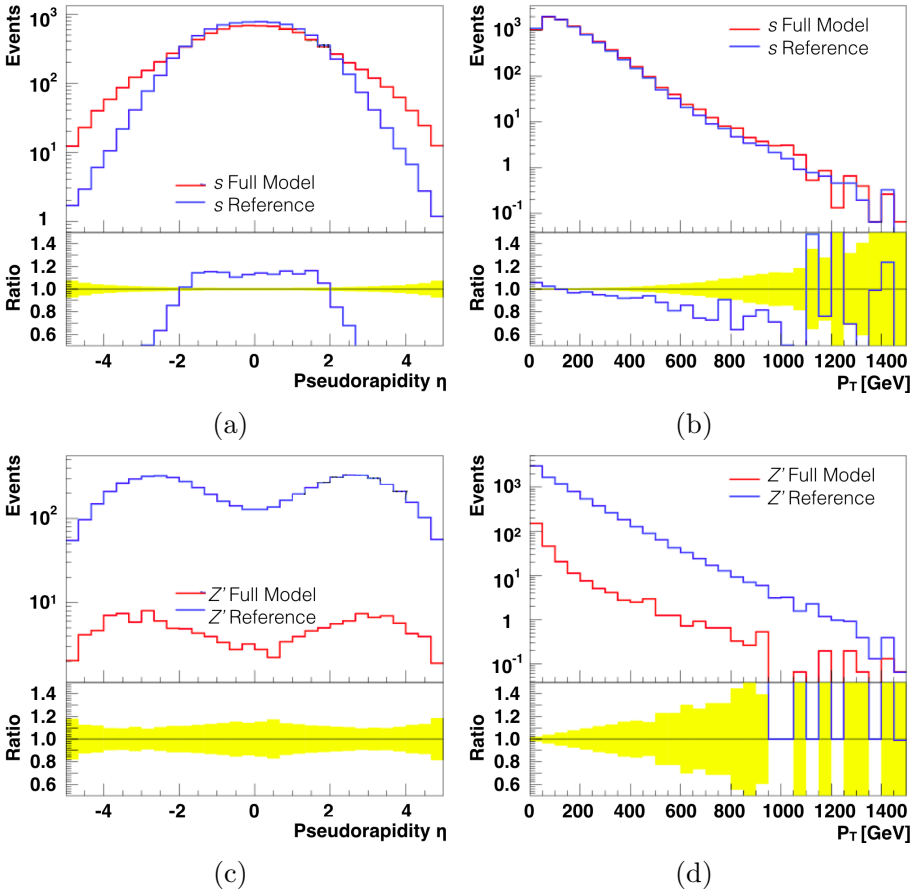


Figure 6.4.: Distributions normalised to an integrated luminosity of 40 fb^{-1} for pseudorapidity η (a) and transverse momentum p_{T} (b) for the Dark Higgs and η (e) and p_{T} (f) for Z' at ME level compared between the reference sample generated with CMSSW from the old UFO file (blue) and the sample based on the new UFO file without Z' specified as mediator (red). The yellow shaded area is the MC statistical uncertainty.

differences in the Dark Higgs distributions obtained by calling the PDG ID as seen in figg. 6.4a and 6.4b. The authors of the Dark Higgs UFO file have been notified about this issue, and intend to fix it in a future version. However, this does not occur when using the standalone software.

When considering distributions that indicate what impact there can be on an analysis, one finds that the fat jet distributions are consistent as shown in figures 6.6a, 6.6b and 6.6c.

The internal Madgraph issue, causing mislabelling of Z' into either a Higgs or a Dark Higgs, is about to be resolved. In the meantime, the best approach is to limit the available processes to the ones that occur via Z'. The impact on the cross-section is negligible, since alternative DM productions associated with Dark Higgs are strongly suppressed. Since the distributions for observables show no impact from the mislabelling, it is not expected that the mislabelling can cause problems in later stages of simulation and analysis. There are possible instances where this mislabelling may cause problems, for example filtering where one selects a specific subset of the simulated processes to keep before proceeding with the subsequent simulation steps. Since the decay products of Z' are set to be DM particles, it is considered unlikely that a need will arise to call the PDG ID of the Z' particle with regards to setting up a filter for production of a specific subset of the signal.

To prove that the approach of limiting the available processes to the ones that involve Z' in the new UFO file has the desired impact on the PDG ID labelling, a comparison has been made between the reference sample and a sample made with the recommended constraints. From figure 6.7 it is seen that the distributions at ME level are consistent with the reference sample, and that the mass of the DM, DM system follows the same trend as observed when the full set of processes in the new UFO are allowed. With the MC generation of Dark Higgs samples validated, official production of fully simulated Dark Higgs samples can take place.

6.2.2. Dark Higgs Sample Production

To begin with, three samples have been produced for different mass points in the (DM, Z') plane with a Dark Higgs mass of 70 GeV. These function as reference samples and are used to estimate the required extent of the search grid to set limits. The gridpacks have been produced the same way as for the validation study. The cut on jet k_{T} at parton shower level used in the matching (qcut) has been adjusted to improve the matching efficiency. The samples are first produced in GENSIM

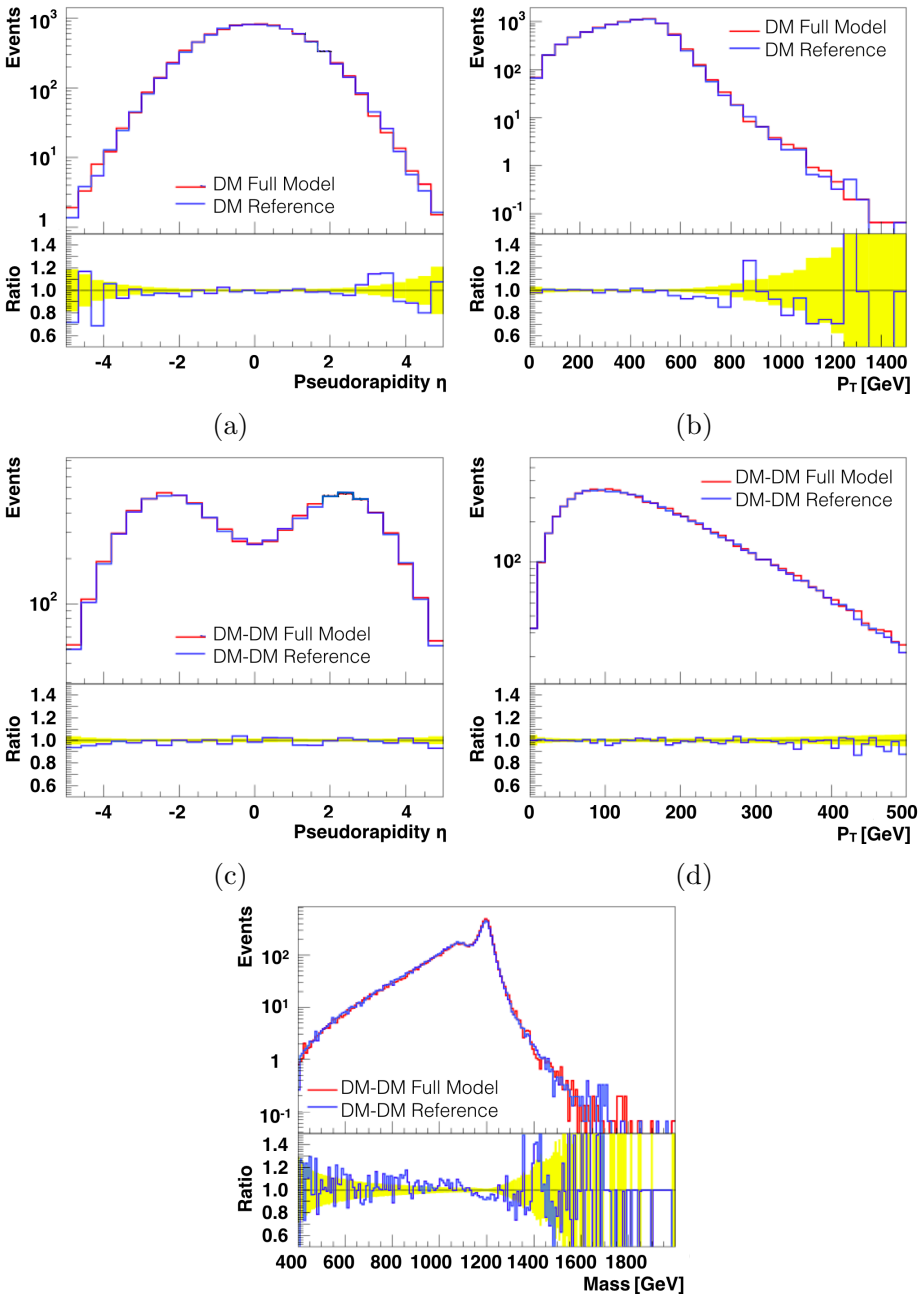


Figure 6.5.: Distributions normalised to an integrated luminosity of 40 fb^{-1} for DM with the pseudorapidity η in (a) and the transverse momentum p_{T} in (b). The following distributions are for Z' derived from the final state DM, DM system: η (c), p_{T} (d) and the mass (e) compared between the reference sample generated with CMSSW from the old UFO file (blue) and the sample based on the new UFO file without Z' specified as mediator (red). The yellow shaded area is the MC statistical uncertainty.

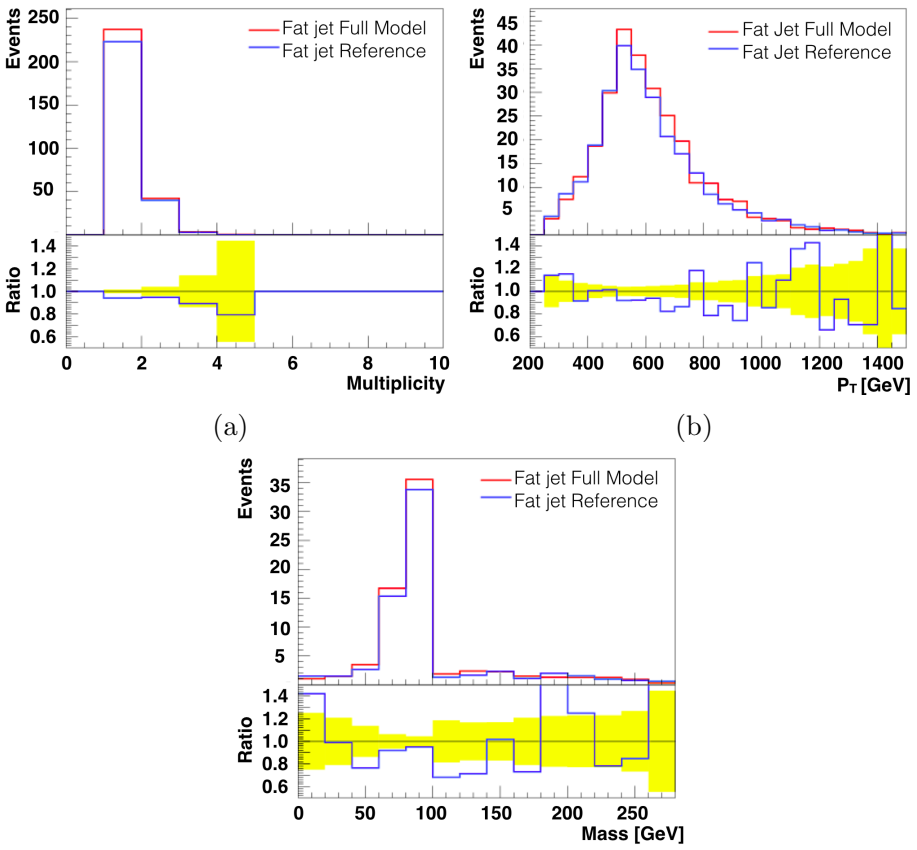


Figure 6.6.: Distributions normalised to an integrated luminosity of 40 fb^{-1} for fat jet multiplicity (a) and p_{T} (b) after an $E_{\mathrm{T}}^{miss} > 500$ GeV cut. The mass of the fat jet (c) after the full cut sequence compared between the reference sample generated with CMSSW from the old UFO file (blue) and the sample based on the new UFO file without Z' specified as mediator (red). The yellow shaded area is the MC statistical uncertainty.

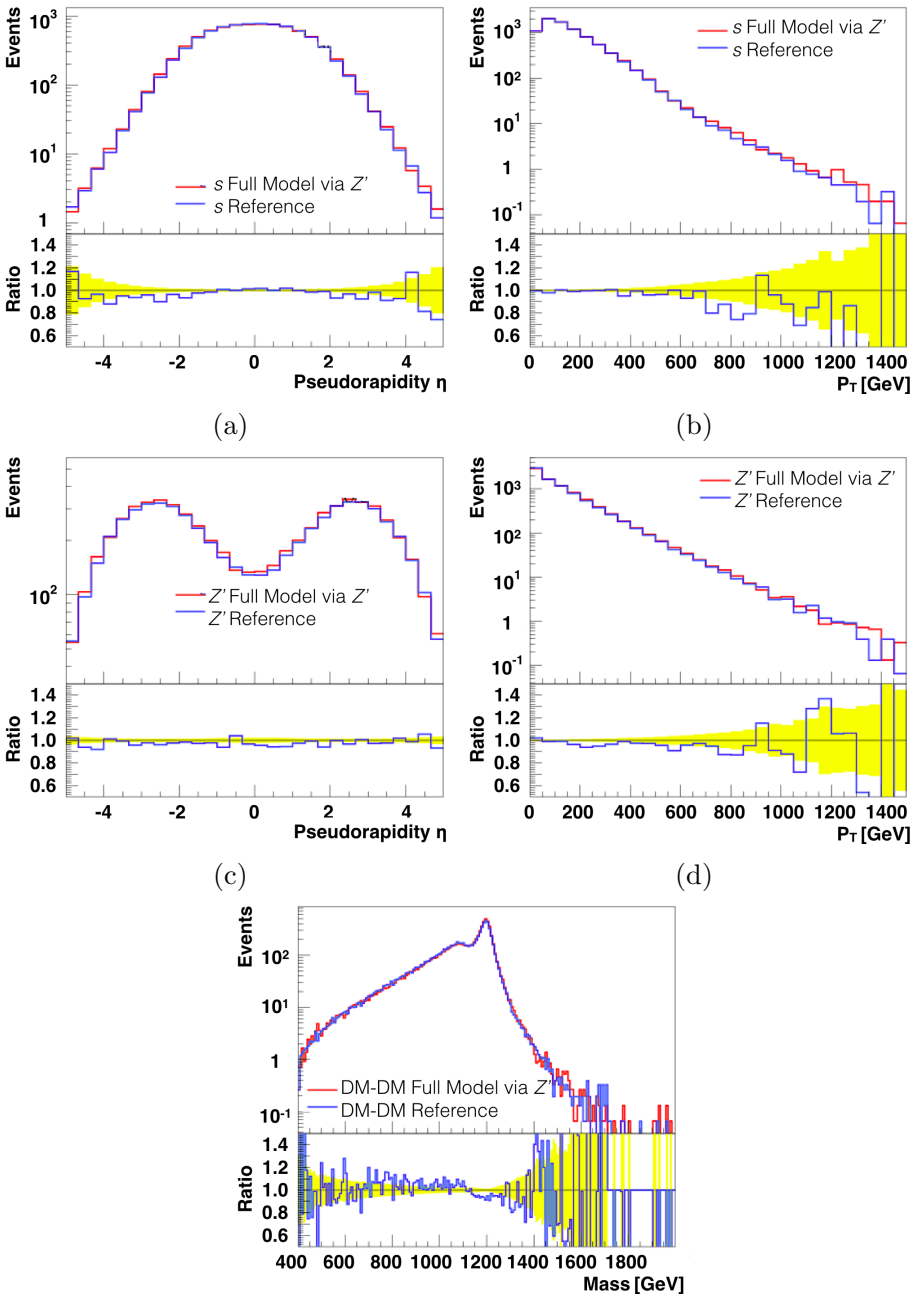


Figure 6.7.: Distributions normalised to an integrated luminosity of 40 fb^{-1} for pseudorapidity η (a) and transverse momentum p_{T} (b) for the Dark Higgs along with η (c) and p_{T} (d) for Z' at ME level. The mass of the final state DM, DM system is given in (e) compared between the reference sample generated with CMSSW from the old UFO file (blue) and the sample based on the new UFO file with Z' specified as mediator (red). The yellow shaded area is the MC statistical uncertainty.

format where detector simulations are done with GEANT4 set to simulate the CMS detector setup during the 2016 data taking. Pileup is then mixed in the process to mimic the 2016 pileup rates. Finally, the information is reduced to MINIAOD format which is suitable for analysis. Each sample is contains $3 \cdot 10^5$ events before matching.

With limits evaluated for the three Dark Higgs samples, an official production has been made. The gridpacks for the official samples are made with Madgraph5aMC@NLO v.2.4.2 and for the hard process the PDF set NNPDF31_nnlo_as_118_nf_4 has been used, with additional PDF sets included via reweighting, including NNPDF30_lo_as_130_nf_4. The production of the GEN events and beyond has been made with 2016 settings, which include the CUETP8M1 tune, to be compatible with the intended analysis on 2016 data. The efficiency of the MLM matching varies with the choice of masses for Z' and DM in the signal. To obtain an estimate of the matching efficiency, GEN samples with 10^4 events for each requested mass point have been produced. The list of requested mass points, their matching efficiencies and their cross-sections are shown in table 6.2.

$m_{Z'}$ [GeV]	m_χ [GeV]	Matching Efficiency [%]	Cross-section Before Matching [pb]
500	100	61.9	5.425
500	200	61.7	1.156
1000	100	56.1	1.166
1000	200	55.9	0.7935
1000	300	53.9	0.4325
1000	400	56.1	0.1663
1000	500	55.4	0.006842
1500	100	53.1	0.2713
1500	200	53.3	0.2270
1500	300	54.7	0.1717
1500	400	53.7	0.1193
1500	500	52.5	0.07414
1500	600	52.7	0.03722
2000	100	49.9	0.07298
2000	200	52.2	0.06508
2000	300	50.5	0.05526
2000	400	52.0	0.04469
2000	500	51.6	0.03463
2000	600	50.1	0.02523
2500	100	49.0	0.02105
2500	200	52.2	0.01947
2500	300	50.5	0.01745
2500	400	53.9	0.01506
2500	500	50.9	0.01266
2500	600	51.6	0.01034
3000	100	50.8	0.006189
3000	200	49.9	0.005842
3000	300	53.0	0.005372
3000	400	48.5	0.004863
3000	500	49.9	0.004282
3000	600	52.7	0.003688

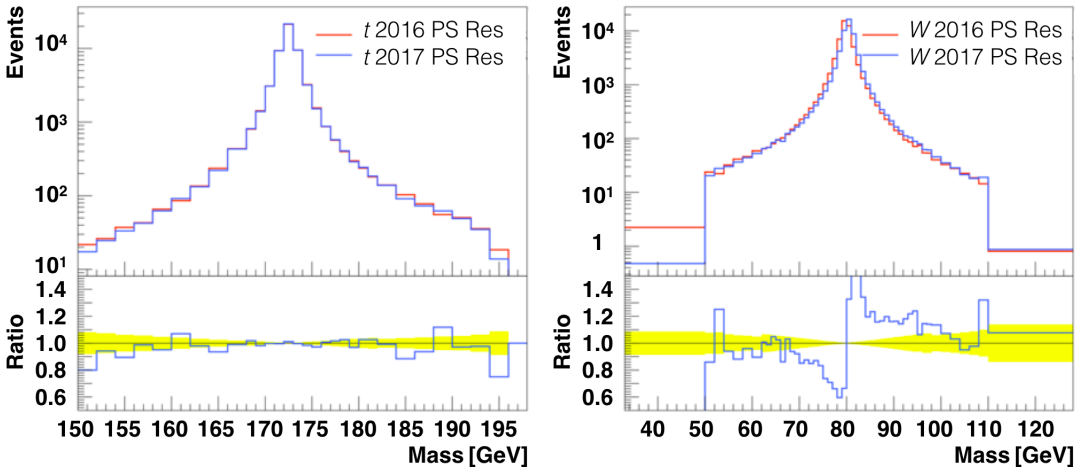
Table 6.2.: List of Dark Higgs signal points with $m_s = 70$ GeV requested and produced.

6.3. Heavy Higgs Samples

The MC production for the second analysis in this thesis, the search for Heavy Higgs bosons (see section 2.3.1) is produced with Madgraph5aMC@NLO v.2.6.0 using NNPDF31_nnlo_as_0118_nf_4. The validation for the Heavy Higgs production is to check that events generated from the new gridpacks are consistent with the earlier production to spot any potential bugs or errors in the setup. For validation purposes, samples are produced at GEN level using CMSSW_9_3_14. Both resonance and interference samples are produced for scalar and pseudoscalar events at the masses 400 GeV and 600 GeV both at a relative width of 2.5%. The showering is done with Pythia 8 using the CP5 tune as recommended by the CMS group. The samples are compared with 2016 GENSIM samples used to make signal samples for an earlier analysis [16]. As seen in fig. 6.8b, there is a shift in the W mass distribution at matrix element level, which is expected since the W mass setting in the 2016 production deviates slightly from the recommended value. The parameters for the W boson have therefore been changed from $M_W = 79.824$ GeV to $M_W = 80.419$ GeV and from $\Gamma_W = 2.085$ GeV to $\Gamma_W = 2.0476$ GeV. The top decay width has also been adjusted from $\Gamma_t = 1.5083$ GeV to $\Gamma_t = 1.4915$ GeV, but the difference in the top mass distribution is within statistical errors (see fig. 6.8a). The changes for the W and top mass distributions result in a small change in cross-section amounting to around 5% for resonance samples. This change in cross-section also transfers to the interference samples, where the negative cross-section decreases further by about 5%.

An investigation into the analysis parameters c_{hel} and $m_{t\bar{t}}$ has been performed to ensure that the distributions are consistent with those from 2016. This is shown for the $m_\Phi = 400$ GeV mass point with the resonance and interference separately displayed for both the scalar and pseudoscalar case with a relative width of 2.5%. The distributions are shown after a cut sequence ensuring that only good events are included that can be used for analysis. The cut sequences for the spin correlation analysis and the validation analysis are slightly different and are given in table 6.3. The $m_{t\bar{t}}$ distributions in fig. 6.9 all have the same feature of a slight shift in the $t\bar{t}$ mass peak, this is probably due to the adjusted W and top masses. It is also difficult to separate the pseudoscalar case in figures 6.9a and 6.9b from the scalar case in figures 6.9c and 6.9d based on the $t\bar{t}$ mass distributions alone. One can make use of one of the spin correlation variables to easier tell the two cases apart. The c_{hel} is one such variable that has proven effective.

The c_{hel} variable, also called $\cos\varphi$ in literature, is defined as the cosine of the angle between the two charged leptons originating from each of their top quarks for a



(a) (b)

Figure 6.8.: Mass distributions at ME level for top quarks (a) and for W bosons (b) for a pseudoscalar with a mass of 400 GeV and a relative width of 2.5% compared between 2017 GEN samples and GENSIM samples from the 2016 production. The yellow shaded area is the MC statistical uncertainty.

Object	Selection
Jets	$\Delta R = 0.4$, $\lvert\eta\rvert < 2.5$, $p_{\mathrm{T}} > 30$ GeV
Electron	$\lvert\eta\rvert < 2.47$, $p_{\mathrm{T}} > 20$ GeV
Muon	$\lvert\eta\rvert < 2.5$, $p_{\mathrm{T}} > 20$ GeV
Veto Electron	$\lvert\eta\rvert < 2.47$, $p_{\mathrm{T}} > 15$ GeV
Veto Muon	$\lvert\eta\rvert < 2.5$, $p_{\mathrm{T}} > 15$ GeV
Channel	**Cut**
ee	$n_e = 2$, $n_e^{\mathrm{veto}} \leq 2$, $n_\mu^{\mathrm{veto}} = 0$, $e_1 = \bar{e}_2$
$\mu\mu$	$n_\mu = 2$, $n_\mu^{\mathrm{veto}} \leq 2$, $n_e^{\mathrm{veto}} = 0$, $\mu_1 = \bar{\mu}_2$
$e\mu$	$n_e = n_\mu = 1$, $n_e^{\mathrm{veto}} \leq 1$, $n_\mu^{\mathrm{veto}} \leq 1$, opposite charge
$ee, e\mu, \mu\mu$	$m_{ll} > 20$ GeV for validation, $m_{ll} > 0$ GeV for spin correlation
$ee, \mu\mu$	81.2 GeV $< m_{ll} <$ 101.2 GeV for validation, not applied for spin correlation
$ee, e\mu, \mu\mu$	$n_{jet} > 2$
$ee, \mu\mu$	$E_{\mathrm{T}}^{miss} > 20$ GeV for validation, not applied for spin correlation
$ee, e\mu, \mu\mu$	$n_b = 2$ for spin correlation, not applied for validation
$ee, e\mu, \mu\mu$	Isolation of leptons within $\Delta R = 0.4$

Table 6.3.: Cut sequence in rivet analysis for Heavy Higgs

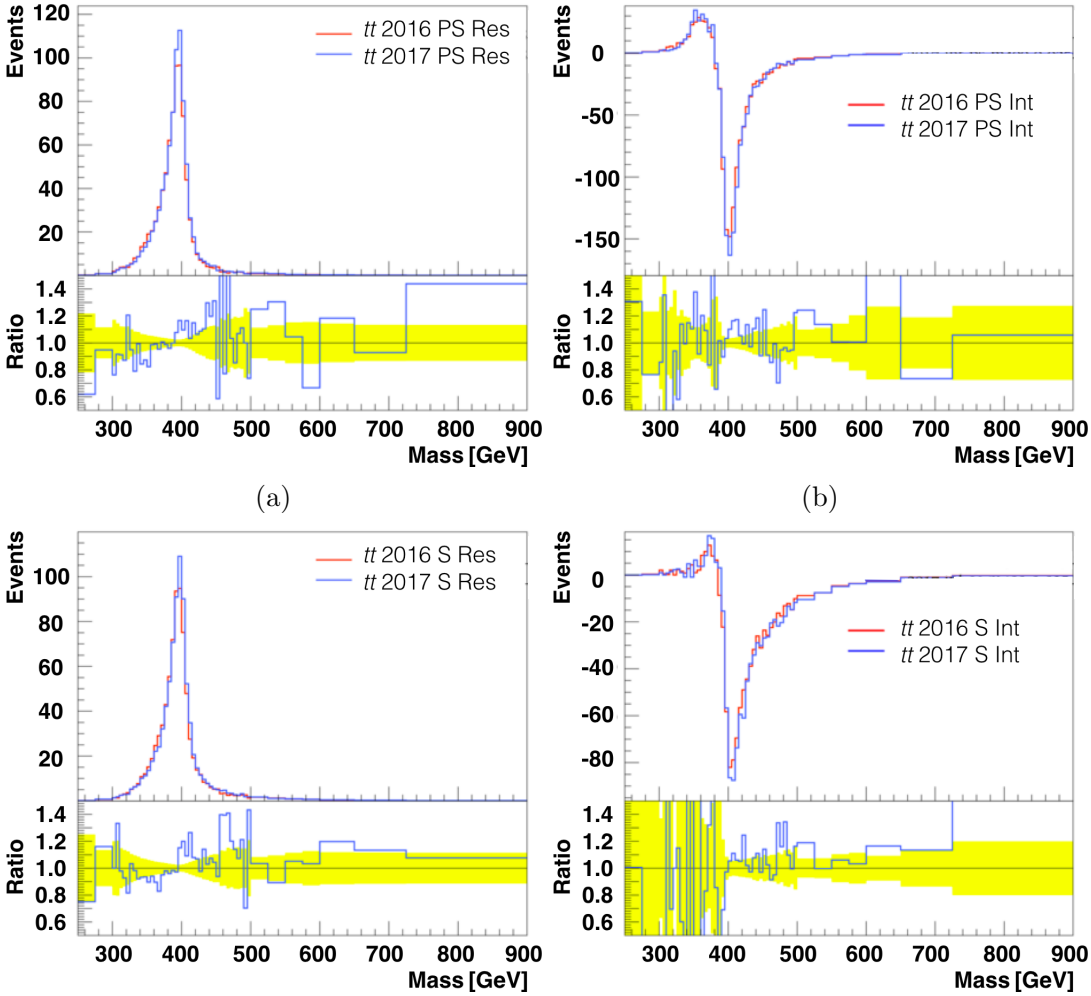


Figure 6.9.: Mass distributions for the $t\bar{t}$ after the application of cuts from the validation study (table 6.3) with the Heavy Higgs mass 0f 400 GeV and a relative width of 2.5% for the resonance of a pseudoscalar (a) and for its interference (b) for the resonance of a scalar (c) and its interference (d) compared between 2017 GEN samples and GENSIM samples from the 2016 production. The yellow shaded area is the MC statistical uncertainty.

$t\bar{t}$ system with dileptonic decays. The direction of an outgoing lepton is taken relative to the direction of the top quark it originates from in the rest frame of this given top quark by boosting from the rest frame of the $t\bar{t}$ system. When the directions of both charged leptons are found in this way, the cosine is obtained by taking the scalar product of their unit vectors [18].
The c_{hel} distributions differ significantly between the scalar and pseudoscalar cases as expected. The resonance for pseudoscalar has a steep positive incline (fig. 6.10a), whereas the resonance for the scalar has a slight decline (fig. 6.10c). The interference has positive and negative events cancelling each other out, and at the given mass of 400 GeV there are far more negative events than positive. The c_{hel} distributions for interference look like their respective resonance distributions with opposite signs (figures 6.10b and 6.10d). A general consistency is observed in the c_{hel} distributions between the new samples and the 2016 production.

Some interference samples have an almost equal amount of negative and positive weights, resulting in a cross-section that is very small even though there are significant effects to be observed, especially in the $t\bar{t}$ mass distribution. The rate of positive to negative events differs by up to 2% in the interference samples compared to the 2016 production. This is sometimes enough to cause a sign change in the cross-section, but when the positive and negative events from the interference samples are treated separately, the samples are consistent.

In the 2016 production, the GEN level weights are scaled in a way that the cross-section can be found by taking the average of the weights. This feature is not recreated for the new Heavy Higgs samples that have GEN level weights centred at ± 1. The LHE weights are given as the total cross-section divided by the number of generated events which are then given a negative sign for negative events.

For production, a set of samples has been chosen in which the total width at $g = 1$ equals the width of the decay into $t\bar{t}$, using the result from the Born approximation:

$$\Gamma_{\Phi \to t\bar{t}} = \frac{3 G_{\mathrm{F}} m_t^2}{4\pi\sqrt{2}} M_\Phi g_{\Phi t}^2 \left(1 - 4\frac{m_t^2}{M_\Phi^2}\right)^p . \text{ [17]} \tag{6.1}$$

Where p is 0.5 for pseudoscalar (A) and 1.5 for scalar (H) boson decays.
These widths are called the natural widths and the intention is to use reweighting on the SM $t\bar{t}$ events to replicate the Heavy Higgs events for analysis. The produced samples are therefore intended as reference samples for the reweighted events.

A filter has been applied to remove events with hadronically decaying tauons from the hard subprocess. The filter efficiency is around 55% and gives a sample with

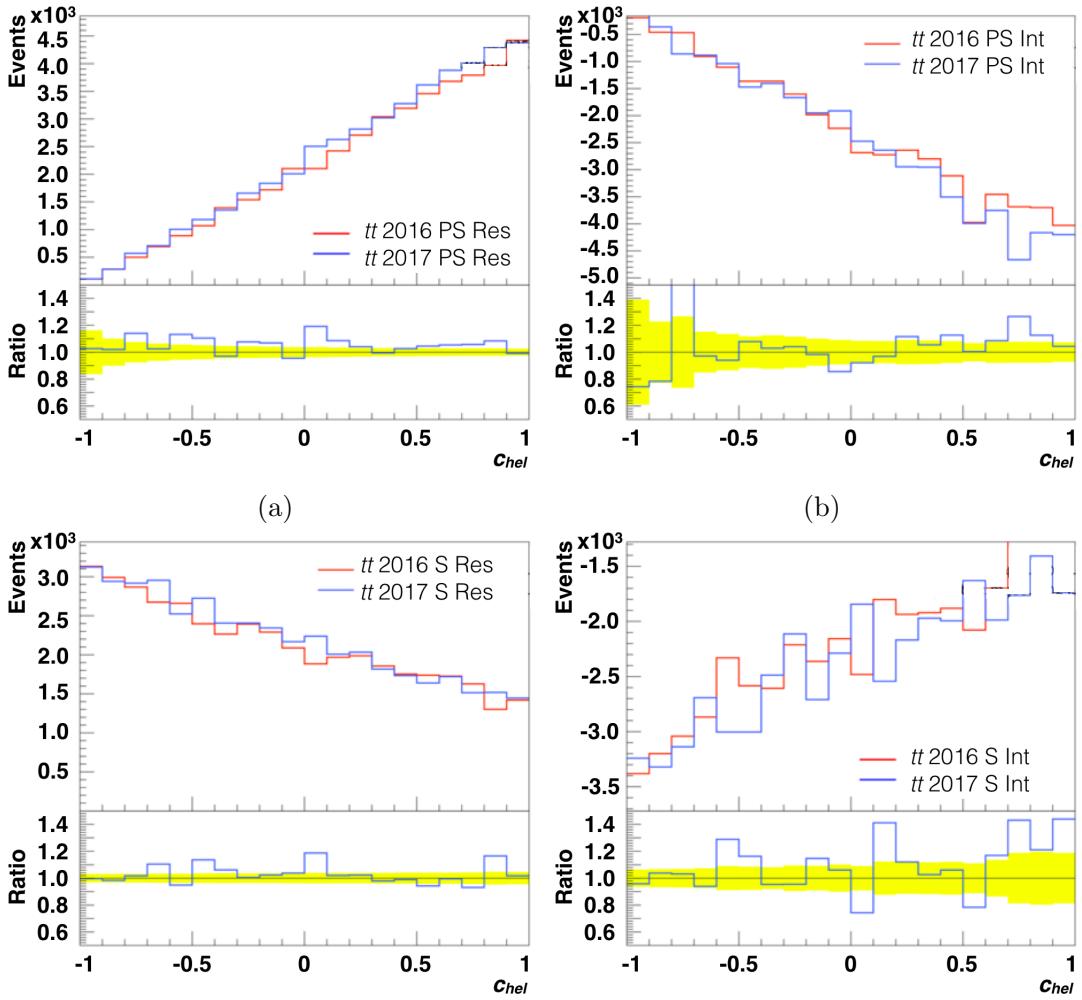


Figure 6.10.: Distributions for c_{hel} from the spin correlation study (table 6.3) with the Heavy Higgs mass 0f 400 GeV and a relative width of 2.5% for the resonance of a pseudoscalar (a) and for its interference (b) for the resonance of a scalar (c) and its interference (d) compared between 2017 GEN samples and GENSIM samples from the 2016 production. The yellow shaded area is the MC statistical uncertainty.

CP	Mass [GeV]	Relative Width [%]	LO Cross-section R [pb]	LO Cross-section I [pb]
A	365	1.9122	1.832	-1.468
A	400	2.9640	1.189	-0.8924
A	600	4.7919	0.1465	-0.0108
A	800	5.2844	0.02918	0.0446
A	1000	5.4974	0.007856	0.0382
H	365	0.20381	0.4850	-0.2679
H	400	0.75905	0.4284	-0.2596
H	600	3.2076	0.08907	-0.0295
H	800	4.3016	0.02023	0.0113
H	1000	4.8431	0.005652	0.0133

Table 6.4.: List of Heavy Higgs signal points requested and produced.

two leptons in the final state, giving higher Monte Carlo statistics in a dileptonic analysis. The samples have been produced with $5 \cdot 10^5$ events for 2017 analysis and $2 \cdot 10^6$ events for 2018 analysis. This event count is after filtering and will serve as an accurate reference for future analyses on Heavy Higgs signals.

Bibliography

[1] Bryan Webber, *'Parton shower Monte Carlo event generators'*, Scholarpedia, 6 (12):10662 (2011).

[2] T. Gleisberg et al., *'SHERPA 1.α, a proof-of-concept version'*, JHEP02 (2004) 056

[3] S. Hoeche et al., *'Matching Parton Showers and Matrix Elements'*, "HERA and the LHC" workshop, CERN/DESY 2004/2005, doi:10.48550/arXiv.hep-ph/0602031, arXiv:hep-ph/0602031.

[4] M. L. Mangano et al., *'Matching matrix elements and shower evolution for top- quark production in hadronic collisions'*, JHEP 01 (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[5] C. Degrande et al., *'UFO - The Universal FeynRules Output'*, Comput.Phys.Commun. 183 (2012) 1201-1214, doi:10.1016/j.cpc.2012.01.022, arXiv:1108.2040v2.

[6] J. Alwall et al., *'A standard format for Les Houches Event Files'*, Comput.Phys.Commun.176:300-304,2007, doi:10.1016/j.cpc.2006.11.010, arXiv:hep-ph/0609017.

[7] T. Sjstrand et al., *'An introduction to Pythia 8.2'*, Comput.Phys.Commun. 191:159-177, 2015, doi:10.1016/j.cpc.2015.01.024.

[8] M. Bahr et al., *'Herwig++ Physics and Manual'*, Eur.Phys.J.C58:639-707,2008, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883 [hep-ph].

[9] S. Agostinelli et al., *'GEANT4: A Simulation toolkit'*,

In: Nucl. Instrum. Meth. A506 (2003), pp. 250? 303. doi:10.1016/S0168-9002(03)01368-8.

[10] A. Buckley et al., *'Rivet user manual'*,
Comput. Phys. Commun. 184 (2013) 2803, arXiv:1003.0694v8.

[11] M. Duerr et al., *'Hunting the dark Higgs'*,
JHEP 1704 (2017) 143, doi:10.1007/JHEP04(2017)143, arXiv:1701.08780v2.

[12] J. Alwall et al., *'The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations'*,
JHEP07(2014)079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301v2.

[13] M. Cacciari et al., *'The anti-k_t jet clustering algorithm'*,
JHEP 0804:063,2008, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189v2.

[14] S. Dooling, *'Differential Cross Section Measurement of Drell-Yan Production and associated Jets with the CMS Experiment at the LHC'*,
PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2014.

[15] M. Tanabashi et al. (Particle Data Group), *'43. Monte Carlo Particle Numbering Scheme'*,
Phys. Rev. D 98, 030001 (2018) and 2019 update, pdg.lbl.gov/2019/reviews/rpp2019-rev-monte-carlo-numbering.pdf

[16] **CMS Collaboration**, *'Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV'*,
JHEP 04 (2020) 171, doi:10.1007/JHEP04(2020)171, arXiv:1908.01115v2.

[17] A. Djouadi et al., *'Fully covering the MSSM Higgs sector at the LHC'*,
arXiv:1502.05653v2 [hep-ph] 22 Jun 2015.

[18] W. Bernreuther et al., *'Top quark pair production and decay at hadron colliders'*,
Nucl. Phys. B 690 81 (2004) 81 [hep-ph/0403035].

7. Search for a Dark Higgs, using the α_{t} Analysis on 2016 Data

In this chapter, a search for a Dark Higgs signal is presented. Details about the investigated Dark Higgs scenario are discussed in section 2.2. The Dark Higgs is predicted to mix with the SM Higgs and thereby decay into two b-quarks and is produced in association with DM particles that will give a signature of missing transverse momentum, as seen in fig. 2.8. Due to the heavy vector boson Z' which acts as the mediator, the signal has a tendency to be boosted.

To search for this model, an existing search for natural and split supersymmetry in final states with jets and missing transverse momentum of the CMS Collaboration [1] has been utilised and transformed into a search for Dark Higgs.

The analysis [1] relies on the measurement of the so-called α_{t} variable. This variable is robust against energy mismeasurements, and controls background due to jet mismeasurements, caused by possible detection inefficiencies or by nonuniformities in the calibration of the calorimeters, that are the dominant source of large missing transverse energy in events from QCD multijet production. The α_{t} variable is able to separate this background from a genuine missing energy signal. In the following, the analysis [1] is referred to as the 'α_{t} analysis'.

The α_{t} analysis is sensitive to signal with a fully hadronic final state associated with significant amounts of missing transverse momentum (E_{T}^{miss}). These properties do indeed apply to the Dark Higgs, and the use of b-tagging in the analysis makes it even more applicable to the Dark Higgs. The analysis has some significant differences from the search suggested in [2], in particular that other features of the Dark Higgs signal are exploited than the highly boosted region. The analysis is a search for signal with jets and E_{T}^{miss} and has access to signal down to E_{T}^{miss}=150 GeV along with a variety of jet multiplicities, based on AK4-jets. The increased acceptance of signal of course comes with the price of higher background, which needs to be suppressed and controlled. For this purpose, the analysis makes use

of the α_t and $\Delta\phi^*_{min}$ variables, which effectively cut down the QCD background by several orders of magnitude to a level around 1% of the remaining background. To control the remaining background, control regions are used to establish data-driven constraints. Due to the properties of the Dark Higgs signal, it is expected to be less sensitive to the cuts imposed in the analysis, so that the relative signal yield to background is enhanced significantly. To further enhance sensitivity to signal, the search region is split into bins according to jet multiplicity, number of b-tagged jets and the total transverse momentum of the selected jets. Finally, the shape of the missing energy distribution from the selected jets is used as an additional enhancement of signal sensitivity.

The analysis itself is designed to cover a wide range of searches for supersymmetric counterparts to SM particles and is also considered for Dark Matter searches. It is therefore rather complex with a wide range of categories rather than cutting down to one category of particular interest.

7.1. MC Generated Background

The MC samples used for estimating the SM background are given in table 7.1.

The W+jets background involves W bosons that are generated in the hard process and decay into either a quark-antiquark pair or a lepton-neutrino set. Events from the channels with a decay into a lepton-neutrino set have a higher chance of getting selected in the signal region due to the escaping neutrino leaving a signature of missing energy in the detector. For these events to be accepted however, the lepton has to avoid triggering the lepton veto by either travelling along the z axis or close to a jet. The lepton can also escape undetected or have a transverse momentum below the veto threshold. The veto conditions for each flavour of leptons are shown in table 7.2. In the case of the W boson decaying to $\tau + \nu_\tau$, the tauon decays into either a lepton or a neutrino and a quark-antiquark pair typically forming pions. The neutral pions decay mainly into a pair of photons, whereas the charged pions generally live long enough to be stopped by detector material. The pions are therefore vulnerable to the photon veto and the single isolated track veto depending on their charge.

The $t\bar{t}$ backgrounds contain a top-antitop pair that decay before they hadronize. The dominant decay of the top quark is into a W boson and a bottom quark. The subsequent decay of the W boson can then either be into a quark-antiquark

pair or a lepton-neutrino set giving rise to three categories for $t\bar{t}$ events, which starting from the most common are: Fully hadronic with both W bosons decaying into quark-antiquark pairs, semileptonic with one W boson decaying into a lepton-neutrino pair and dileptonic where both W bosons decay into a lepton-neutrino set.

The $Z \to \nu\nu$ background is tricky, since the final state escapes detection and occurs as a significant amount of missing energy. It is one of the dominant backgrounds in this search along with $t\bar{t}$ and W + jets after a full selection and the rate of this background relies heavily on the theoretical assumption of lepton universality since this background cannot be directly measured experimentally.

The QCD processes cover events where only strong interactions between gluons and quarks or between gluons occur. Since QCD occurs in the underlying event by default, the multijet QCD background is simply where heavy SM or BSM particles have not been generated either because of insufficient momentum transfer or simple bad luck. In events where leptons and photons are absent, it is natural to have an overwhelming QCD background, which is effectively reduced by application of the $\Delta\phi^*_{min}$ and α_{t} cuts described in section 7.2. After the application of these cuts, the remaining QCD background is impossible to determine reliably by simple use of MC simulations due to statistical limitations one expects from a large sample where only very few events pass selection. On top of that, the theoretical uncertainty is also large for the QCD background. To obtain more reliable estimates on the QCD background, three control regions have been used to get data-driven corrections on these estimates as described in section 7.2.

The Drell-Yan background, which can either be a $Z \to l^+l^-$ decay or a $\gamma \to l^+l^-$ decay. Although this background is small in the signal region of the analysis, it is important in determining the yield and kinematics of the $Z \to \nu\nu$ background by use of the $\mu\mu$ + jets control region described in section 7.2.

Processes that involve the generation of two vector bosons are also present in the background. They can either be two W bosons, two Z bosons or a W boson and a Z boson.

Events with a single top also occur in the background that are generated through either the t-channel, the tw-channel or the s-channel. There are also some more complex backgrounds where a $t\bar{t}$ pair is produced along with an additional boson that can either be the Higgs, a photon, W or Z.

For simulations, the hadronic showering is done with Pythia 8 using the CUETP8M1 tune on all samples except for the inclusive $t\bar{t}$ sample, which is using the CUETP8M4 tune. All background samples have been matched using MLM matching, except for the $t\bar{t}+W$, $t\bar{t}+\gamma$ and $t\bar{t}+H$ samples where the FXFX method is used instead.

Category	Processes	ME Generator	Cross-section [pb]
$t\bar{t}$	$t\bar{t}$	Powheg	381.76
	$t\bar{t}$ semileptonic	Madgraph	
	$t\bar{t}$ dileptonic		
	$t\bar{t}$ $H_T > 600\,\text{GeV}$		
W + Jets	W + jets	Madgraph	61526.7
$Z \to \nu\nu$	$Z \to \nu\nu$	Madgraph	458.624
Multijet	QCD	Madgraph	$3.009 \cdot 10^7$
Drell-Yan	DY, $m_{ll} > 50\,\text{GeV}$	Madgraph	6025.2
ttH	$t\bar{t}+H$	Madgraph	0.5085
ttV	$t\bar{t}+W$	Madgraph	0.6105
	$t\bar{t}+Z$		0.7826
	$t\bar{t}+\gamma$		3.697
Diboson	WW	Pythia 8	113.898
	WZ		47.13
	ZZ		16.523
Single Top	t t-channel	Powheg	136.02
	$\bar{t}$ t-channel		80.95
	t tw-channel		35.6
	$\bar{t}$ tw-channel		35.6
	t s-channel	Madgraph	3.34368

Table 7.1.: List of MC samples used to estimate the SM background.

7.2. Selection Cuts

To find events that are relevant for analysis, a series of trigger algorithms are used with one trigger applied on $n_j \geq 2$, has cuts on H_T and α_t that are below those of the selection cuts in the analysis where events with $H_T > 900$ GeV are also allowed regardless of their α_t value. To include events where $n_j = 1$, there is a trigger with the requirements of $H_T^{miss} > 120$ GeV, $E_T^{miss} \geq 120$ GeV and at least

one jet with $p_T > 20$ GeV, $|\eta| < 5.2$. The trigger efficiencies are determined from control regions and are H_T dependent. In the 200 GeV $\leq H_T <$ 600 GeV region, the trigger efficiencies range from 97.4-97.9%. At $H_T \geq 600$ GeV, the trigger efficiency reaches 100%.

The selection requirements are listed in table 7.2, beginning with the objects for selection or veto. Key parameters for selections and cuts are explained below: The scalar sum of the magnitudes of the transverse momentum (p_T) of jets in a given event is defined as the variable H_T, whereas the variable H_T^{miss} refers to the magnitude of the negative vector sum of the transverse momenta of the jets. In table 7.2, ΔR indicates the isolation requirement given as the distance parameter $R = \sqrt{\eta^2 + \phi^2}$. The relative isolation, I^{rel} requirement evaluates the p_T within the given ΔR around the identified lepton. The following relation is used: $I^{rel} = (p_T^{\Delta R} - p_T^l)/p_T^l$, where p_T^l is the transverse momentum of the given lepton, $p_T^{\Delta R}$ is the total transverse momentum within the given ΔR cone. The parameter I^{track} is similar to I^{rel}, where the track p_T is used instead of p_T^l. The fraction of energy attributed to charged hadrons in the leading jet, $f_{h_\pm}^{j1}$ is indicative of the jet reconstruction quality.

Object	Selection
Jets	AK4, $p_T > 40$ GeV, $\|\eta\| < 2.4$
Soft Jets	AK4, $p_T > 25$ GeV, $\|\eta\| < 2.4$
Muons	$p_T > 30$ GeV, $\|\eta\| < 2.1$, $\Delta R < 0.5$
Veto Photons	$p_T > 25$ GeV, $\|\eta\| < 2.5$, $\Delta R < 0.3$
Veto Electrons	$p_T > 10$ GeV, $\|\eta\| < 2.5$, $I^{rel} < 0.1$, $0.05 < \Delta R < 0.2$
Veto Muons	$p_T > 10$ GeV, $\|\eta\| < 2.5$, $I^{rel} < 0.2$, $\Delta R < 0.3$
Veto Single Isolated Tracks	$p_T > 10$ GeV, $\|\eta\| < 2.5$, $I^{track} < 0.1$, $0.05 < \Delta R < 0.2$
Cuts for Signal Region	
None of the veto particles present in the event	
$E_T^{miss} > 150$ GeV	
$H_T > 200$ GeV	
$P_T^{jet1} > 100$ GeV	
$0.1 < f_{h_\pm}^{j1} < 0.95$	
Veto on events containing jets with $p_T > 40$ GeV and $\|\eta\| > 2.4$	
$H_T^{miss} > 200$ GeV	
$H_T^{miss}/E_T^{miss} < 1.25$	

Table 7.2.: Object selection.

The selected events are then categorised according to H_T, the number of selected

jets n_{jet} and the number of b-tagged jets n_b as shown in table 7.3.
Events where the sub-leading jet has 40 GeV $< p_{\text{T}} <$ 100 GeV are categorised as asymmetric, indicated by $\geq 2a$ in table 7.3.
The H_{T}^{miss} distribution for each bin is also stored to further enhance the separation

n_{jet}	n_b	H_{T}^{min} 1 [GeV]	H_{T}^{min} 2 [GeV]	H_{T}^{min} 3 [GeV]	H_{T}^{min} 4 [GeV]	H_{T}^{min} 5 [GeV]
1	0	200	400	600	900	-
1	1	200	400	600	-	-
2	0	200	400	600	900	1200
2	1	200	400	600	900	1200
2	2	200	400	600	-	-
3	0	200	400	600	900	1200
3	1	200	400	600	900	1200
3	2	200	400	600	900	1200
3	3	200	400	600	-	-
4	0	-	400	600	900	1200
4	1	-	400	600	900	1200
4	2	-	400	600	900	1200
4	3	-	400	600	900	-
5	0	-	400	600	900	1200
5	1	-	400	600	900	1200
5	2	-	400	600	900	1200
5	3	-	400	600	900	-
5	≥ 4	-	400	-	-	-
≥ 6	0	-	400	600	900	1200
≥ 6	1	-	400	600	900	1200
≥ 6	2	-	400	600	900	1200
≥ 6	3	-	400	600	900	1200
≥ 6	≥ 4	-	400	-	-	-
$\geq 2a$	0	200	400	600	900	-
$\geq 2a$	1	200	400	600	900	-
$\geq 2a$	2	200	400	600	900	-
$\geq 2a$	3	200	400	600	-	-

Table 7.3.: Event binning scheme used in analysis.

between signal and background. The H_{T}^{miss} binning follows the H_{T} binning starting with 200 GeV $\leq H_{\text{t}}^{miss} <$ 400 GeV up to and including the bin where the given H_{T} range applies for H_{T}^{miss} too, up to $H_{\text{T}}^{miss} \geq$ 900 GeV.

7.2.1. The α_t and $\Delta\phi^*_{min}$ Variables

Cuts are also applied on the parameters α_t and $\Delta\phi^*_{min}$ as shown in table 7.4. These variables are effective at suppressing the QCD multijet background by removing events with the H_T^{miss} mainly originating from misreconstructed jet energy.

H_T Bin [GeV]	α_t Cut	$\Delta\phi^*_{min}$ Cut
200-250	0.65	0.5
250-300	0.6	0.5
300-350	0.55	0.5
350-400	0.53	0.5
400-900	0.52	0.5
≥ 900	-	0.5

Table 7.4.: H_T specific cuts.

The α_t variable is calculated for a dijet system by the following equation:

$$\alpha_t = \frac{1}{2} \cdot \frac{H_T - \Delta H_T}{\sqrt{H_T^2 - (H_T^{miss})^2}}, \tag{7.1}$$

where ΔH_T is the difference in magnitude of the transverse momentum between the two jets. For events with $n_{jet} > 2$, the multiple jets are combined into a pseudo-jet pair where ΔH_T is the difference between the two scalar sums of the p_T of the jets forming each pseudo-jet. The combination of jets into pseudo-jets that minimises ΔH_T is selected for calculating the α_t variable [3]. When one has a pair of jets that are back to back with equal p_T, the α_t value is 0.5. If the jets are back to back and one jet has higher p_T than the other jet, the α_t value is less than 0.5. When the jets are equal in p_T and not back to back, increasing the likelihood of actual E_T^{miss}, the α_t value is greater than 0.5. The α_t variable has several advantages that make it a more favourable choice for QCD suppression than many other variables. The QCD distribution in α_t is largely unaffected by variations of the amount of soft gluon radiation in the given event. This feature is important when considering the large inaccuracies in predicting the QCD background, both with respect to the composition and the total amount. Another appealing feature is that the errors from the jets are propagated to the variable with a relatively small impact. It is also effective in reducing the $Z \to \nu\nu$ background [4].

$\Delta\phi^*_{min}$ is the azimuthal angle between a selected jet transverse momentum and the direction of H_T^{miss}, computed without the p_T contribution from the selected

jet. This process is repeated for all jets in the system to find the minimum value, which then has to be above the cut value of 0.5.
In events with only one jet, the soft jets with selection criteria shown in table 7.2 are included in the computation of $\Delta\phi^*_{min}$ to enable a comparable cut to be set. To demonstrate the effectiveness of the α_t and $\Delta\phi^*_{min}$ variables in suppressing the QCD multijet background a figure from a similar analysis [5] is shown in fig. 7.1, where a combined suppression rate reaches six orders of magnitude.

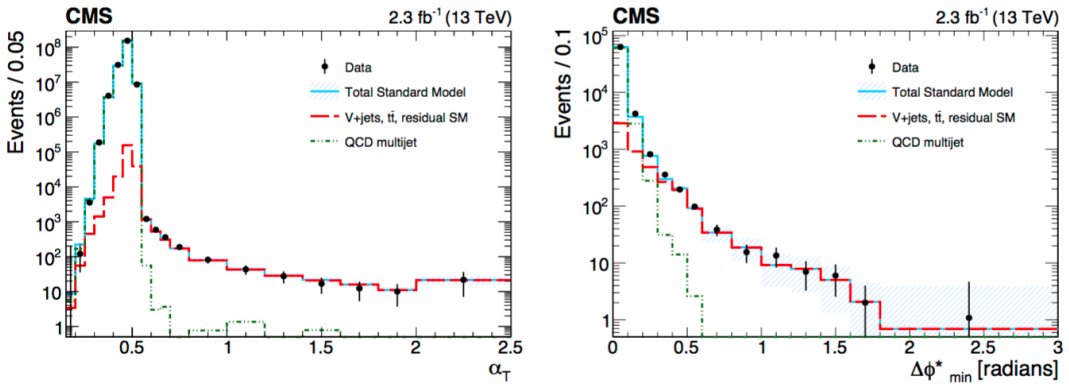


Figure 7.1.: Distributions of the QCD multijet background (green dashed lines) and the SM background from other sources (red solid line) in terms of α_t (left) and $\Delta\phi^*_{min}$ (right). The α_t distribution has a cut on $\Delta\phi^*_{min} = 0.5$ applied when $\alpha_t \geq 0.55$. The $\Delta\phi^*_{min}$ distribution is shown in a region where no cut in α_t is applied due to high H_T. Both figures taken from source [5].

The cut value of α_t is set by requiring the QCD multijet background to be lower than 1% of the total background after the application of both the α_t and $\Delta\phi^*_{min}$ cuts.

7.2.2. Control Regions

Control regions are also defined to give data-driven estimates of the SM background in the form of correction factors to the yields of different backgrounds along with determining the shape of the H_T^{miss} distributions and the associated uncertainties.

The control regions are as follows:
The μ + jets region, requiring one muon where the transverse mass of the muon -

E_{T}^{miss} system lies within the range of 30 to 125 GeV. Cuts on $\Delta\phi^*_{min}$ and α_t are omitted. This region has a high rate of W + jets events along with single top and diboson processes.

The $\mu\mu$ + jets region, requiring two oppositely charged muons. The two muon system is required to have a mass within 25 GeV of the mass of the Z boson. This region is mainly containing Drell-Yan events that are used to estimate the $Z \to \nu\nu$ background in the signal region.

A single muon trigger is used for both the μ + jets and $\mu\mu$ + jets regions, requiring at least one muon with $p_{\mathrm{T}} > 24$ GeV and $|\eta| < 2.1$. This trigger has an efficiency of 90% for the μ + jets region and 99% for the $\mu\mu$ + jets region. The binning in the μ + jets control region follows that of the signal region with the exception that the event count is integrated over H_{T}^{miss}, whereas the $\mu\mu$ + jets region has a reduced the n_b bins to $n_b = 0$ and $n_b \geq 1$ in addition to integrating the event count over H_{T}^{miss}.

QCD Control Regions

The QCD background has a very unreliable estimate from MC simulations alone, so control regions are used to better estimate the QCD background. To avoid assumptions based on MC simulations in determining and validating the transfer factor from one control region to the signal region, a set of three control regions based on two variables with a low enough correlation to consider them uncorrelated are used. There is no overlap in events between the control regions or between the signal region and the control regions. This enables the use of what is referred to as the ABCD method, where A is the signal region, B is the control region from which data is used to estimate the number of QCD events in the signal region with the transfer factor based on the ratio between the C and D control regions [6]. The three regions for estimating the QCD background, without relying on MC based transfer factors, are chosen where cuts have been inverted on either the $H_{\mathrm{T}}^{miss}/E_{\mathrm{T}}^{miss}$ ratio or the $\Delta\phi^*_{min}$ variable or both. The regions are as follows:

1) $1.25 < H_{\mathrm{T}}^{miss}/E_{\mathrm{T}}^{miss} < 3.0$

2) $0.25 < \Delta\phi^*_{min} < 0.5$

3) $1.25 < H_{\mathrm{T}}^{miss}/E_{\mathrm{T}}^{miss} < 3.0$, and $0.25 < \Delta\phi^*_{min} < 0.5$

The QCD involving low p_T gluon radiation in the direction of E_T^{miss} will be dominant in region 1, whereas QCD with jets in the direction of E_T^{miss} will dominate region 2. Region 3 has QCD with both jets and low p_T gluon radiation in the direction of E_T^{miss}.

From the control regions, transfer factors for the backgrounds: $t\bar{t}$, W+jets, $Z \to \nu\bar{\nu}$ and the multijet QCD background are obtained. The transfer factors are derived from the ratio between the MC simulations of the SM background in the signal region and the control region with the exception of the QCD control regions as explained earlier. These transfer factors are then applied to the data from the respective control region to estimate the yield in the signal region of the dominant SM processes of the respective control region. Prior to deriving the transfer factors, an MC estimate of the less dominant backgrounds in the given control region is subtracted from both data and the MC of the total SM background in that control region.

7.3. Corrections and Systematic Uncertainties

In addition to the statistical errors on data, there are systematic sources of uncertainty which are either of experimental or theoretical nature. Many experimental sources are associated with corrections that have been made with a finite accuracy. Others are related to essential properties with finite accuracy such as the jet energy. The theoretical errors involve (apart from finite sized MC simulations) systematic uncertainties from the MC simulations. They are also propagated to the transfer factors discussed in the previous section.

7.3.1. Theoretical Uncertainties

The QCD renormalisation factor (μ_R) and factorisation factor (μ_F) explained in section 2.1.4 are loosely determined, so the effects from varying each QCD scale factor by a factor 2 are taken as systematic uncertainties. These uncertainties apply to all the MC simulations since they originate from proton-proton collisions.

To account for uncertainties involved in determining the parton distribution functions (PDF) that play a key role in calculating production cross-sections of the particles of interest in proton-proton collisions as outlined in section 2.1.4, a set of

100 replicas are used to establish the possible variations due to alternative functions that fit the data used to establish the given PDF [7]. The PDF replicas are applied as weights on event basis as mentioned in section 6.1.1.

Cross-section predictions have a finite precision, and for $W+\text{jets}$ and $t\bar{t}$ the cross-section uncertainty becomes significant. This is simply because these backgrounds are dominant in the $\mu+\text{jets}$ control region and in most of the signal region.

NLO corrections on QCD and Electroweak processes are applied, the systematic uncertainties of these corrections are set at 100%.

The initial state radiation (ISR) in MC simulations defined in section 6.1.1 has a significant inaccuracy in simulations of $t\bar{t}$ events. This effect has been explored in a different study [8], wherefrom weights have been applied. The systematic uncertainty from this source is taken to be half the difference from unity of these weights.

7.3.2. Experimental Uncertainties

The triggers that have been applied on data have efficiencies that differ slightly from those of the MC simulations. To account for this discrepancy, scale factors have been applied with associated uncertainties originating from limited statistics of the MC simulations and systematic effects. These uncertainties in the trigger scale factors for the signal region have been accounted for accordingly.

Scale factors for the lepton selection, reconstruction and isolation have been established with a similar procedure along with the uncertainties, which are applied to the control regions. The scale factors regarding the lepton veto with their respective errors have been applied to the relevant signal and control regions.

The b-tagging efficiency and the mis-tagging rate on light flavour quarks between data and MC simulation differ across the signal region to a varying degree. These discrepancies are accounted for with scale factors derived from a different study [9]. The uncertainties on these scale factors are the sources of the b-tagging uncertainty and the mis-tagging uncertainty.

The jet energies are corrected according to MC simulations that are then verified in data in different signal regions as outlined in section 4.5.1, and as stated in [10], there are uncertainties in these corrections. To account for these uncertainties and

uncertainty in jet energy resolution, the variations based on these uncertainties are propagated from the control regions.

The estimated pileup rate is corrected by reweighting to match the expected data distribution for pileup. The total inelastic cross-section is known with an accuracy of around 5%, which can still result in either overestimating or underestimating the pileup rate to an extent that starts affecting the rate of energy contamination in jets for example. The total inelastic cross-section in the MC simulations is therefore varied by an amount of 5% to fully understand and account for the impact on the analysis caused by the limitations in predicting the pileup rate.

7.3.3. Uncertainties Estimated from Data

The following systematics are evaluated with the use of closure tests, probing n_j dependent trends as well as H_T dependence. They are given as: α_t extrapolation, $\Delta\phi*_{min}$ extrapolation, W polarisation, single isolated track veto, and the $n_b \geq 1$ event rate for the $Z \to \nu\nu$ background.

The reconstructed variables α_t and $\Delta\phi*_{min}$ rely on parameters with finite precision such as jet energy and jet position. These errors propagated to the variables need to be checked since the analysis relies heavily on these variables. The α_t and $\Delta\phi*_{min}$ extrapolations are tested individually in both the μ + jets and the $\mu\mu$ control regions by predicting the cut efficiency in data from the efficiency in MC simulations. The difference between the predicted and measured efficiency for each variable are then taken as systematic uncertainties.

A suspected deviation from the assumptions on the rate of positively charged W bosons compared to negative W bosons in MC simulations need to be checked. The resulting effect of this W polarisation needs to be accounted for. This is done in the μ + jets control region, where the events with positive muons are counted to predict the events with negative muons. Any difference between the prediction and the measurement is then considered the systematic uncertainty concerning the W polarisation.

Just as the lepton vetoes and selection efficiencies are accounted for with scale factors and systematic uncertainties. The same is done for the single isolated track veto, where a closure test is needed to establish the systematic uncertainty. The μ + jets control region is also used to test the tagging efficiency of the single isolated track veto where a prediction of the veto efficiency is made through MC

simulations and checked in the data. Any discrepancy from this prediction is then a systematic uncertainty with regards to the veto on single isolated tracks..

The $Z \to \nu\nu$ background becomes dominant in some of the b-tagged parts of the signal region. Thus, it is important to check the validity of the MC simulations of these events with respect to the fraction with at least one b-tagged jet. In the $\mu\mu$ control region, the predicted rate of $Z \to \mu\mu$ events from MC simulations in the $n_b \geq 1$ channel is compared with the observed data to find the systematic uncertainty in the b-tagged $Z \to \nu\nu$ events and to check for biases in the b-tagging rate of the MC simulations.

Since the H_{T}^{miss} distribution is used as the discriminating distribution in the analysis, it is important to have confidence in its accuracy. Therefore the uncertainties of the H_{T}^{miss} distribution shapes are established with closure tests. The shape of the H_{T}^{miss} distribution in the background is also estimated with a closure test in the control regions, first by varying the jet multiplicity n_{jet} while integrating over H_{T} and n_b, followed by varying H_{T} while integrating over n_{jet} and n_b. The H_{T}^{miss} distributions are assumed to have linear dependency and are therefore fitted to linear functions that are used to correct the shape of the H_{T}^{miss} distribution in the signal region. Any non-closure between data and simulation is then considered the uncertainty of the shape of the H_{T}^{miss} distribution. The total uncertainty ranges from around 10% to 50% depending on the jet multiplicity and H_{T}.

7.3.4. Signal Uncertainties

For the Dark Higgs signal, systematics on the jet energy corrections and the b-tagging and mis-tagging as well as the muon veto efficiency are applied as experimental uncertainties. The experimental uncertainties on the signal are treated the same way as for the background systematics, using the same corrections and uncertainties. Theoretical uncertainties are applied to the signal by scaling the observed limit with the total theoretical uncertainty. For the Dark Higgs, the parton distribution function (PDF) and the μ_F and μ_R scales have been considered for each mass point. The PDF uncertainties average around 10% whereas the scale uncertainties average about 12%. These uncertainties are based on variations on the cross-section, and checks have been made on the variations of signal yield in the analysis as well with no significant deviation from the cross-section estimate.

7.4. Bins of Particular Interest in the Signal Region

For the Dark Higgs search, the jet multiplicities; $n_{\text{jet}} = 1$, $n_{\text{jet}} = 2$ and $n_{\text{jet}} \geq 2$ asymmetric are of interest. The number of b-tagged jets being one or two depending on the particular mass parameters of the Dark Higgs samples.

A portion of the Dark Higgs events will be highly boosted, meaning that the two b-tagged jets will be so close together that they will often fit within a single AK4-jet cone. This results in the events appearing in the $n_{\text{jet}} = 1$, $n_b = 1$ channel with relatively high H_{T}. This trend is observed in fig. 7.2. One has to keep in mind that the binning in fig. 7.2 is finer than in the actual analysis. It is also noted that this channel has a relatively high background, especially in the low H_{T} region. This gives two bins of interest for the $n_{\text{jet}} = 1$, $n_b = 1$ channel, the $400\text{GeV} \leq H_{\text{T}} < 600\text{GeV}$ and $H_{\text{T}} \geq 600\text{GeV}$ bins. The $400\text{GeV} \leq H_{\text{T}} < 600\text{GeV}$ bin has an uncertainty of about 14%, whereas the $H_{\text{T}} \geq 600\text{GeV}$ bin has an uncertainty around 19%.

Apart from the b-tagged monojet category, the jet categories with $n_b = 2$ are of interest. An example is given for the 2 jet category that with $n_b = 2$ has a very high signal-background ratio in fig. 7.3a for part of the $200\text{ GeV} \leq H_{\text{T}} < 400\text{ GeV}$ bin. To demonstrate a case where the H_{T}^{miss} binning comes into play, an example from the asymmetric category with $n_b = 2$ is given in fig. 7.3b with part of the $400\text{ GeV} \leq H_{\text{T}} < 600\text{ GeV}$ bin. With the vast majority of the background having $H_{\text{T}}^{miss} < 400$ GeV and the signal being biased towards $H_{\text{T}}^{miss} \geq 400$ GeV, the H_{T}^{miss} shape greatly enhances the signal sensitivity.

7.5. Statistical Method for Analysis

A likelihood fit is applied to Asimov datasets based on the background only and signal + background hypotheses with the uncertainties listed in section 7.3 as nuisance parameters. These likelihood fits are then used for limit setting, where a strength parameter, r is determined at which the signal is excluded at 95% confidence level.

The Asymptotic method is described in detail in the article [11] and summarised below:

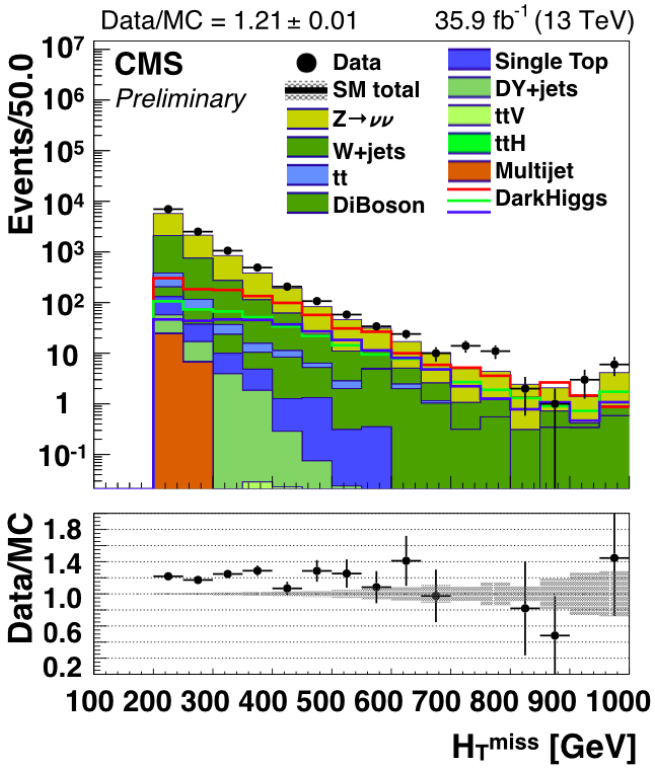
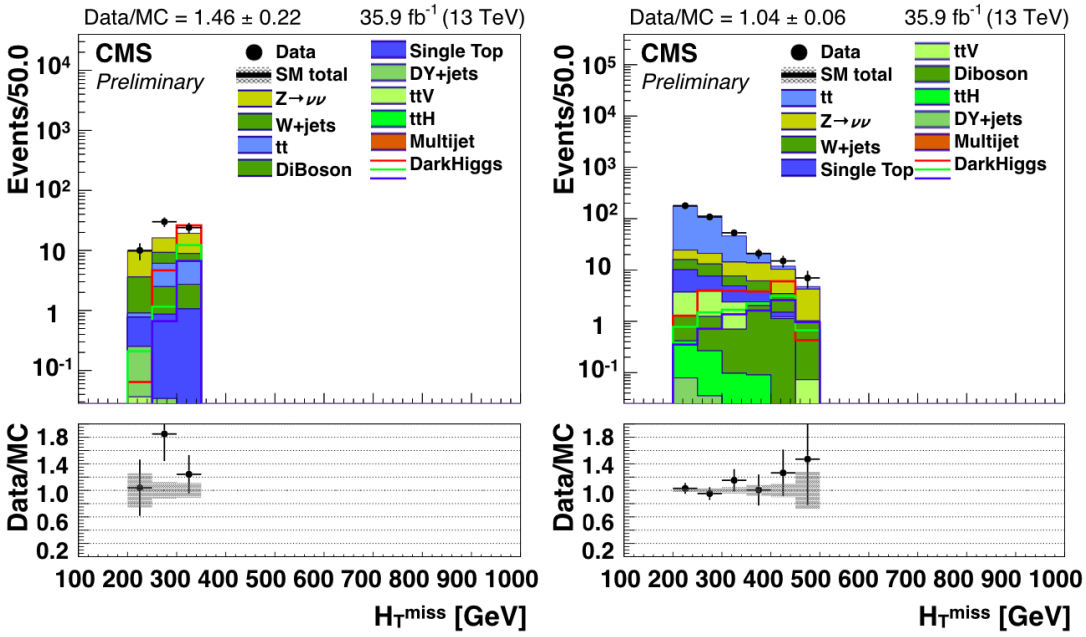


Figure 7.2.: SM background and examples of signal for the Dark Higgs with a mass of 70 GeV and masses for Z' and DM respectively being: 500 GeV and 100 Gev (red line), 1000 GeV and 200 GeV (green line) and 1500 GeV and 100 GeV (blue line). Shown in the 1 jet 1 b-tag region with statistical errors only.



(a) (b)

Figure 7.3.: SM background and examples of signal for the Dark Higgs with a mass of 70 GeV and masses for Z' and DM respectively being: 500 GeV and 100 GeV (red line), 1000 GeV and 200 GeV (green line) and 1500 GeV and 100 GeV (blue line). Shown in the 2 jet 2 b-tag region with $H_T = [300 : 350]$ GeV (a) and in the asymmetric multijet 2 b-tag region with $H_T = [400 : 500]$ GeV (b) with statistical errors only.

The likelihood functions are the product of the Poisson probability of each bin as shown in eq. 7.2, where constraints from the control regions also are included.

$$L(r, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(rs_j + b_j)^{n_j}}{n_j!} e^{-(rs_j+b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}, \tag{7.2}$$

where r is the signal strength parameter, $\boldsymbol{\theta}$ is a vector with the nuisance parameters, N is the number of bins in the signal region, n_j is the observed number of events in bin j of the signal region, M is the number of bins in the control regions, m_k is the observed number of events in bin k of the control regions, u_k is the expected number of events in bin k of the control regions. The parameters s_j and b_j are the expected values of signal and background events in bin j of the signal region, they are defined using equations 7.3 and 7.4 below:

$$s_i = s_{\text{tot}} \int_{bin\ i} f_s(x; \boldsymbol{\theta}_s)\, \mathrm{d}x, \tag{7.3}$$

$$b_i = b_{\text{tot}} \int_{bin\ i} f_b(x; \boldsymbol{\theta}_b)\, \mathrm{d}x, \tag{7.4}$$

where f_s is the probability density function for the signal, x is the variable of the given histogram which in this analysis is H_{T}^{miss}. The total expected signal event count is given as s_{tot}, whereas the shape of f_s is determined by the parameter $\boldsymbol{\theta}_s$. Equivalents apply to the background, and the total set of nuisances given as $\boldsymbol{\theta}$ in eq. 7.2 can be written as $\boldsymbol{\theta} = (b_{\text{tot}}, \boldsymbol{\theta}_b, \boldsymbol{\theta}_s)$ containing the free nuisance parameters for signal and background. The normalisation of the signal s_{tot} is fixed at the nominal value for limit setting. A new variable q is derived from the likelihoods using eq. 7.5 below:

$$q(r) = -2 \log \frac{L(r, \hat{\hat{\boldsymbol{\theta}}}(r))}{L(\hat{r}, \boldsymbol{\theta}(\hat{r}))}, \tag{7.5}$$

where $\hat{\hat{\boldsymbol{\theta}}}$ has the values of $\boldsymbol{\theta}$ that maximise $L(r)$, whereas $\hat{\boldsymbol{\theta}}$ and $\hat{r}$ are set to give the maximum likelihood, given the data. An Asimov dataset is generated to match the nominal background only hypothesis with the expected statistical fluctuations in data as the statistical error. This enables a simpler equivalent of q which can be established without the use of real data from the signal region.

$$q_A(r) = -2 \log \frac{L_A(r, \hat{\hat{\boldsymbol{\theta}}}(r))}{L_A(r', \boldsymbol{\theta})}, \tag{7.6}$$

where L_A is the likelihood function when the Asimov dataset is used in the place of real data and r' is the mean value of r in the Asimov dataset, which takes the value of zero when using the background only hypothesis for the Asimov data.

New probability density functions are then derived based on the variable q_A given the two different hypotheses, r and r'. The median of q_A, given r', is found whereafter the probability of assuming the value of q_A and higher is found, given r. This procedure is repeated for different values of r until a probability of 5% is reached. This process is repeated for one and two standard deviations, upwards and downwards from the median of q_A given r'. If the final value of r is greater than one, it means that a higher signal strength than r times the nominal can be excluded. This means that in order to exclude the nominal hypothesis at 95% confidence level, one has to reach a value for r lower than one.

7.6. Analysis Result

The procedure stated in section 7.5 is followed for a variety of signal points in the $M_{Z'}/M_\chi$ plane. The signal points used are fixed at a Dark Higgs mass of $m_s = 70$ GeV, with a DM coupling of $g_\chi = 1$ and a coupling to quarks of $g_q = 0.25$. The chosen value for the Dark Higgs has the highest expected sensitivity of the three examples shown in fig. 2.9, whereas the couplings are presented with the conventional settings [12]. The resulting limit for each signal point is given in colour scale. Interpolation is used to fill the gaps between the points, and a contour is drawn to separate the region where the limit is above one, from where it is below one. The resulting limits are given in fig. 7.4 on the interpolated colour scale for the median expected limit. For the observed limit, the value of q_A is obtained from the data for which a strength parameter for the given signal hypothesis is determined. The uncertainty band on the observed limits is based on the theoretical uncertainties on the signal that are applied by scaling the observed limits. From the observed limits it seems that there are upwards fluctuations in data corresponding to a little less than one standard deviation from the median expected limit. This result therefore gives no hints of new physics and the observed exclusion contour for the Dark Higgs hypothesis is certain.

The observed exclusion contour reaches DM masses up to around 300 GeV and Z' masses of up to 1700 GeV. This result falls short of the expected result for a dedicated search in the $E_{\mathrm{T}}^{miss} > 500$ GeV region, where the promised sensitivity reaches DM masses up to 450 GeV and Z' masses up to 2400 GeV [2]. Considering the fact that this analysis is not designed specifically with the Dark Higgs in mind, the result is actually rather impressive and the first experimental result on searching for the Dark Higgs, with all the relevant backgrounds and systematic uncertainties accounted for.

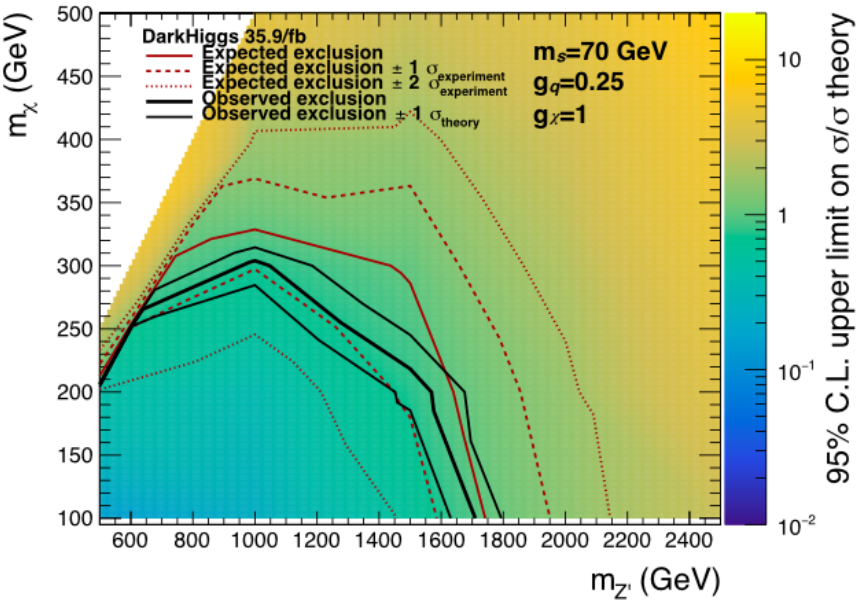
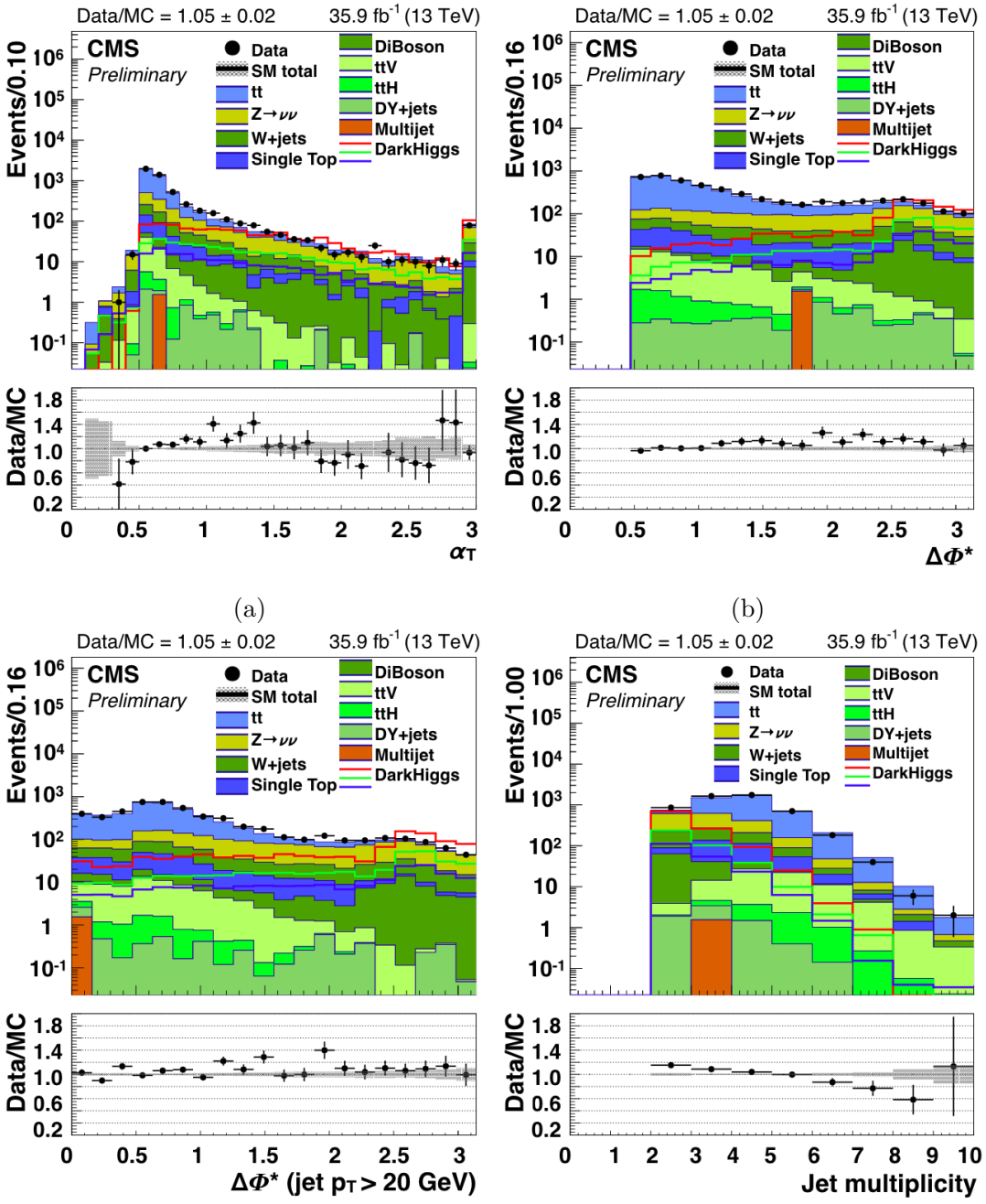


Figure 7.4.: Limit plot for the Dark Higgs with a fixed Dark Higgs mass of $m_s = 70\ GeV$ and the couplings $g_\chi = 1$ and $g_q = 0.25$. The exclusion contours are given at a 95% confidence level for the expected limit (red line) and the observed limit (black thick line). The variations of the expected limit are indicated by dashed red lines for one standard deviation and dotted red lines for two standard deviations. The standard deviation of the observed limit based on the theoretical uncertainties on signal is indicated by thin black lines. The background colours indicate the central value for the expected limit.

7.7. Perspective and Outlook

It has been demonstrated that a significant sensitivity on a new signal can be obtained from an existing analysis, which in this case contains many of the elements necessary to detect the Dark Higgs signal. The resulting limit is also the first limit set on Dark Higgs signal using experimental data.

There are several variables that are used in the analysis where a high signal-background ratio can be obtained. When looking at the variables used in the analysis, one notices that it is not only the H_{T}^{miss} variable where one can get a good signal to background ratio. This is shown for events containing two b-tagged jets for all possible jet multiplicity categories integrated over H_{T} in figure 7.5. For the α_{t} variable, the signal to background ratio is greatly enhanced at $\alpha_{\mathrm{t}} > 1.5$ as seen in figure 7.5a. A similar sensitivity is observed in the $\Delta\phi^*_{\mathrm{min}}$ distribution around $\Delta\phi^*_{\mathrm{min}} > 2.2$, see figure 7.5b. One interesting thing to note is that for a two-jet system, $\Delta\phi^*_{\mathrm{min}} > 2.2$ corresponds to $\Delta\phi_{\mathrm{jet}} < \pi - 2.2 < 1$ which is the border of an AK10-jet cone in the case where the two jets have the same rapidity. By including jets with transverse momentum down to $p_{\mathrm{T}} = 20$ GeV in the $\Delta\phi^*_{\mathrm{min}}$ evaluation, one gets a further enhancement of the signal to background ratio shown in figure 7.5c. By limiting the jet multiplicity to the two b-tagged jets also improves sensitivity to a comparable level as seen in figure 7.5d. It is clear from these figures that it is possible to modify the analysis in a way that the Dark Higgs sensitivity is enhanced. It is therefore not unrealistic to reach sensitivities close to what is predicted in [2], when also using mass reconstruction. Since the conclusion of this analysis, the CMS detector has collected significantly more data that will also help enhancing the sensitivity of a future Dark Higgs analysis.



(a) (b)

(c) (d)

Figure 7.5.: SM background and examples of signal for the Dark Higgs with a mass of 70 GeV and masses for Z' and DM respectively being: 500 GeV and 100 GeV (red line), 1000 GeV and 200 GeV (green line) and 1500 GeV and 100 GeV (blue line). Shown for events with 2 b-tagged jets in the variables: α_t (a), $\Delta\phi^*_{\text{min}}$ (b), $\Delta\phi^*_{\text{min}}$ including all jets with $p_t > 20$ GeV (c) and jet multiplicity (d).

Bibliography

[1] **CMS Collaboration**, *'Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s}$ = 13 TeV in final states with jets and missing transverse momentum'*,
JHEP 05 (2018) 025, doi:10.1007/JHEP05(2018)025, arXiv:1802.02110v2.

[2] M. Duerr et al., *'Hunting the dark Higgs'*,
JHEP 1704 (2017) 143, doi:10.1007/JHEP04(2017)143, arXiv:1701.08780v2.

[3] **CMS Collaboration**, *'Search for supersymmetry in pp collisions at 7 TeV in events with jets and missing transverse energy'*,
Phys. Lett. B 698 (2011) 196, doi:10.1016/j.physletb.2011.03.021, arXiv:1101.1628.

[4] **CMS Collaboration**, *'SUSY searches with dijet events'*,
CMS PAS SUS-08-005, http://cds.cern.ch/record/1149915/files/SUS-08-005-pas.pdf.

[5] **CMS Collaboration**, *'A search for new phenomena in pp collisions at $\sqrt{s}$ = 13 TeV in final states with missing transverse momentum and at least one jet using the α_T variable'*,
Eur. Phys. J. C 77 (2017) 294, doi:10.1140/epjc/s10052-017-4787-8, arXiv:1611.00338.

[6] M. Barrett, *'Background estimation strategies in CMS'*,
TOP 2010 3rd International Workshop on Top Quark Physics, https://agenda.irmp.ucl.ac.be/event/538/contributions/2924/attachments/...
...1535/1748/barrett_Top2010.pdf.

[7] J. Butterworth et al., *'PDF4LHC recommendations for LHC Run II'*,
J. Phys. G 43 (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.

[8] **CMS Collaboration**, *'Search for top-squark pair production in the single-lepton final state in pp collisions at* $\sqrt{s} = 8\text{TeV}$*'*,
Eur. Phys. J. C 73 (2013) 2677, doi:10.1140/epjc/s10052-013-2677-2, arXiv:1308.1586.

[9] **CMS Collaboration**, *'Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV'*,
JINST 13 (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[10] **CMS Collaboration**, *'Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV'*,
JINST 12 (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[11] G. Cowan et al., *'Asymptotic formulae for likelihood-based tests of new physics'*,
Eur. Phys. J. C 71 (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[12] D. Abercrombie et al., *'Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum'*,
Phys. Dark Univ. 26 (2019) 100371, doi:10.1016/j.dark.2019.100371, arXiv:1507.00966.

8. Search for a Heavy Higgs in $t\bar{t}$ Pair Production using 2017 Data

The analysis, being a search for anomalies in the $t\bar{t}$ distributions with focus on probing hypotheses beyond the Standard Model (BSM) involving an additional (pseudo)scalar boson with a mass greater than $2m_t$. For this purpose, two variables are selected for analysis, the $m_{t\bar{t}}$ and c_{hel} variables. The $m_{t\bar{t}}$ distribution is an obvious choice since the signal will manifest itself as a resonance with interference in this distribution. The other variable c_{hel} has different slopes for $t\bar{t}$ events originating from pseudoscalars, scalars and purely from SM processes as demonstrated in fig. 8.1. This is due to different spin correlations of the top quarks in the events. The c_{hel} variable is defined as the cosine of the angle between the two leptons in the rest frame of their respective parent (anti)top quark, reached by a rotation free boost from the $t\bar{t}$ rest frame [1]. The charged leptons from top (anti-top) decays are almost 100% efficient analysers of the top polarisation [1]. Therefore, $t\bar{t}$ spin correlation can be measured most effectively using dileptonic $t\bar{t}$ events, for example by utilising the c_{hel} variable which is the most sensitive variable compared to known alternative spin correlation variables. The addition of the c_{hel} variable in the analysis both enhances the signal sensitivity and increases the ability to distinguish scalar signal from a pseudoscalar signal. To make effective use of the $m_{t\bar{t}}$ variable, it is necessary to obtain a high concentration of $t\bar{t}$ events with precise modelling of the expected yield and distribution. The c_{hel} variable relies on each top quark in the $t\bar{t}$ system decaying leptonically, therefore the dileptonic $t\bar{t}$ events are of interest in this analysis. Several methods and checks have been applied to ensure confidence in the precision of the predicted analysis distributions. A large set of systematic sources have been considered to ensure that their effects on the analysis are taken into account. Smoothing has been applied in some cases where statistical fluctuations overshadow the actual effects from variation of a given systematic source. A control region has been used to constrain the $V+\text{jets}$ background yield. Control distributions have been checked for inconsistencies between data and MC simulated predictions. With these measures in place, precise and reliable estimates of the analysis distributions can be made,

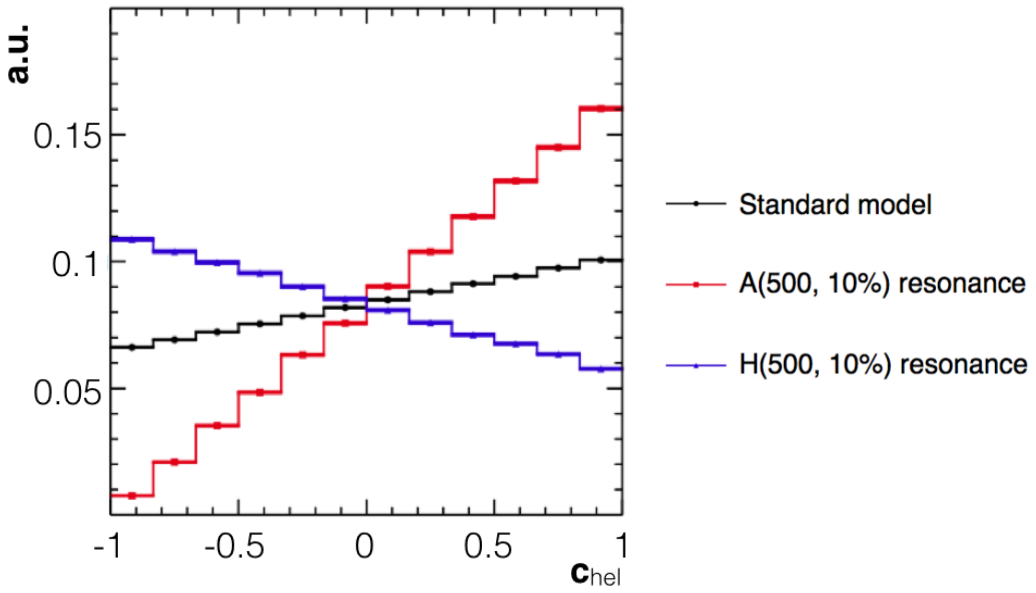


Figure 8.1.: c_{hel} distributions for $t\bar{t}$ events at generator level originating from a heavy scalar (blue), pseudoscalar (red) and from SM processes (black). Figure taken from source [2].

so that a search for the signal hypotheses can take place. To minimise the effect of negative components of the interference cancelling out the positive resonance in the same bin, a fine binning for the $m_{t\bar{t}}$ distribution has been selected. The statistics in each bin need to be sufficiently large to ensure that the sensitivity is not compromised by significant statistical uncertainties from either data or MC simulations. The binning in the c_{hel} distribution therefore only contains 5 bins to keep a decent resolution on $m_{t\bar{t}}$ without reducing the statistics in each bin too severely.

Between 2016 and 2017, there have been many changes that are relevant to the analysis. A new tracker has been installed in the CMS detector and the instantaneous luminosity has increased significantly. This means that the use of additional tools for handling pileup related effects has become necessary. Additional corrections and their corresponding systematic sources have also been recommended for use in analysis, such as the pre-firing of the level 1 trigger.

8.1. Kinematic Reconstruction

To evaluate variables like $m_{t\bar{t}}$ and c_{hel}, it is necessary to do kinematic reconstruction on data since the detected objects are decay products of the decayed top quarks. The kinematic reconstruction uses an algorithm stated by L. Sonnenschein [3]. It assumes a known top mass and W mass, sets the lepton and neutrino masses to zero and sets the mass of the b-quark to 4.8 GeV. The unknown variables to be solved are the components of the momentum for the neutrino and antineutrino in the $t\bar{t}$ system. To solve these variables, it is assumed that all the missing transverse energy is attributed to the neutrino-antineutrino pair. With the constraints obtained from the assumptions on the $t\bar{t}$ system, the unknown variables can be reduced to a fourth order polynomial of one variable, which has up to four solutions. The assignment of jets to the two b-quarks of the $t\bar{t}$ system introduces another element of ambiguity that increases the number of possible solutions with a factor $N_j!$ where N_j is the jet multiplicity.

To prevent reconstruction failure due to errors in measurement of the jet and lepton momenta, these are smeared in terms of magnitude and direction within the accuracies of the measurements with 100 variations. In addition to increasing the likelihood of generating a solution, the smearing is also used to determine which jets go with the two leptons. This is done by generating weights according to equation 8.1. For events where two or more jets are b-tagged, the combinations with only one b-tagged jet are discarded prior to computing their weight.

$$\omega = \sum_{i=1}^{100} \omega^i_{l\bar{b}} \omega^i_{\bar{l}b}, \tag{8.1}$$

where ω is the weight, $\omega^i_{l\bar{b}}$ is the probability for the b candidate to originate from a top when deriving the reconstructed mass of the $\bar{l}b$ system given the variation indexed with i. The same procedure is followed for the antiparticle candidates to find $\omega^i_{\bar{l}b}$.

The assignment of jets to b and $\bar{b}$ with the highest weight is selected for the remaining stages of the reconstruction. To find the best solution for the neutrino momenta, the reconstructed mass of the $t\bar{t}$ system is evaluated, and the solution that minimises $m_{t\bar{t}}$ is selected. This solution is true in 60% of the cases and is more accurate than taking an average of the solutions that are likely to be true, even when this average is weighted [4]. An average of the p_{T} of the $t\bar{t}$ system is taken as a final step in the reconstruction to end up with one solution for the $t\bar{t}$ candidate.

8.2. Event Selection and Cuts

Dileptonic $t\bar{t}$ events are when both the top and antitop quark decay leptonically in the process: $t \to b + W^+, W^+ \to l^+ + \nu_l$, where the process for the antitop is charge conjugated and l represents the lepton of any flavour (e, μ, τ). There are three dileptonic channels: the dielectron channel (ee), the dimuon channel $(\mu\mu)$ and the electron-muon channel $(e\mu)$. Since tauons have a limited lifetime and only travel a short distance before decaying, the tauon decay products determine which channel the event belongs to. If a tauon decays hadronically, the event is less likely to be accepted and is considered a $t\bar{t}$ background event. To obtain a high concentration of dileptonic $t\bar{t}$ events, we have a series of cuts and selection criteria as stated in table 8.1. This cut sequence is applied to events that pass at least one of the triggers for either muons or electrons. This can either be one of the dilepton triggers or one of the single lepton triggers. The jets are AK4 with jet ID tight lepton veto and PU jet ID tight as outlined in section 4.5.2. The b-tagging is done with deepCSV at the medium working point, explained in section 4.5.3. The electron selection uses a working point based on multivariate analysis which accepts around 90% of the electrons, outlined in section 4.4. It is a fairly loose working point, increasing the yield of $t\bar{t}$ events with a manageable increase in other background processes. The muon ID is evaluated at the medium working point with isolation at the loose working point, which are outlined in section 4.3.

Object	Selection Requirement
Jets	$p_{\mathrm{T}} > 20$ GeV $\lvert\eta\rvert < 2.4$
Electrons	$p_{\mathrm{T}} > 20$ GeV $\lvert\eta\rvert < 2.4$ Id: MVA90
Muons	$p_{\mathrm{T}} > 20$ GeV $\lvert\eta\rvert < 2.4$ Cutbased Id: M Iso L
Cut Parameter	**Cut Value**
Lepton Quantity	$n_l = 2$, $p_{\mathrm{T}}^{l1} > 25$ GeV
Dilepton Mass	$m_{ll} > 20$GeV
Remove Z-window	106GeV $< m_{ll} <$ 76GeV, only for ee and $\mu\mu$ events
Jet Multiplicity	$n_j \geq 2$, $p_{\mathrm{T}}^{j1} > 30$ GeV, $p_{\mathrm{T}}^{j2} > 30$ GeV
Missing Transverse Energy	$E_{\mathrm{T}}^{miss} > 40$ GeV, only for ee and $\mu\mu$ events
B-tagging	$n_b \geq 1$
Kinematic Reconstruction	Has at least one real solution

Table 8.1.: Object selections and cuts used in the analysis.

Events without any solution to the kinematic reconstruction are discarded as a final selection cut.

For evaluating the effectiveness of the selection cuts and for estimating the amount of residual background processes, these SM processes are divided into the following categories: diboson (VV), V+jets, single top (t) and $t\bar{t}$+V. The processes included in each category are given in table 8.2, along with the generators used to produce the processes at ME level.
The showering is done with Pythia 8 using the CP5 tune. All background samples have been matched using the MLM matching described in section 6.1.1, except for the $t\bar{t} + W$ samples where the FXFX method is used instead [5].

Category	Processes	ME Generator	Cross-section [pb]
$t\bar{t}$	$t\bar{t}$ dilepton	Powheg	381.76
	$t\bar{t}$ lepton+jets		
$t\bar{t} + V$	$t\bar{t} + W$	Madgraph	0.6465
	$t\bar{t} + Z$		0.7536
Single Top	t t-channel	Powheg	136.02
	$\bar{t}$ t-channel		80.95
	t tw-channel		71.7
	$\bar{t}$ tw-channel		
	t s-channel		6.35
V + Jets	Drell-Yan, $10\,\text{GeV} < m_{ll} < 50\,\text{GeV}$	Madgraph	18610
	Drell-Yan, $m_{ll} > 50\,\text{GeV}$		6225.42
	W + Jets		61526.7
Diboson	WW	Pythia 8	75.8
	WZ		27.6
	ZZ		12.14

Table 8.2.: List of MC samples used to estimate the SM background.

A control region is used for the purpose of rescaling the V + jets events with the same cut sequence as in table 8.1 up to and including the $m_{ll} > 20$ GeV cut, from there, two more cuts are applied: 76 GeV $< m_{ll} <$ 106 GeV and $n_j \geq 2$, where $p_{\text{T}}^{j1} > 30$ GeV and $p_{\text{T}}^{j2} > 30$ GeV.

8.3. Corrections and Scale Factors

Scale factors are applied to compensate for differences in efficiency between data and MC simulations. Among these are the trigger scale factors, which have been derived by comparing trigger efficiencies on MC simulations and data that pass the

triggers for events with large amounts of missing transverse energy (MET trigger), following the procedure given in [6]. For an improved resolution of the trigger scale factors, the scale factors are determined as a function of p_{T} rather than η. Uncertainties originating from limited statistics of the MC simulations and possible correlations between the MET trigger and the lepton triggers are included in the systematic error of the triggers. These are applied to adjust the expected trigger efficiency in the ee and $e\mu$ channels. When applying the trigger scale factor in the $\mu\mu$ channel, the overall consistency between data and MC simulation worsens. This defeats the purpose of the trigger scale factor in the $\mu\mu$ channel, wherein the trigger scale factor is omitted as a consequence. Lepton scale factors have been applied to match the selected working points for isolation and identification of each lepton, applied in all channels. Scale factors for b-tagging are also applied, including the scale factor for the mis-tagging of light flavoured jets.

Corrections to account for the muon pre-firing of the level one trigger have been applied as well. These pre-firing corrections are at the percent level and have been included to more accurately model the $\mu\mu$ channel.

A data-driven correction is applied on the simulated $V+$jets background from the above mentioned control region.

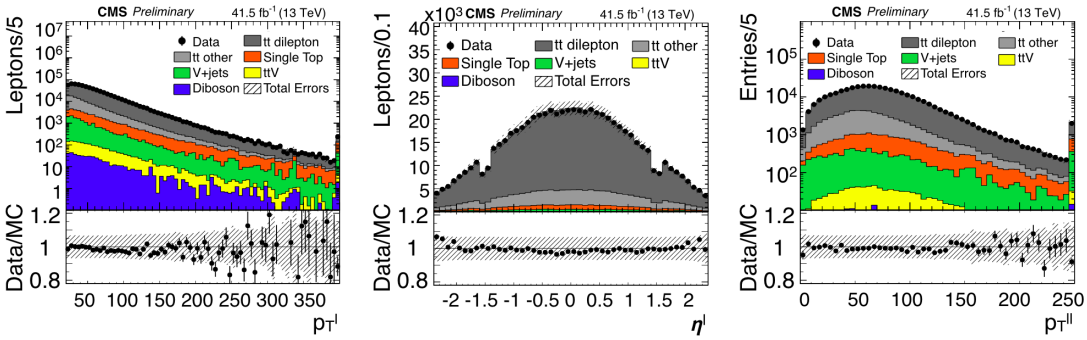
Jet energy corrections and jet energy resolution following the procedure outlined in section 4.5.1 as established for 2017 data are applied on data and MC simulations.

To ensure that the corrections and scale factors are applied correctly, and for reference for similar searches on 2017 data, control distributions are given for each channel as well as for the combined channel. The control plots shown in figures 8.2, 8.3, 8.4 and 8.5 include all the cuts except for the requirement of a solution in the kinematic reconstruction, similar control distributions at preceding steps of the cut sequence are found in appendix A. A slight downwards slope in the data/MC ratio is observed in the lepton p_{T} distribution up to $p_{\mathrm{T}}^{l} = 150$ GeV whereafter the trend is blurred by statistical fluctuations (fig. 8.2a). This trend is visible in the $e\mu$ and $\mu\mu$ channel (figg. 8.4a, 8.5a), whereas the data/MC distribution for the lepton p_{T} distribution is flat in the ee channel (fig. 8.3a). Fluctuations in the data/MC distributions for rapidity and pseudorapidity occur consistently over the whole range of η (or y for rapidity), see figg. 8.2b, 8.2e and 8.2g. These fluctuations cannot be explained by statistics only, but they lie well within the systematic uncertainties.
There is a distinct upwards slope in the data/MC distribution for p_{T}^{j}, which is

present in all channels (figg. 8.2f, 8.3f, 8.4f, 8.5f). There is also an upwards slope in the missing E_{T} data/MC distribution (fig. 8.2j), mainly present in the $e\mu$ channel (fig. 8.5j). These slopes lie within the systematic uncertainties and are most likely connected to the top p_{T} discrepancy between data and MC simulation, which has been observed since 2015 [8].
The missing E_{T} has not been aligned, and therefore gives rise to a wave-like distribution for the azimuthal angle of the missing E_{T} in fig. 8.2k. A similar shape is observed in the data/MC distribution, which is generally within the systematic uncertainties. The event yield at $N^{jet} \geq 6$ is significantly underestimated, but still within errors for the combined channel (fig. 8.2h). The b-tagged jet multiplicity is only correctly estimated at $N^{bjets} \leq 2$ (fig. 8.2i). Since the analysis does not give special treatment to events with high b-tagged jet multiplicities, that total around 5% of the events, this error is considered acceptable.
Poor data/MC agreement is observed in the vertex multiplicity in figures 8.2l, 8.3l, 8.5l and 8.4l. This is due to an unsuccessful attempt to implement an improved approach to pileup reweighting, since the existing approach to pileup reweighting also gives poor data/MC agreement. Unfortunately, the pileup systematic variations rely on reweighting as well and therefore simply take the nominal value. The discrepancies between data and MC is roughly covered by other systematics though, and the tight PU jet ID working point minimises the impact of a wrong estimate of pileup.

The signal for this analysis consists of dileptonic $gg \to \Phi \to t\bar{t}$ processes, where Φ can be either a scalar or pseudoscalar particle. Both the resonance and the interference with the SM $t\bar{t}$ production are accounted for in separate samples. Three free parameters are varied to create a search grid. These are the charge-parity (CP), the mass and the width. The full list of signal points with the selected parameters is given in table 8.3.

Correction factors on the simulated signal and background are applied to go from leading order (LO) to next to next to leading order (NNLO), they are called K-factors. For the SM MC samples, the NNLO cross-sections are already available, but the K-factors on signal have been derived privately, using Sushi-1.6.1 [7]. The NNLO cross-section is calculated using the type 2 Yukawa coupling, NNPDF31_nnlo_as_0118_nf_4 and input parameters in what is called the physical basis, consisting of the Higgs masses, an off-diagonal term called m_{12}, the four Higgs coupling terms λ_6 and λ_7 that are set to zero and $\sin(\beta - \alpha)$. The width used in the signal samples is calculated along with the cross-sections. The key input parameter for obtaining the correct width is called $\tan\beta$, while the rest of the input parameters are adjusted to fit with the Higgs masses according to the



(a) Transverse momentum of leptons

(b) Pseudorapidity of leptons

(c) Transverse momentum of dileptons

(d) Mass of dileptons

(e) Rapidity of dileptons

(f) Transverse momentum of jets

(g) Pseudorapidity of jets

(h) Jet multiplicity

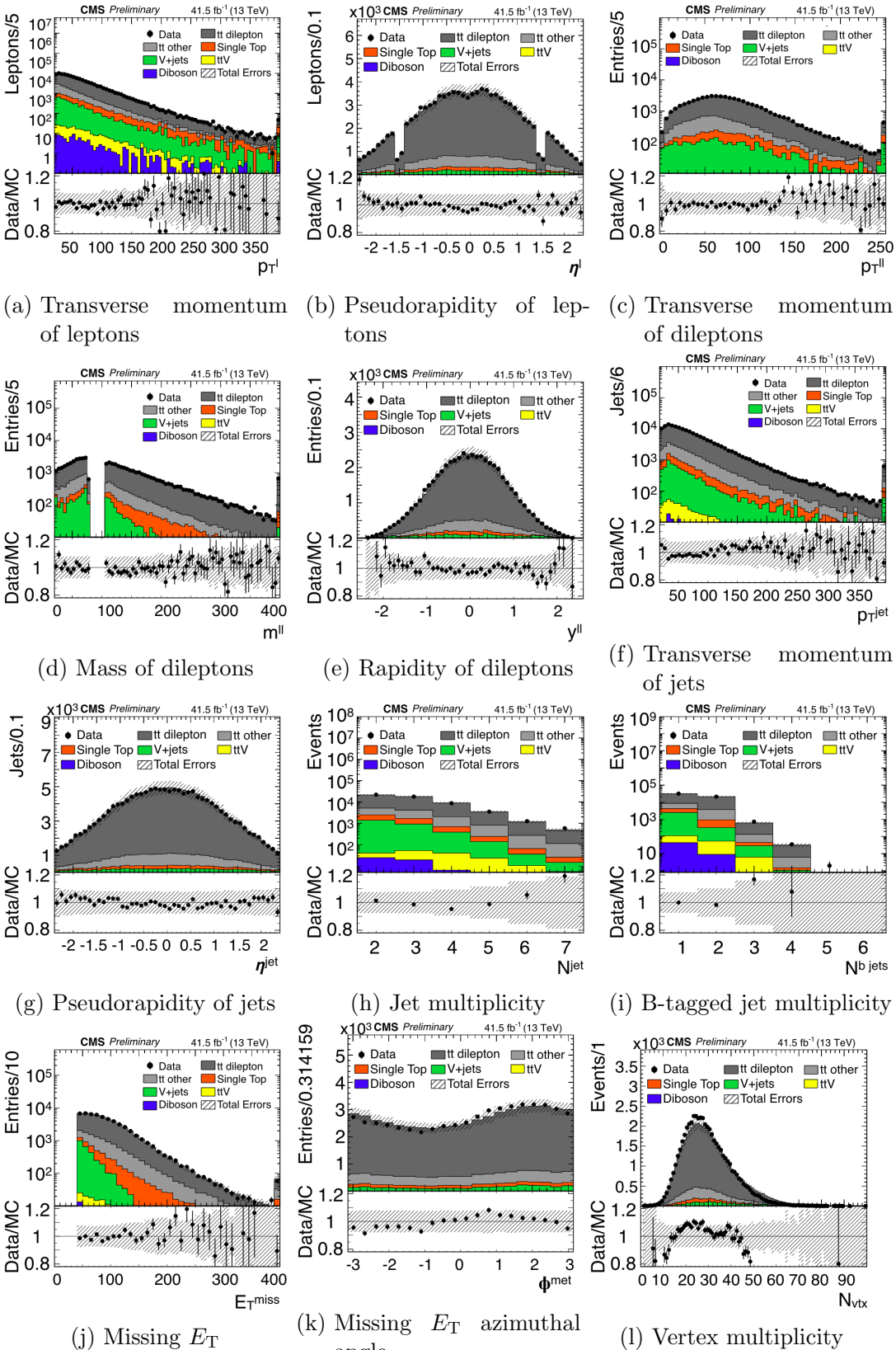
(i) B-tagged jet multiplicity

(j) Missing E_T

(k) Missing E_T azimuthal angle

(l) Vertex multiplicity

Figure 8.2.: Control distributions for the combined channel with all cuts prior to Kin-reco applied



(a) Transverse momentum of leptons

(b) Pseudorapidity of leptons

(c) Transverse momentum of dileptons

(d) Mass of dileptons

(e) Rapidity of dileptons

(f) Transverse momentum of jets

(g) Pseudorapidity of jets

(h) Jet multiplicity

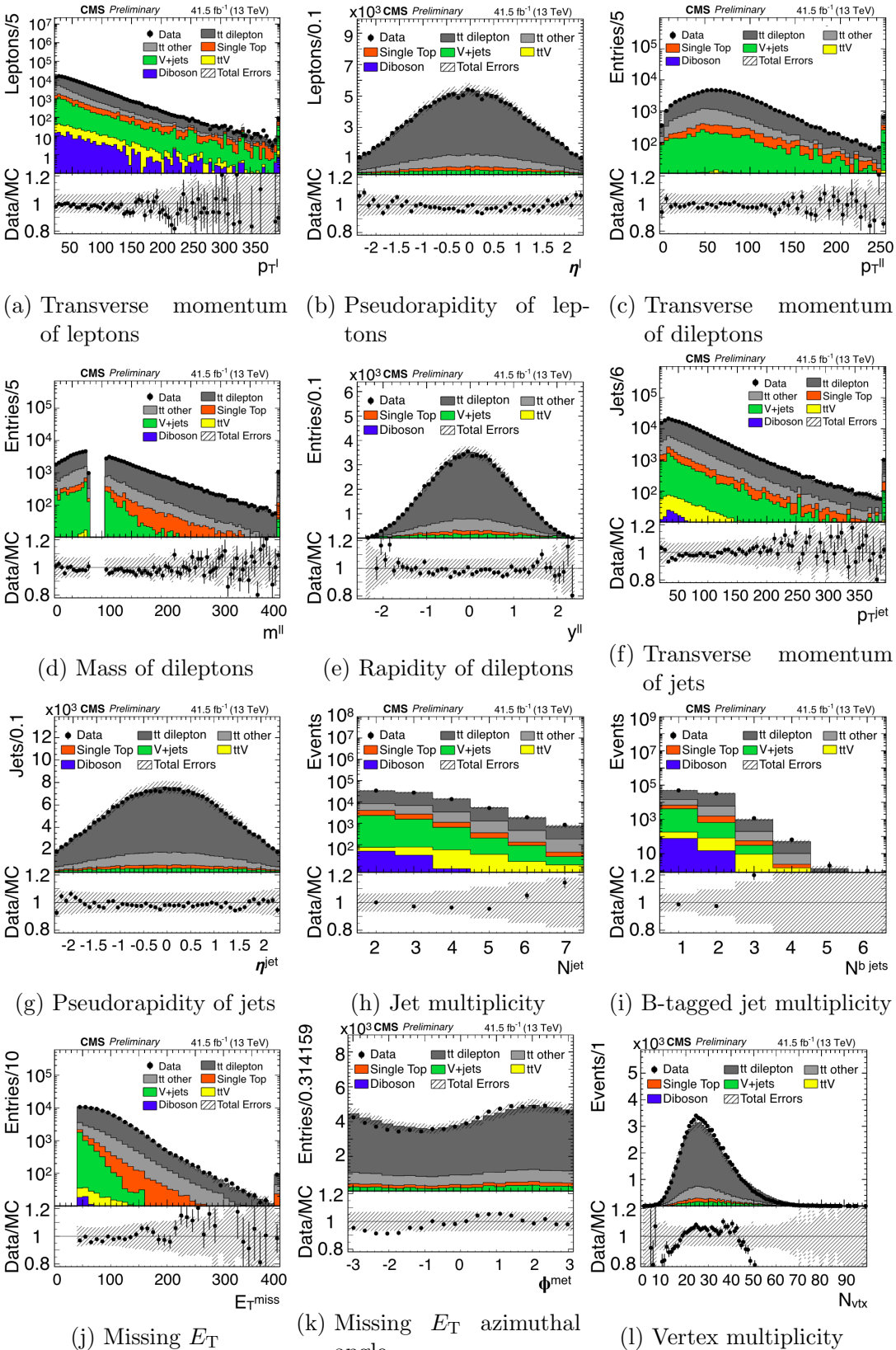
(i) B-tagged jet multiplicity

(j) Missing E_T

(k) Missing E_T azimuthal angle

(l) Vertex multiplicity

Figure 8.3.: Control distributions for the *ee* channel with all cuts prior to Kin-reco applied



(a) Transverse momentum of leptons

(b) Pseudorapidity of leptons

(c) Transverse momentum of dileptons

(d) Mass of dileptons

(e) Rapidity of dileptons

(f) Transverse momentum of jets

(g) Pseudorapidity of jets

(h) Jet multiplicity

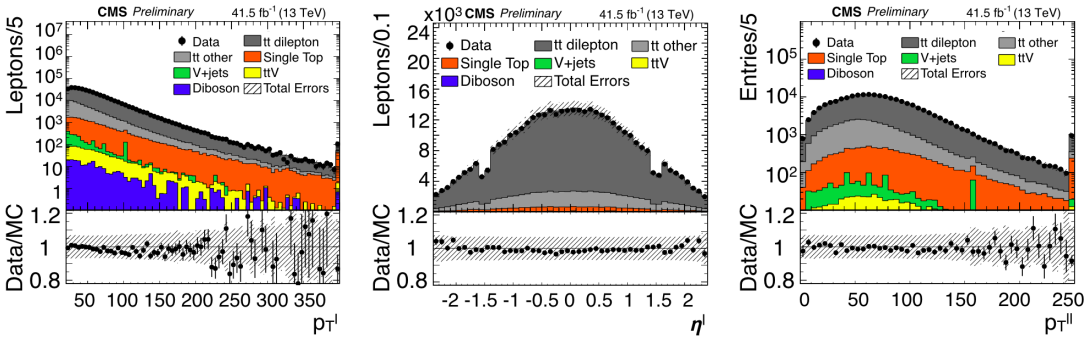
(i) B-tagged jet multiplicity

(j) Missing E_T

(k) Missing E_T azimuthal angle

(l) Vertex multiplicity

Figure 8.4.: Control distributions for the $\mu\mu$ channel with all cuts prior to Kin-reco applied



(a) Transverse momentum of leptons

(b) Pseudorapidity of leptons

(c) Transverse momentum of dileptons

(d) Mass of dileptons

(e) Rapidity of dileptons

(f) Transverse momentum of jets

(g) Pseudorapidity of jets

(h) Jet multiplicity

(i) B-tagged jet multiplicity

(j) Missing E_T

(k) Missing E_T azimuthal angle

(l) Vertex multiplicity

Figure 8.5.: Control distributions for the $e\mu$ channel with all cuts prior to Kin-reco applied

CP	Mass [GeV]	Relative Width [%]	K-factor R	K-factor I
A	365	1.9122	2.50076	2.0694
A	400	2.9640	2.37786	2.0179
A	600	4.7919	2.18698	1.9352
A	800	5.2844	2.12482	1.9075
A	1000	5.4974	2.08649	1.8902
H	365	0.2038	2.52968	2.0813
H	400	0.7591	2.39676	2.0259
H	600	3.2076	2.20193	1.9418
H	800	4.3016	2.12818	1.9090
H	1000	4.8431	2.08801	1.8909

Table 8.3.: List of signal points used in the analysis along with the derived K-factors.

hMSSM using 2HDMC-1.7.0 [9]. As the LO basis for the K-factors, the cross-section for the $gg \to \Phi$ process with the same widths and top-couplings, g_t as in the NNLO result are derived using Madgraph5aMC@NLO. Cross checks have been made by varying the PDFs used, as well as the Yukawa coupling type. These changes give a variation of around 1% each. A discrepancy between the $\Phi \to t\bar{t}$ branching ratio derived internally in Madgraph5aMC@NLO and the calculated branching ratio using equation 6.1 from chapter 6 has been observed to be up to 10% for pseudoscalar samples and 3% for scalar samples. For the pseudoscalar samples, there is a clear proportionality with the Heavy Higgs mass. This discrepancy is probably just due to the omission of terms involving the running coupling (α_s) in equation 6.1.

The derived K-factors apply to the resonance part of the signal only, so for the signal interference, the geometrical mean (square root of the product) of the signal resonance K-factor and the SM LO to NNLO K-factor is used. To keep the PDF basis for the signal corrections consistent, the SM $t\bar{t}$ cross-section is derived at LO and NNLO with top++ [10]. The K-factor for the SM top production is found to be 1.71.

8.4. Systematics

Uncertainties are evaluated for the experimental corrections and efficiencies along with theoretical estimates. Most of the systematic variations have an impact on the

shape of the distributions of background and signal where applicable in addition to the normalisation. The systematic uncertainties are listed in table 8.5 with the treatment of the given systematic in a likelihood fit for limit setting.

Muon and Electron Identification and Isolation Scale Factors

The scale factors, correcting the mismatch between data and MC simulations with regards to identification and isolation efficiencies for muons and electrons, are established with a finite precision. The scale factors relating to muons are treated separately from the electron scale factors. Variations in these scale factors are then applied according to their precision to determine their effects on the event rate and shape of the analysis distributions.

Jet Energy Resolution (JER)

To match the jet energy resolution in data, the jet p_{T} resolution in MC simulations is smeared as outlined in section 4.5.1. Among the sources of error in the JER are the uncertainty of the remaining jet energy corrections, residual ISR and FSR radiation and the fact that the assumption that the JER is Gaussian does not hold in the tails of the distributions for high p_{T} jets. The scale factor on the JER is varied with the estimate of one standard deviation in the MC simulations, which are impacted in both the shape of the distributions of selected events and the rate of selected events.

Jet Energy Scale (JES)

he energy measurement of jets is calibrated by applying jet energy scale correction factors which are connected to three categories of sources of uncertainty, pileup offset, relative data/MC corrections and absolute data/MC corrections. The procedure for jet energy corrections is described in section 4.5.1., which is essential to review to understand the systematic uncertainties arising from said procedure. The error from some of the sources vary significantly over the pseudorapidity of the jets and are therefore divided in bins of $|\eta|$ called barrel (BB) in the range $|\eta| < 1.3$, inner part of the endcap (EC1) in the range $1.3 \leq |\eta| < 2.5$, forward part of the endcap (EC2) in the range $2.5 \leq |\eta| < 3.0$ and the forward calorimeter

(HF) in the range $3.0 \leq |\eta| < 5.2$. The full list of systematic sources on jet energy scale is given in table 8.4. Each of these sources affect the event rate and shape of the analysis distribution.

Missing E_{T} Unclustering

The uncertainty on the missing E_{T} due to energy deposited outside the range of the reconstructed energy clusters is evaluated as the p_{T} resolution for the given PF candidates contributing to this unclustered energy. This resolution depends strongly on the flavour of the PF candidates contributing to the unclustered energy. A more detailed description is given in [13]. This systematic affects both the event rate and the shape of the analysis distribution.

Renormalisation and Factorisation Scale

The QCD renormalisation scale (μ_R), explained in section 2.1.4, has the default value of 1 in units of mass at the scale of the heaviest particles in the given process, which is set to m_t in $t\bar{t}$ events and $M_\phi/2$ in Signal events. It is varied down to 0.5 and up to 2 for the hard process. The QCD factorisation scale (μ_F), explained in section 2.1.4, is set and varied in the same manner as μ_R for the hard interaction [14].

L1 Trigger Pre-firing

Due to the narrow time interval between bunch crossings (25 ns), there is a slight possibility for a muon from a bunch crossing, when activating the muon stations, to activate the L1 muon trigger for the next bunch crossing. This effect is called pre-firing of the L1 trigger. The uncertainty associated with the corrections for this pre-firing of the level 1 trigger is taken into account, contributing to both the rate and the shape uncertainty.

Source	Description
Pileup Offset	
PileUpDataMC	Data vs. MC simulation offset with the random cone (RC) method
PileUpPtRef	True offset vs. RC $\bigotimes$ absolute p_{T}
PileUpPt(BB, EC1, EC2, HF)	True offset vs. RC $\bigotimes$ relative η
Relative Data/MC	
RelativeJER (EC1, EC2, HF)	Jet p_{T} resolution
RelativeFSR	ISR + FSR correction
RelativeStatFSR	ISR + FSR statistical uncertainty
RelativeStat(EC2, HF)	Statistical uncertainty
RelativePt(BB, EC1, EC2, HF)	Log-linear vs. flat fit $\times$ 50%
RelativeBal	Difference in calibration fits between MPF and jet p_{T} balance methods
TimeEta	Relative η time dependence
Absolute Data/MC	
AbsoluteScale	Lepton scale, $\pm$ 0.11%
AbsoluteMPFBias	MPF bias, $\pm$ 0.28% (from ν's $\bigoplus$ ISR acceptance, 0.2% $\bigoplus$ 0.2%)
AbsoluteStat	Statisitical uncertainty vs. p_{T}
SinglePionECAL	Single pion response in ECAL, $\pm$ 4.2%
SinglePionHCAL	Single pion response in ECAL, $\pm$ 1.5%
Fragmentation	Jet fragmentation in Pythia 6.4 vs. Herwig++2.3
TimePt	Absolute p_{T} time dependence
Flavour	
FlavorQCD	Uncertainties on flavour dependent jet responses in studied dijet events

Table 8.4.: Jet energy scale uncertainty sources as stated in [11] or [12].

Trigger Scale Factor (SF)

The trigger scale factors have uncertainties that originate from statistical limitations and possible correlations with the MET-trigger, are propagated to the trigger systematic. This systematic contributes to both the event rate and the shape of the analysis distribution.

Pileup

Since pileup reweighting has not been implemented, the pileup variations assume the nominal value. The impact from pileup variations in the analysis is not expected to be one of the major systematic sources though. With pileup reweighting, the total inelastic cross-section is varied by 4.6% from the nominal value.

B-tagging

Variations in establishing the scale factors for b-tagging efficiency also contribute to the systematic uncertainties, which have been accounted for. Similarly, the b-tagging rate of light flavoured jets (LF b-tagging) is corrected with the establishment of scale factors with errors as a separate systematic source. Both of which influence both the event rate and the shape of the analysis distribution.

Parton Distribution Function

To estimate the uncertainty of the PDF function, a set of 100 variations has been used. These variations are evaluated for the SM $t\bar{t}$ events, which are evaluated in each bin of the analysis region to determine the discrepancy from the nominal count of $t\bar{t}$ events. The variations which have the 16th highest and 16th lowest event count are then taken as the upward and downward variation respectively. Thereby the envelope of the PDF set is taken for each bin in the analysis. The PDF envelope affects both the event rate and the shape of the analysis distribution.

Reweighting of the Top Quark Transverse Momentum

The top p_{T} reweighting parameters come from an exponential fit of the slope of the p_{T} distribution for the top quark, appearing in the data/MC ratio [8]. Since the top p_{T} reweighting has not been applied, the fitting parameters are applied as systematic up and down variations instead. This systematic is set to only affect the shape of the analysis distribution.

Top Quark Mass

The top mass is varied between 169.5 GeV and 175.5 GeV for the top mass systematic. This is done with independent $t\bar{t}$ samples and affect both the shape of the analysis distribution and the event rate.

Damping Parameter in Matching (Hdamp)

In the matching of the ME and PS states, there is need of a damping parameter h_{damp}. This parameter has a nominal value of 1.379 m_t and is varied between 0.874 m_t and 2.31 m_t in dedicated samples. This follows the recommendation for 2017 data that follows the procedure stated in [15]. This systematic affects the event rate and the shape of the distribution in analysis.

Parton Shower Variations (ISR and FSR)

To account for the uncertainties of the parton showers at the ISR and FSR stages, explained in section 6.1.1, the μ_R parameter is varied. These variations are given as weights in 2017 MC samples that involve top quarks. In both cases the μ_R parameter is varied with a factor 0.5 and 2 of the nominal renormalisation scale of the parton shower.

PDF α_s

Variations in α_s are applied between $\alpha_s = 0.117$ and $\alpha_s = 0.119$ using the PDFs: NNPDF30_nlo_as_0117 and NNPDF30_nlo_as_0119.

Underlying Event

Variations in the CP5 tune for simulating the underlying events are applied in dedicated samples with variations of parameters in Pythia 8 on parton interactions stated in [16].

Cross-sections, Luminosity and DY Normalisation

Uncertainties for the integrated luminosity, cross-sections for SM MC categories and the uncertainty of the scaling of the event rate in the V + jets category are accounted for as normalisation scales (norm in table 8.5) to the distributions, since they have no impact on the shapes of the distributions themselves.

The systematic uncertainties along with the statistical uncertainty for each bin are treated as nuisance parameters in the likelihood fits. To probe their impact on the resulting limit and to check that the pull is centred at zero and ideally that no constraints on the systematics arise from fitting the likelihood functions.
These impacts and pulls for the systematics are given for a benchmark sample with $M_\Phi = 500$ GeV, CP-odd (pseudoscalar) and relative width of 2.5% in figures 8.6, 8.7, 8.8, 8.9, 8.10 and 8.11. These plots represent the situation on the systematics after smoothing has been applied to mitigate constraints arising from the likelihood fits.

8.4.1. Smoothing of Systematics

To address the fact that some of the systematics are over-constrained in the likelihood fit, due to statistical fluctuations on top of the systematic variations, these systematics are smoothed for both signal and background for the relevant systematics. These are : MET unclustering, JER and JES. The JES is divided into subcategories to facilitate a more accurate estimate. Some systematics only apply

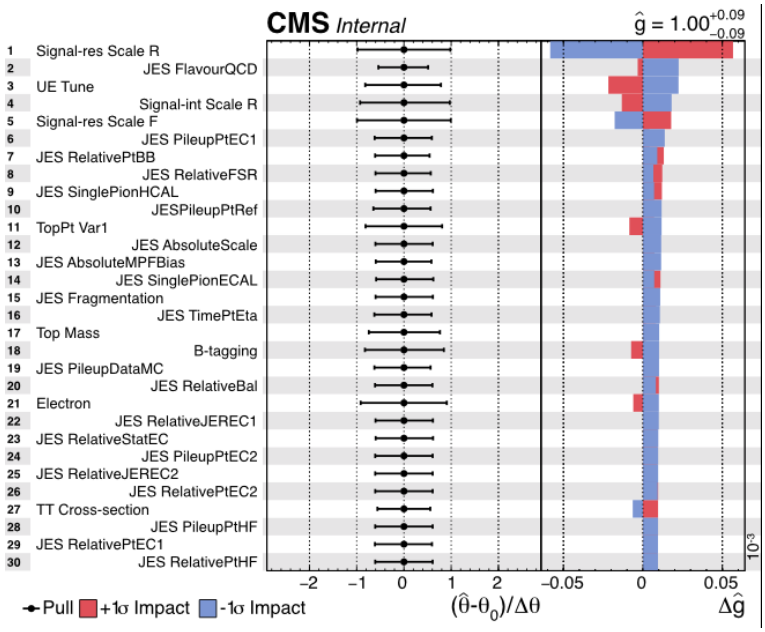


Figure 8.6.: Impact from systematics on pseudoscalar signal with mass 500 GeV and relative width of 2.5%.

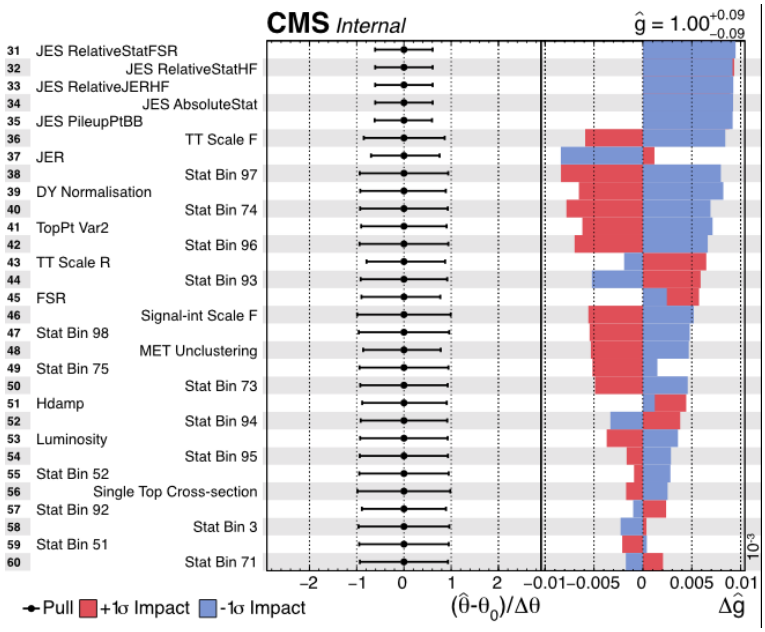


Figure 8.7.: Impact from systematics on pseudoscalar signal with mass 500 GeV and relative width of 2.5%.

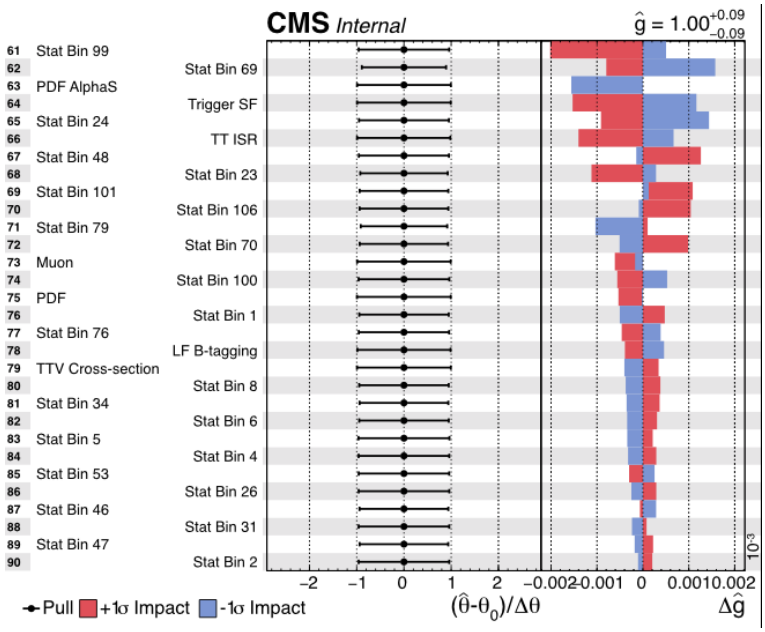


Figure 8.8.: Impact from systematics on pseudoscalar signal with mass 500 GeV and relative width of 2.5%.

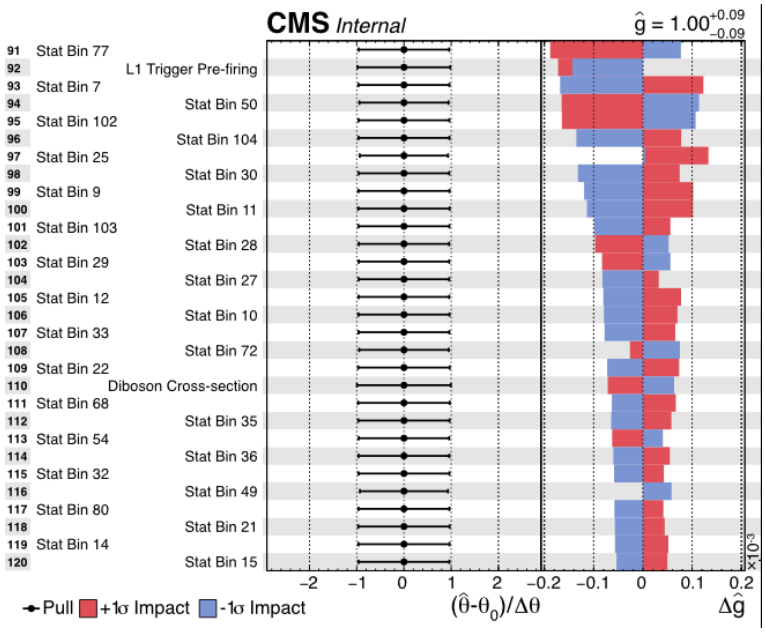


Figure 8.9.: Impact from systematics on pseudoscalar signal with mass 500 GeV and relative width of 2.5%.

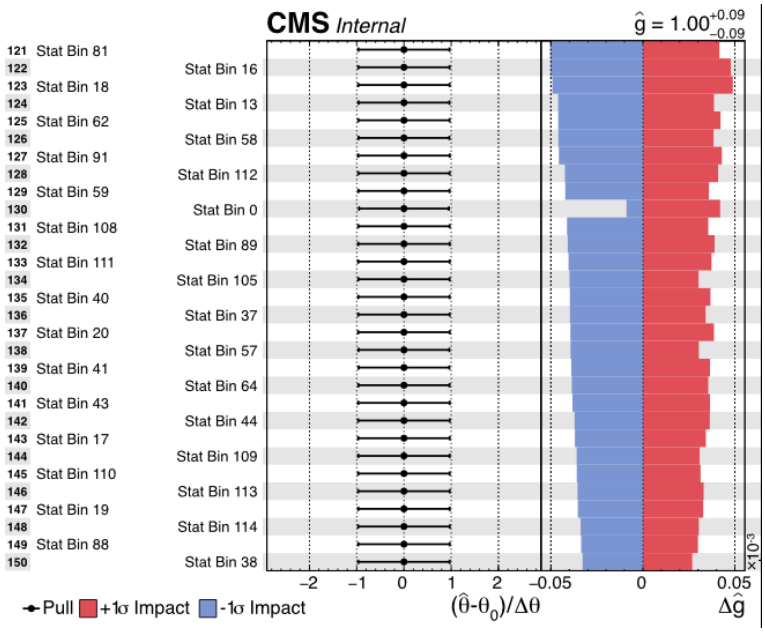


Figure 8.10.: Impact from systematics on pseudoscalar signal with mass 500 GeV and relative width of 2.5%.

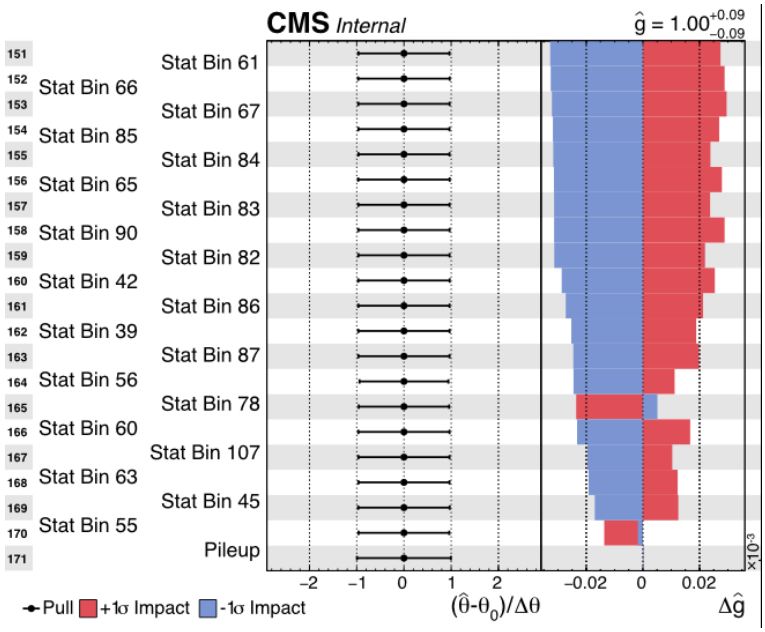


Figure 8.11.: Impact from systematics on pseudoscalar signal with mass 500 GeV and relative width of 2.5%.

Systematic	Process	Type
Electron	All	shape
Muon	All	shape
Jet Energy Resolution	Signal, $t\bar{t}$	shape
Jet Energy Scale	Signal, $t\bar{t}$	shape
MET Unclustering	Signal, $t\bar{t}$	shape
Scale R	Signal, $t\bar{t}$	shape
Scale F	Signal, $t\bar{t}$	shape
L1 Trigger Pre-firing	All	shape
Trigger SF	All	shape
Pileup	All	shape
B-tagging	All	shape
LF B-tagging	All	shape
PDF	$t\bar{t}$	shape
TopPt Var1	$t\bar{t}$	shape
TopPt Var2	$t\bar{t}$	shape
Top Mass	$t\bar{t}$	shape
Hdamp	$t\bar{t}$	shape
ISR	$t\bar{t}$	shape
FSR	$t\bar{t}$	shape
PDF α_s	$t\bar{t}$	shape
UE Tune	$t\bar{t}$	shape
$t\bar{t}$ Cross-section	$t\bar{t}$	norm
$t\bar{t}+V$ Cross-section	$t\bar{t}+V$	norm
Single Top Cross-section	t	norm
Diboson Cross-section	VV	norm
DY Normalisation	$V+$ jets	norm
Luminosity	Signal, $t\bar{t}$, $t\bar{t}+V$, t, VV	norm

Table 8.5.: Systematics used in the analysis.

to the $t\bar{t}$ samples. These are: Top mass, ISR and FSR along with tuning of the underlying event and the h_{damp} matching parameter. The smoothing is done with the LOWESS method [17] to maintain features from the systematics while suppressing fluctuations. With this method, a scatterplot is made from the events with a very fine binning to produce a scatterplot of random fluctuations. The scatterplot is then fitted with an overall trend, which is then the smoothed variation. The $m_{t\bar{t}}$ and c_{hel} distributions are used for the systematic samples as well as the nominal sample, and a 2d-binning in the $c_{hel}, m_{t\bar{t}}$ plane of 10,100 is made in such a manner that each bin contains an approximately equal number of events for the nominal

samples. For cross validation purposes, 100 partitions are made. For each bin in c_{hel}, the relative difference between the $m_{t\bar{t}}$ distribution for the nominal samples and the given systematic samples is then fitted according to overall trends in the distribution.

8.5. Analysis Results

The analysis itself is performed in the $m_{t\bar{t}}$, c_{hel} - plane on all three dileptonic channels added together. The c_{hel} variable is one of the spin correlation variables with a strong discrimination power between scalar and pseudoscalar scenarios [18]. It also improves the sensitivity of the analysis compared to relying only on $m_{t\bar{t}}$.

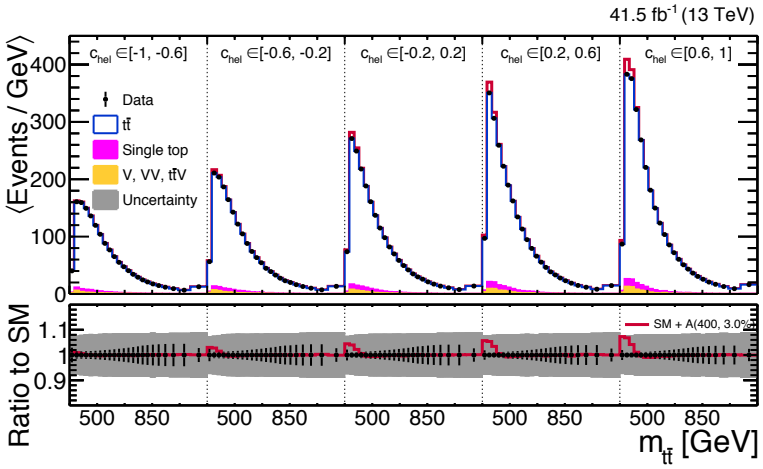


Figure 8.12.: Analysis region with the overlaid signal of a pseudoscalar with a mass of 400 GeV and a width of 3% (red line) with the prefit systematic nuisances (grey shaded area) and an Asimov dataset with statistical errors (black points).

The 2D plane is unrolled to facilitate analysis with the Higgs-Combine tool [19]. In fig. 8.12 we see an example of a signal hypothesis with the systematic errors before fitting and in fig. 8.13 we see the same signal hypothesis after fitting according to the background only hypothesis. The fitting procedure generally follows the procedure stated in section 7.5 with use of the asymptotic method based on fits to Asimov datasets. There are some modifications made to incorporate the interference part of the signal, which can lead to a deficit in event count in some bins. The interference part of the signal also scales differently with regards to the

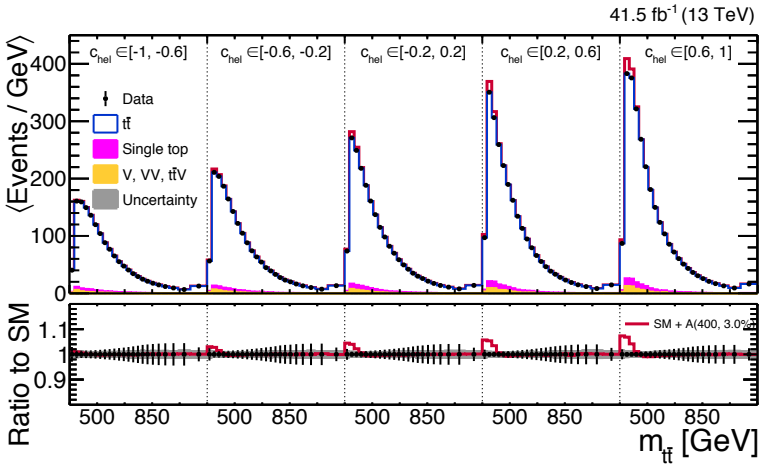


Figure 8.13.: Analysis region with the overlaid signal of a pseudoscalar with a mass of 400 GeV and a width of 3% (red line) with the systematic nuisances after fitting to the background only hypothesis (grey shaded area) and an Asimov dataset with statistical errors (black points).

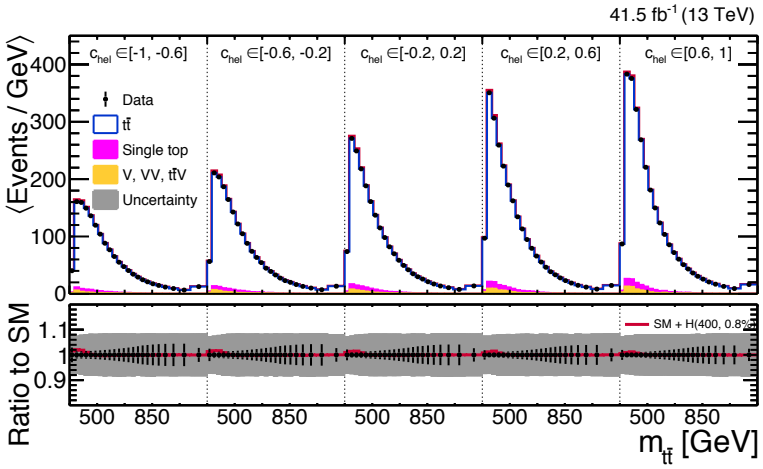


Figure 8.14.: Analysis region with the overlaid signal of a scalar with a mass of 400 GeV and a width of 0.8% (red line) with the prefit systematic nuisances (grey shaded area) and an Asimov dataset with statistical errors (black points).

coupling parameter g than the resonance, with the resonance scaling as $\sigma_R \propto g^4$ and the interference scaling as $\sigma_I \propto g^2$ as explained in section 2.3.1. Therefore the results are given in terms of a scaling factor to the Heavy Higgs coupling rather than the signal cross-section. To reach the 95% confidence level, a scan is made over different values of g and the closest value for g to 95% confidence level after passing the threshold is selected as the limit.

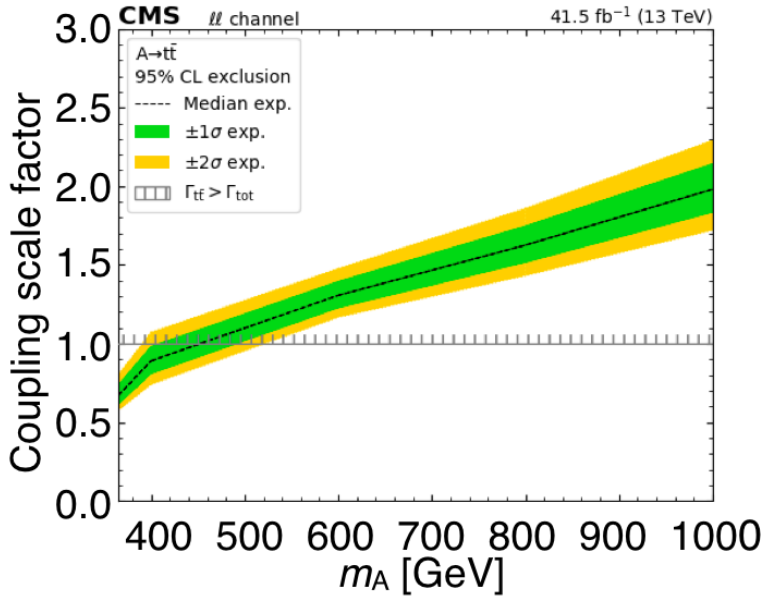
Each signal hypothesis is evaluated individually, and the resulting limit is summarised in figure 8.15a for pseudoscalar hypotheses and in figure 8.16a for scalar hypotheses. Since these results are usually presented at a fixed width, the widths used for limit setting at the given masses are shown in figures 8.15b and 8.16b.

There are some mass points that have widths comparable to previous analyses that can be used to evaluate the effectiveness of this search. One such point is the pseudoscalar with a mass of 600 GeV that has a width close to 5%. The expected limit for this mass point is estimated to exclude couplings down to $g_{At\bar{t}} = 1.3$ which is comparable with the estimate from 2016 on the dilepton channel only at 1.35 [2].

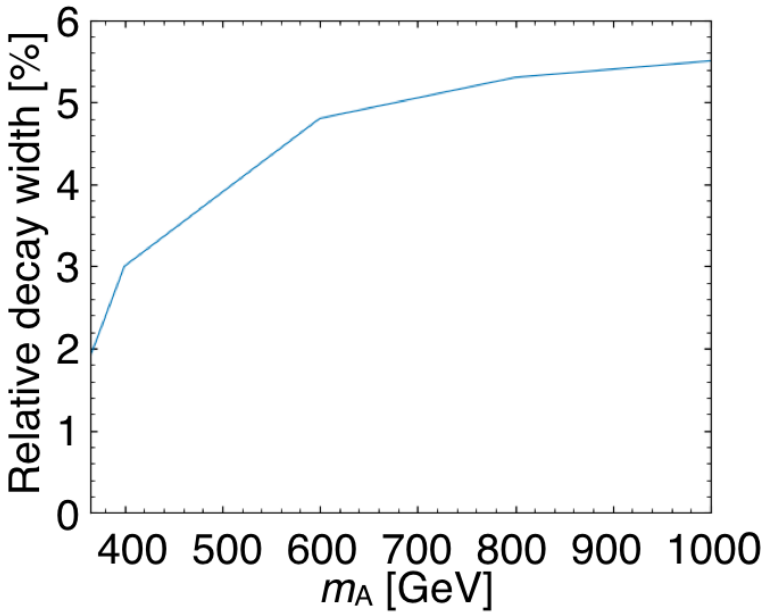
8.6. Perspective and Outlook

Agreement between data and the SM prediction is achieved in all considered control regions and all relevant systematics are considered. Smoothing has been applied to a set of systematics in the SM background to mitigate the effects of MC statistical fluctuations therein. Expected limits have been set that demonstrate that the sensitivity of the analysis is comparable to the result from 2016. To improve the sensitivity in the analysis, one can either increase the signal yield or one can reduce the statistical and systematic errors in the background, signal and the detector. The signal yield can be increased by adding signal samples for the semileptonic channel as well, which will lead to a projected increase in signal yield of $\sim 5\%$. The MC statistics are at a satisfactory level for the dominant SM backgrounds in this analysis, and the MC statistics of the signal have improved by about a factor 2 by filtering signal events prior to simulation.

Due to an excess observed in the 2016 analysis for the exchange of a pseudoscalar Higgs boson at a mass of 400 GeV, with a local significance of $3.5\,\sigma$ and a global significance of $1.9\,\sigma$, it has not been possible to obtain permission to unblind the 2017 data analysis and look into the data in the signal region. This has to

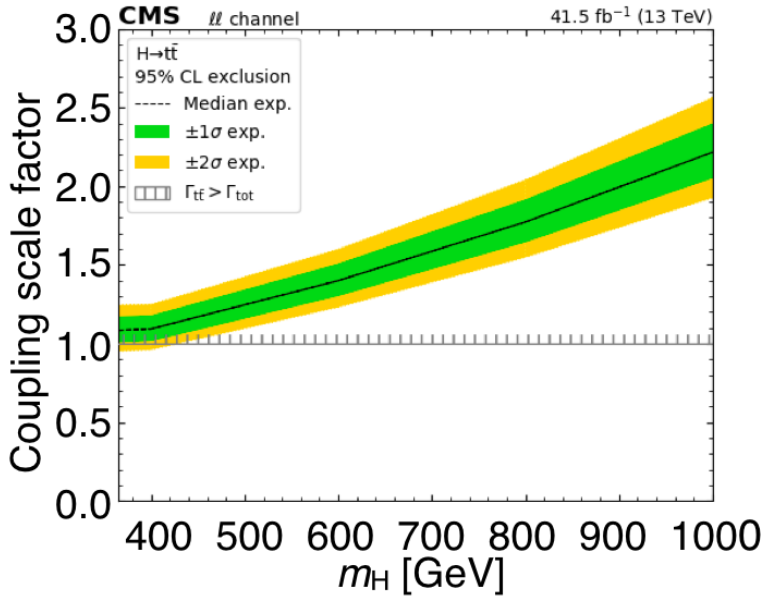


(a)

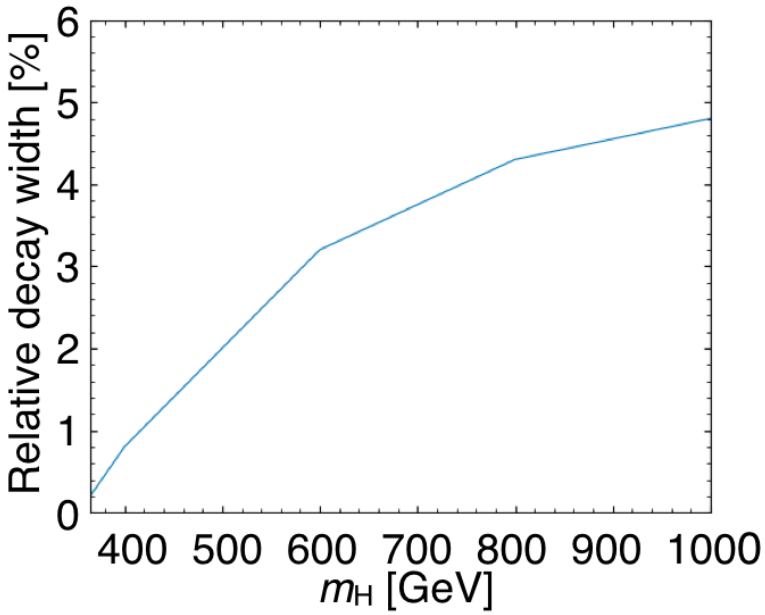


(b)

Figure 8.15.: Expected limit for pseudoscalar signal at 95% confidence level, given in terms of the coupling parameter g as a function of pseudoscalar mass, with the median limit (black line), one σ band (green) and two σ band (yellow) (a). The relative decay width as a function of pseudoscalar mass (b), following the criterion that the branching fraction into a $t\bar{t}$ becomes greater than unity when the coupling parameter $g > 1$, leading to an unphysical result.



(a)



(b)

Figure 8.16.: Expected limit for scalar signal at 95% confidence level, given in terms of the coupling parameter g as a function of scalar mass, with the median limit (black line), one σ band (green) and two σ band (yellow) (a). The relative decay width as a function of scalar mass (b), following the criterion that the branching fraction into a $t\bar{t}$ becomes greater than unity when the coupling parameter $g > 1$, leading to an unphysical result.

wait until the 2018 data analysis is finished and all datasets for 2016-2018 can be combined.

Bibliography

[1] W. Bernreuther et al., *'Top quark pair production and decay at hadron colliders'*, Nucl. Phys. B 690 81 (2004) 81 [hep-ph/0403035].

[2] A. Anuar, *'Top Quark Polarization Properties in Searches for New Phenomena with the CMS Detector at the LHC'*,
PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2019, doi:10.3204/PUBDB-2020-00203.

[3] L. Sonnenschein, *'Analytical solution of $t\bar{t}$ dilepton equations'*,
Phys.Rev.D73:054015,2006; Erratum-ibid.D78:079902,2008, doi: 10.1103/PhysRevD.73.054015 10.1103/PhysRevD.78.079902, arXiv:hep-ph/0603011v3.

[4] I. Korol, *'Measurement of Double Differential $t\bar{t}$ Production Cross Sections with the CMS Detector'*,
PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2016. doi:10.3204/DESY-THESIS-2016-011.

[5] R. Frederix et al., *'Merging meets matching in MC@NLO'*,
J. High Energ. Phys. 2012, 61 (2012), doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[6] T. Arndt et al., *'Measurement of the 2016 Trigger Efficiencies for a dilepton selection for a ttbar analysis'*,
CMS Analysis Note CMS AN-2016/392, 2016.

[7] R. V. Harlander et al., *'SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM'*,
Comput.Phys.Commun. 184 (2013) 1605-1617, doi:10.1016/j.cpc.2013.02.006, arXiv:1212.3249v2.

[8] **CMS Collaboration**, *'Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV'*,
Eur. Phys. J. C75 (2015), no. 11, 542, doi:10.1140/epjc/s10052-015-3709-x, arXiv:1505.04480.

[9] D. Eriksson et al., *'2HDMC - Two-Higgs-Doublet Model Calculator'*,
Comput.Phys.Commun.181:189-205,2010, doi:10.1016/j.cpc.2009.09.011, arXiv:0902.0851v2.

[10] M. Czarkon et al., *'Top++: a program for the calculation of the top-pair cross-section at hadron colliders'*,
Comput.Phys.Commun. 185 (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675v4.

[11] **CMS Collaboration**, *'Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV'*,
JINST 12 (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[12] M. Rieger, *'Search for $t\bar{t}H$ Production in the $H \to b\bar{b}$ Decay Channel: Using Deep Learning Techniques with the CMS Experiment'*,
ISBN 3030653803, 9783030653804.

[13] **CMS Collaboration**, *'Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector'*,
Journal of Instrumentation, Vol. 14, July 4 2019, doi:10.1088/1748-0221/14/07/p07004.

[14] T. Plehn, *'LHC Phenomenology for Physics Hunters'*,
arXiv:0810.2281v2 [hep-ph] 7 Feb 2009.

[15] **CMS Collaboration**, *'Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV'*,
Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016.

[16] **CMS Collaboration**, *'Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements'*,
Eur. Phys. J. C 80, 4, 3 Jan 2020 doi:10.1140/epjc/s10052-019-7499-4.

[17] W. S. Cleveland, *'Robust locally weighted regression and smoothing scatter-*

plots',
J. Am. Stat. Assoc. 74 (1979) 829, doi:10.2307/2286407.

[18] **CMS Collaboration**, *'Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at* $\sqrt{s} = 13$ TeV*'*,
JHEP 04 (2020) 171, doi:10.1007/JHEP04(2020)171, arXiv:1908.01115v2.

[19] **ATLAS and CMS Collaborations, and LHC Higgs Combination Group**, *'Procedure for the LHC Higgs boson search combination in Summer 2011'*,
Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011.

9. Summary

In this thesis, two searches for different scenarios of physics beyond the Standard model are presented. One involving WIMP DM particles, Z' DM mediators and Dark Higgs bosons. The other involving Heavy Higgs bosons that can act as DM mediators decaying into top quark pairs.

The first search for mono Dark Higgs production using the 2016 dataset of $35.9\,\text{fb}^{-1}$ explores the signature where the missing transverse momentum is balanced against a pair of b-jets. Regions with high values of the scalar sum of the transverse momenta of jets, H_T, are investigated. In the multijet regions of the analysis, the background is expected to peak at low values of the magnitude of the negative vector sum of the jet transverse momenta H_T^miss, whereas the signal is expected to be concentrated at high values.

Since no excess in data is visible at large values of H_T^miss and overall agreement with the SM expectation is observed, exclusions on the signal hypotheses are made. Upper expected and observed limits at 95% C.L. on the signal cross-sections are extracted and translated into the excluded regions of the mass of the DM particles and the mass of the Z' DM mediator. DM mediator masses of up to 1700 GeV and DM particle masses up to 300 GeV can be excluded at 95% C.L. These limits represent the first constraints ever set on Dark Higgs models using experimental data.

The second search explores the top-antitop quark pair production in the dileptonic final states with large missing transverse energy and two jets whereof at least one jet is b-tagged, utilising $41.5\,\text{fb}^{-1}$ of 2017 data. The production of heavy scalar and pseudoscalar bosons is investigated. Since Heavy Higgs production with a subsequent decay into a $t\bar{t}$ pair interferes with SM $t\bar{t}$ production, the signal would appear as a peak-dip structure rather than a pure resonance peak in the invariant $t\bar{t}$ mass distribution. This unusual signature provides additional challenges due to the sign change in the signal. Events with an intermediate exchange of a heavy (pseudo)scalar Higgs boson can be distinguished even further from pure SM production since they will significantly change the correlation of spin between the top

quark and the antitop quark. On such variable sensitive to the $t\bar{t}$ spin correlation, c_{hel}, is investigated simultaneously with the $m_{t\bar{t}}$ distribution. The usage of the c_{hel} distribution not only enhances the sensitivity to find heavy Higgs bosons, but will in addition distinguish the case of a scalar from the case of a pseudoscalar coupling to the top quark. Such information is essential in the case of discovery to determine the physical consequences thereof including determining whether there is proof of DM production or other exotic physics in play.

In the analysis, agreement between data and the SM prediction is achieved in all considered control regions and all relevant systematics are considered. Expected upper limits on the coupling strength between the heavy Higgs boson and the top quark are set as a function of the mass hypothesis for the heavy scalar or pseudoscalar Higgs boson. For comparable $t\bar{t}$ decay widths of the Heavy Higgs, these limits are similar to the expected limits derived in the respective analysis using 2016 data involving $35\,\mathrm{fb}^{-1}$ of data [1][2].

Due to an excess observed in the 2016 analysis for the exchange of a pseudoscalar Higgs boson at a mass of 400 GeV, with a local significance of $3.5\,\sigma$ and a global significance of $1.9\,\sigma$, it has not been possible to obtain permission to unblind the 2017 data analysis and look into the data in the signal region. This has to wait until the 2018 data analysis is finished and all datasets for 2016-2018 can be combined. Consequently, the answer cannot be given whether the result from 2016 is a fluctuation or if it is leading to a ground-breaking discovery.

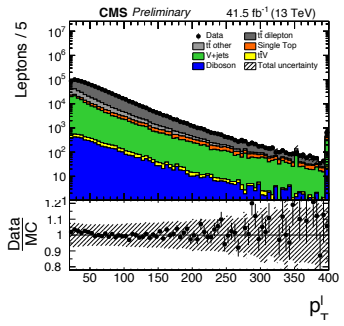
Bibliography

[1] CMS Collaboration, *'Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV'*,
JHEP 04 (2020) 171, doi:10.1007/JHEP04(2020)171, arXiv:1908.01115v2.

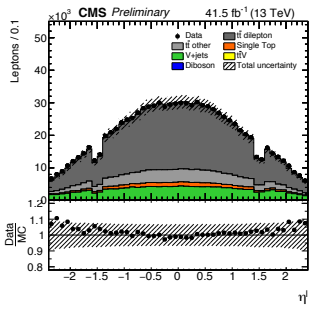
[2] A. Anuar, *'Top Quark Polarization Properties in Searches for New Phenomena with the CMS Detector at the LHC'*,
PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2019, doi:10.3204/PUBDB-2020-00203.

A. $t\bar{t}$ Control Plots

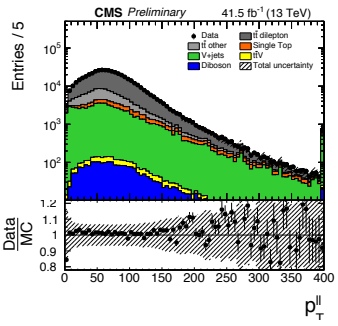
This appendix contains the control plots for the 2017 Heavy Higgs search at the preceding selection steps with respect to the control plots reported in Chapter 7.



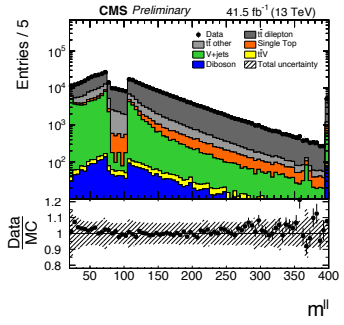
(a) Transverse momentum of leptons



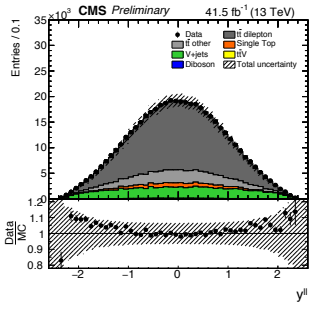
(b) Pseudorapidity of leptons



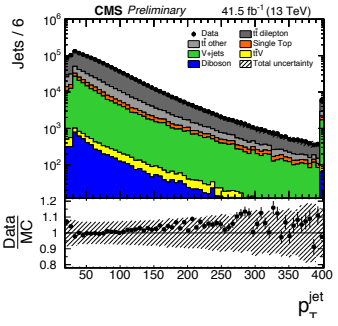
(c) Transverse momentum of dileptons



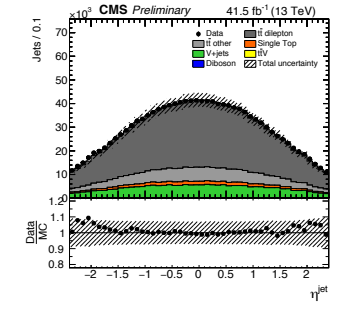
(d) Mass of dileptons



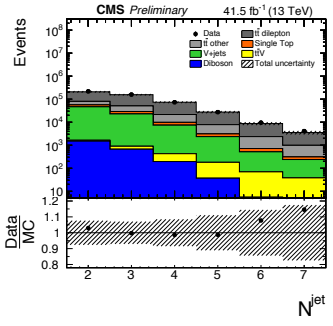
(e) Rapidity of dileptons



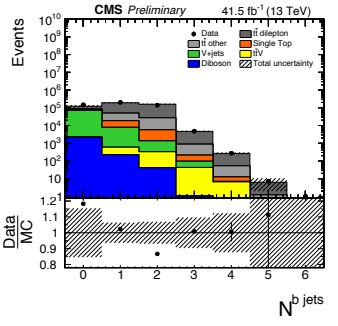
(f) Transverse momentum of jets



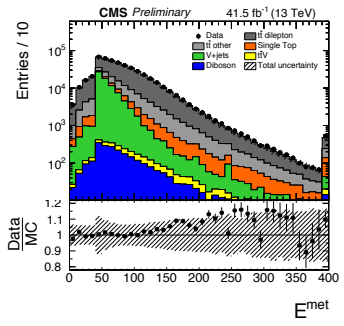
(g) Pseudorapidity of jets



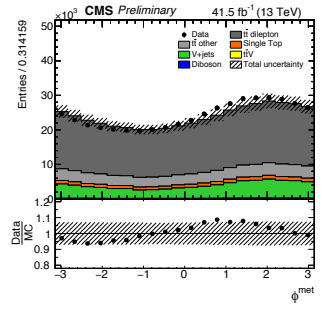
(h) Jet multiplicity



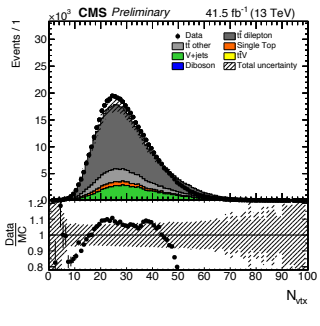
(i) B-tagged jet multiplicity



(j) Missing E_T

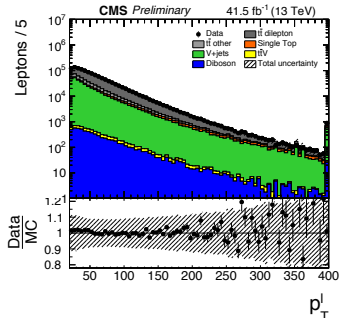


(k) Missing E_T azimuthal angle

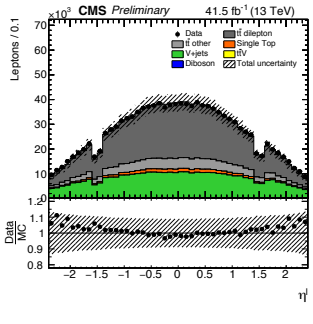


(l) Vertex multiplicity

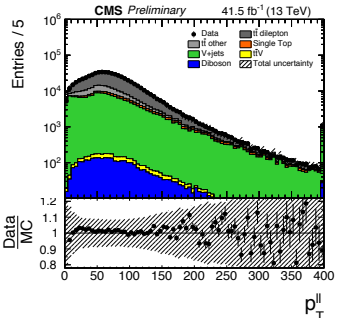
Figure A.1.: Control distributions for the combined channel with all cuts prior to b-tagging applied



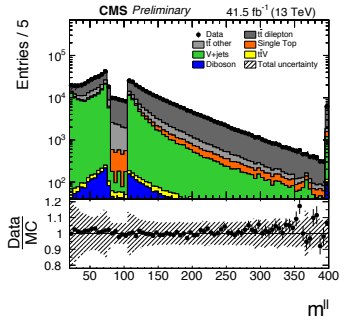
(a) Transverse momentum of leptons



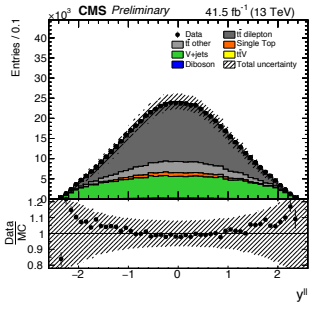
(b) Pseudorapidity of leptons



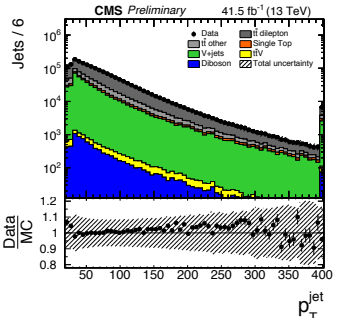
(c) Transverse momentum of dileptons



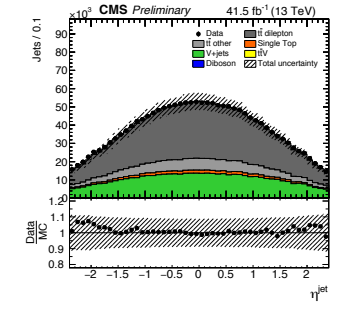
(d) Mass of dileptons



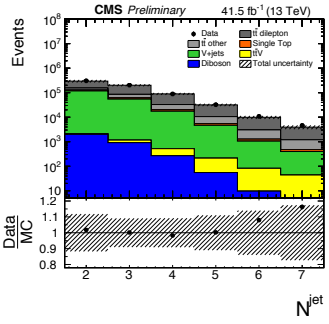
(e) Rapidity of dileptons



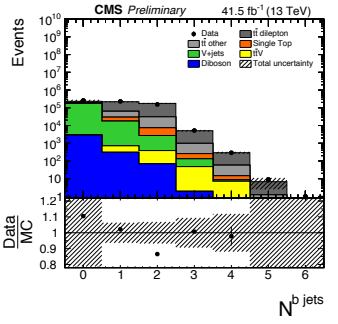
(f) Transverse momentum of jets



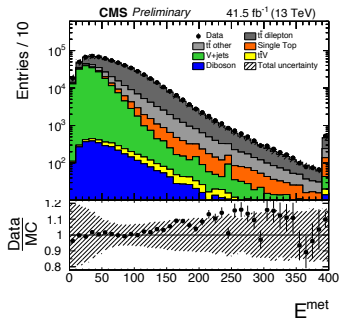
(g) Pseudorapidity of jets



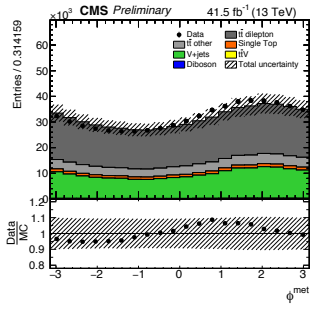
(h) Jet multiplicity



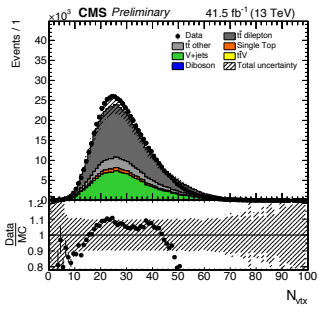
(i) B-tagged jet multiplicity



(j) Missing E_T

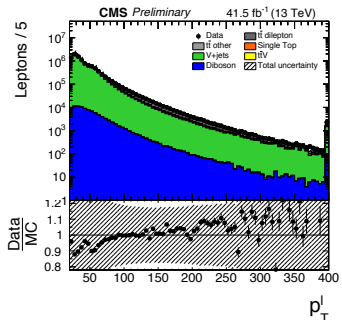


(k) Missing E_T azimuthal angle

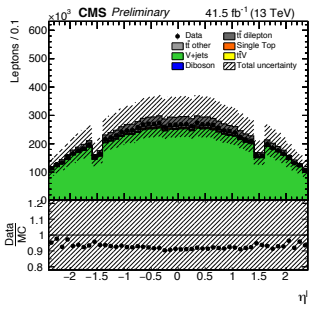


(l) Vertex multiplicity

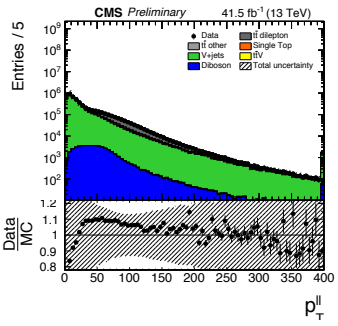
Figure A.2.: Control distributions for the combined channel with all cuts prior to MET applied



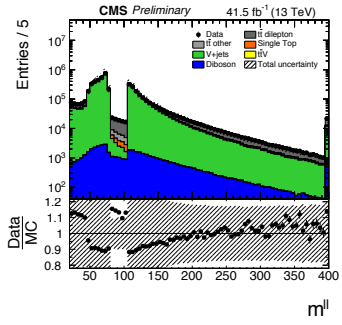
(a) Transverse momentum of leptons



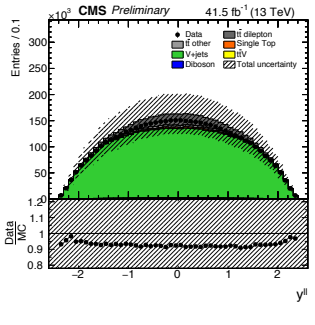
(b) Pseudorapidity of leptons



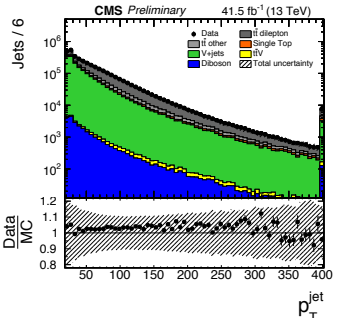
(c) Transverse momentum of dileptons



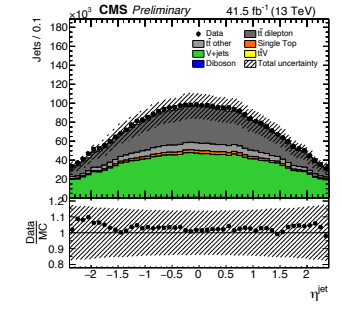
(d) Mass of dileptons



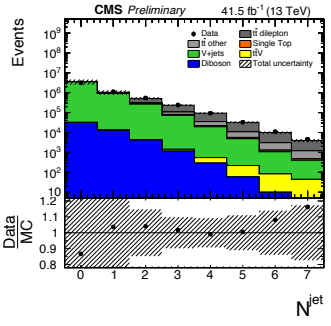
(e) Rapidity of dileptons



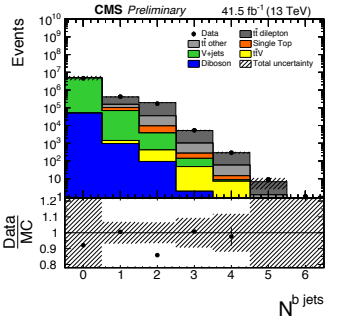
(f) Transverse momentum of jets



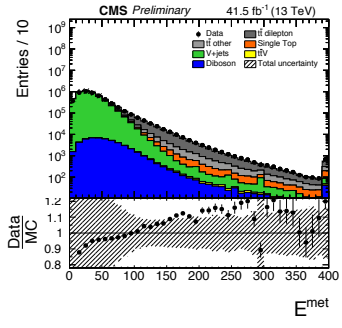
(g) Pseudorapidity of jets



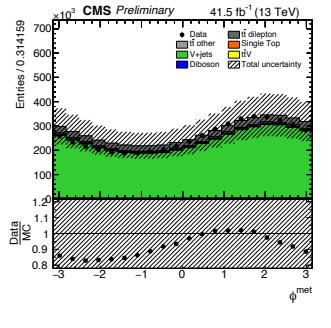
(h) Jet multiplicity



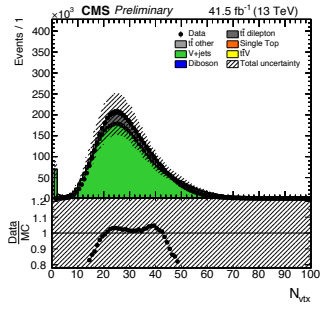
(i) B-tagged jet multiplicity



(j) Missing E_{T}

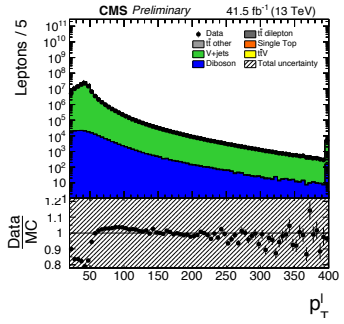


(k) Missing E_{T} azimuthal angle

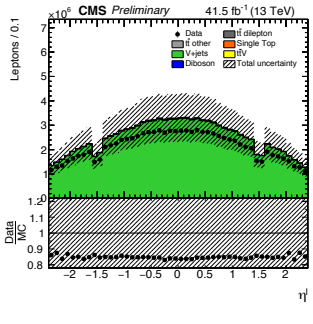


(l) Vertex multiplicity

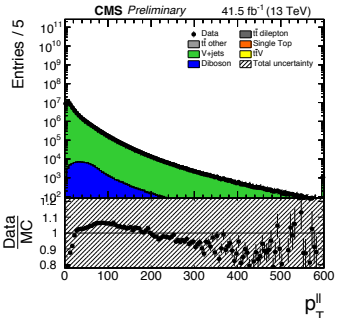
Figure A.3.: Control distributions for the combined channel with all cuts prior to jet selection applied



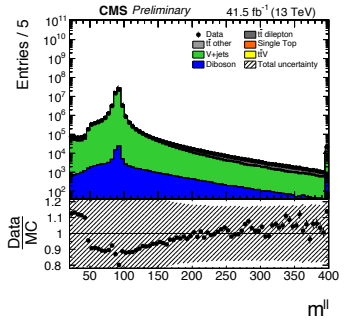
(a) Transverse momentum of leptons



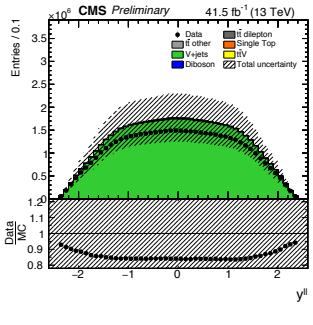
(b) Pseudorapidity of leptons



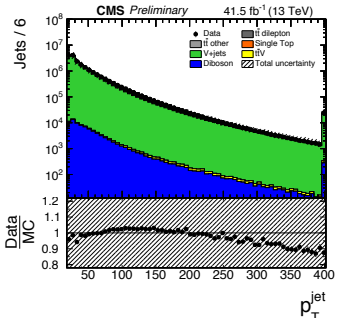
(c) Transverse momentum of dileptons



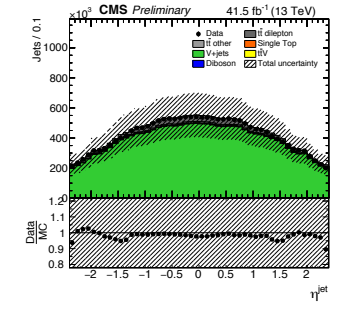
(d) Mass of dileptons



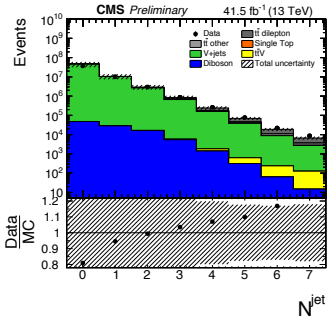
(e) Rapidity of dileptons



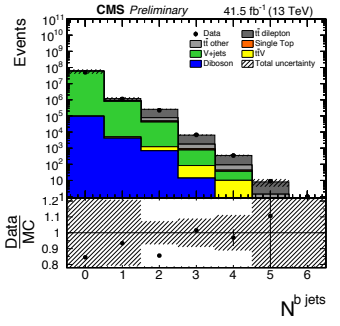
(f) Transverse momentum of jets



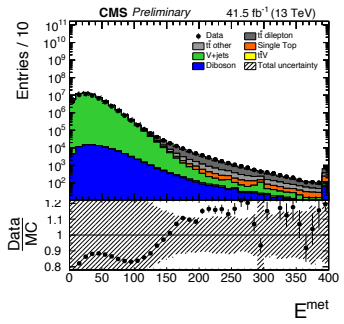
(g) Pseudorapidity of jets



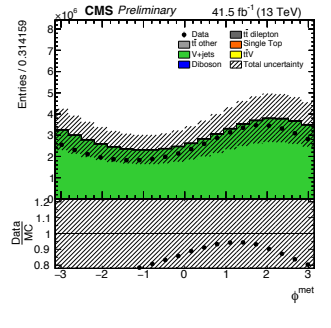
(h) Jet multiplicity



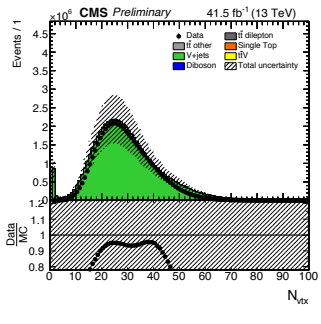
(i) B-tagged jet multiplicity



(j) Missing E_{T}

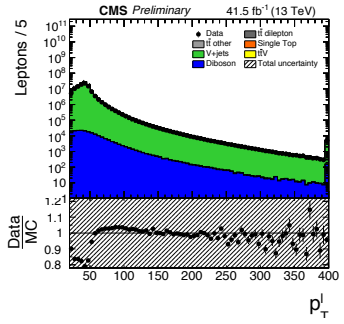


(k) Missing E_{T} azimuthal angle

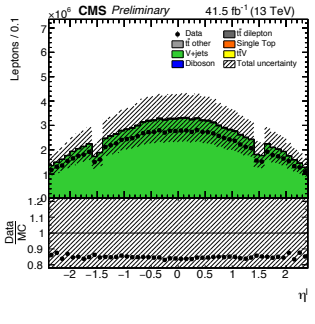


(l) Vertex multiplicity

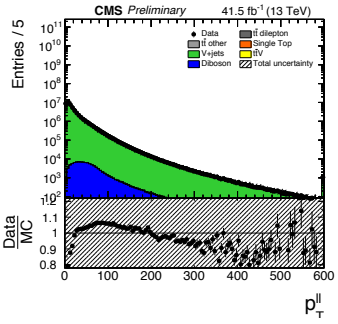
Figure A.4.: Control distributions for the combined channel with all cuts prior to Z-window applied



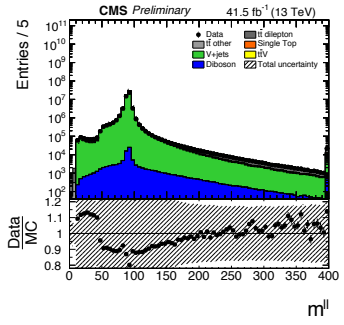
(a) Transverse momentum of leptons



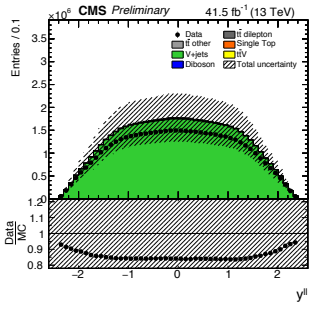
(b) Pseudorapidity of leptons



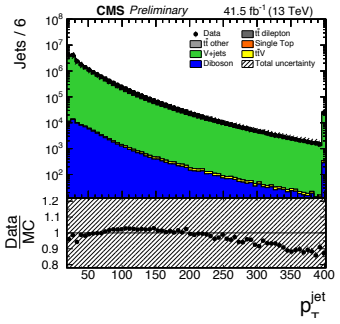
(c) Transverse momentum of dileptons



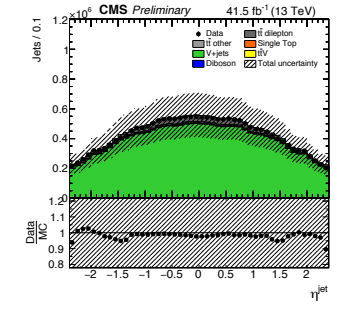
(d) Mass of dileptons



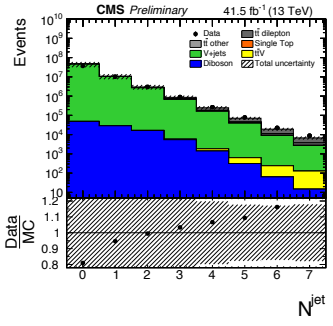
(e) Rapidity of dileptons



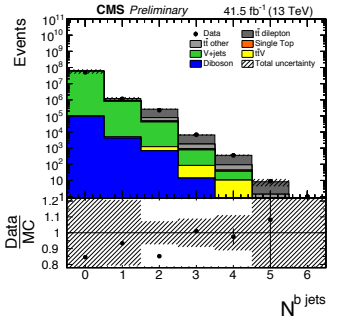
(f) Transverse momentum of jets



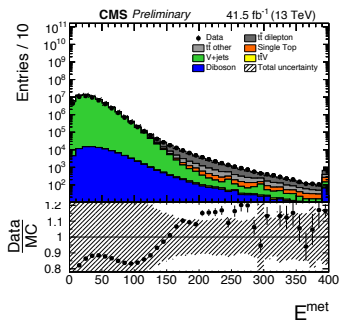
(g) Pseudorapidity of jets



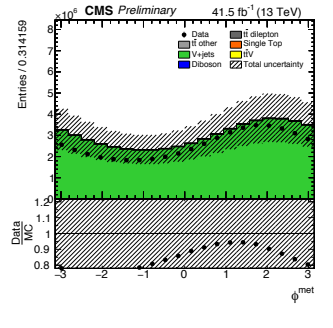
(h) Jet multiplicity



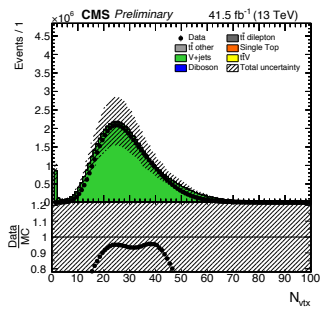
(i) B-tagged jet multiplicity



(j) Missing E_T

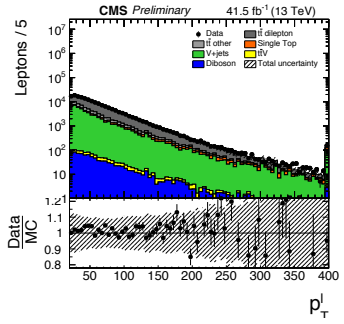


(k) Missing E_T azimuthal angle

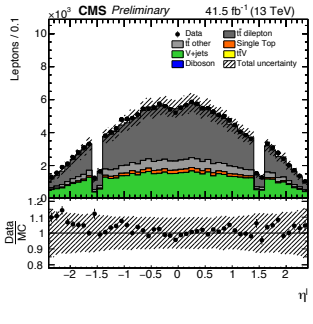


(l) Vertex multiplicity

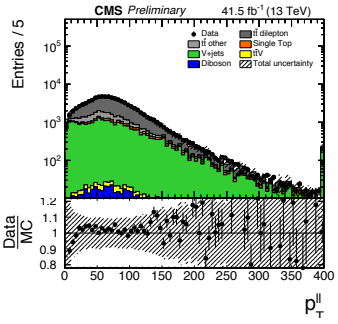
Figure A.5.: Control distributions for the combined channel with all cuts prior to dilepton mass applied



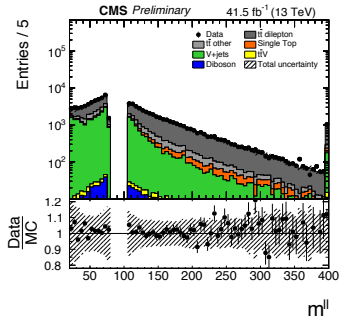
(a) Transverse momentum of leptons



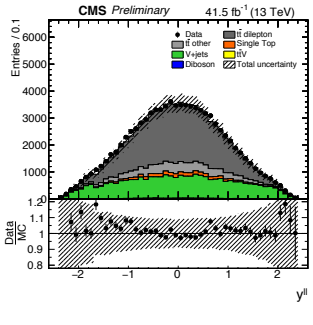
(b) Pseudorapidity of leptons



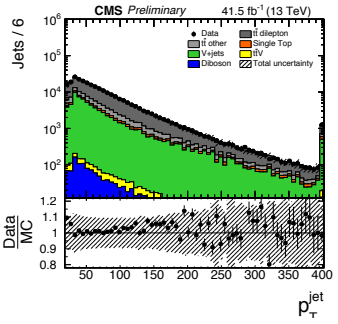
(c) Transverse momentum of dileptons



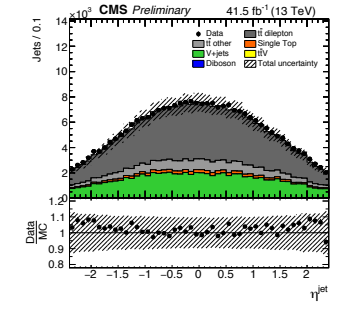
(d) Mass of dileptons



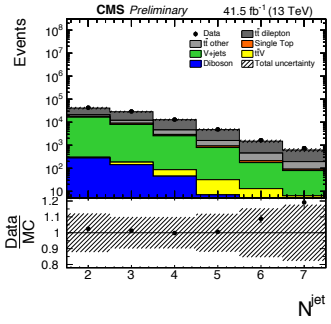
(e) Rapidity of dileptons



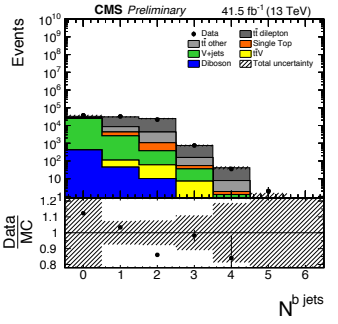
(f) Transverse momentum of jets



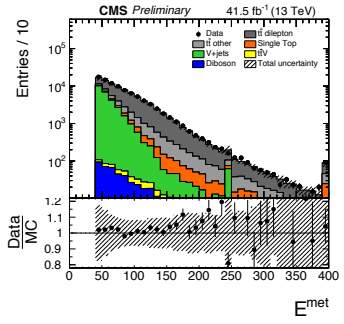
(g) Pseudorapidity of jets



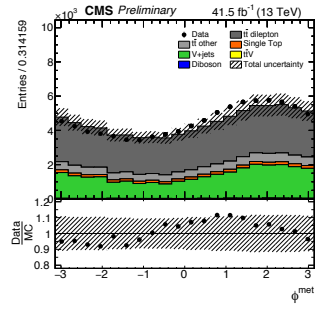
(h) Jet multiplicity



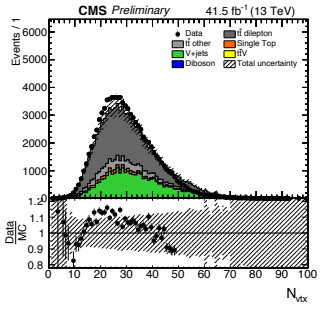
(i) B-tagged jet multiplicity



(j) Missing E_{T}

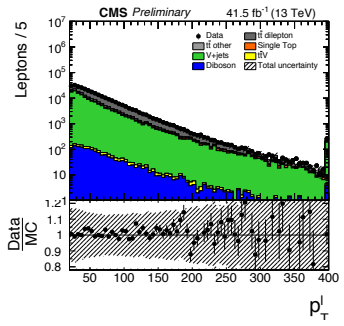


(k) Missing E_{T} azimuthal angle

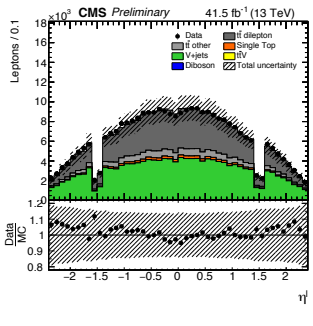


(l) Vertex multiplicity

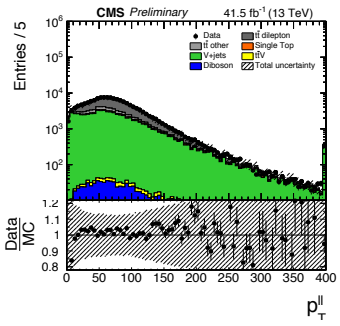
Figure A.6.: Control distributions for the *ee* channel with all cuts prior to b-tagging applied



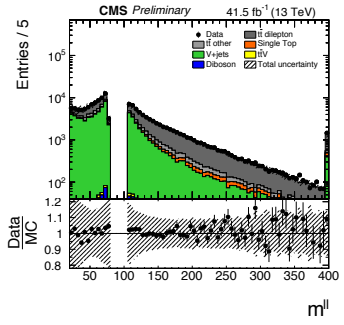
(a) Transverse momentum of leptons



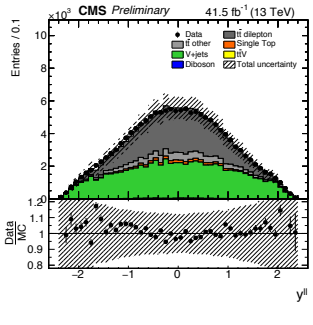
(b) Pseudorapidity of leptons



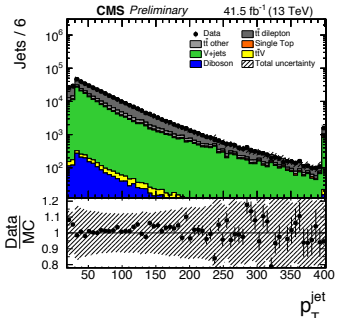
(c) Transverse momentum of dileptons



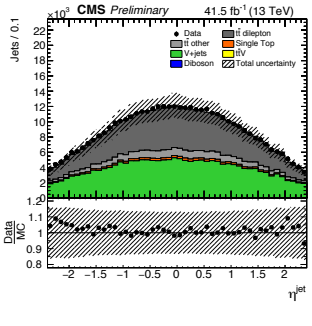
(d) Mass of dileptons



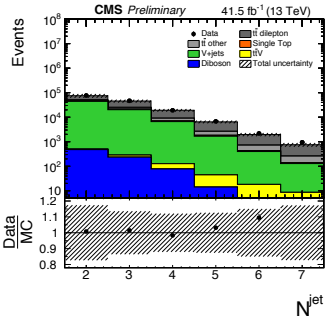
(e) Rapidity of dileptons



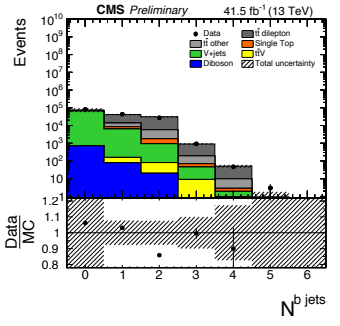
(f) Transverse momentum of jets



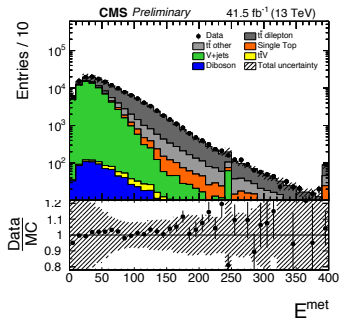
(g) Pseudorapidity of jets



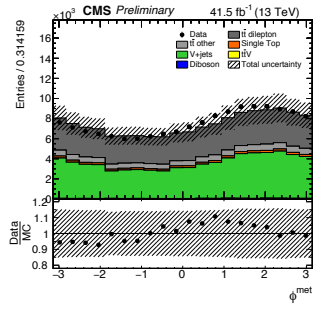
(h) Jet multiplicity



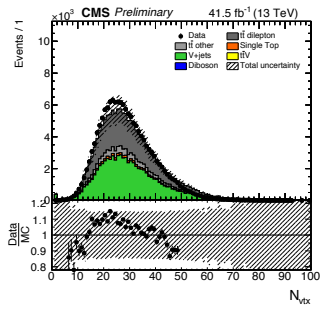
(i) B-tagged jet multiplicity



(j) Missing E_T

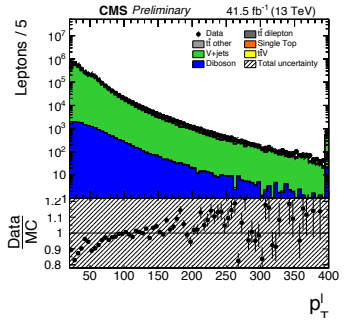


(k) Missing E_T azimuthal angle

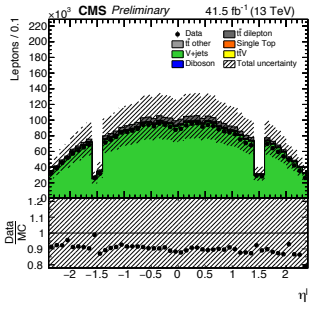


(l) Vertex multiplicity

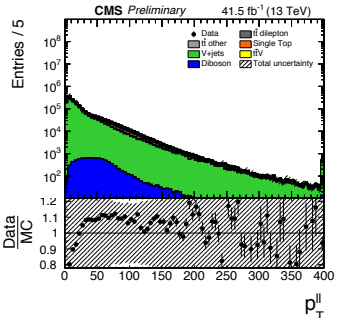
Figure A.7.: Control distributions for the *ee* channel with all cuts prior to MET applied



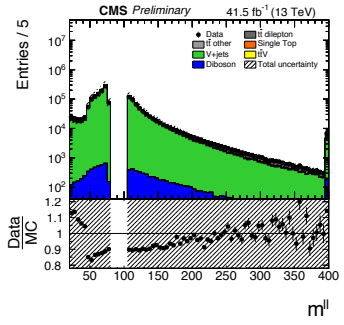
(a) Transverse momentum of leptons



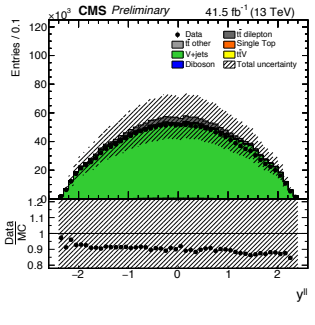
(b) Pseudorapidity of leptons



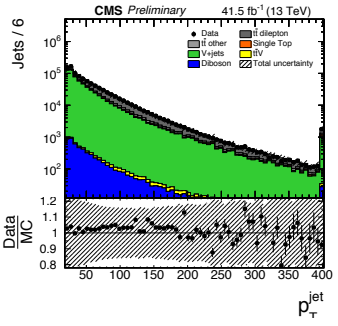
(c) Transverse momentum of dileptons



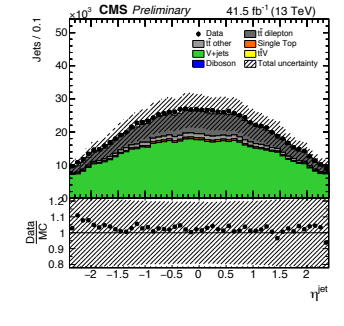
(d) Mass of dileptons



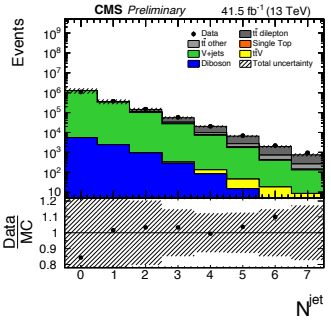
(e) Rapidity of dileptons



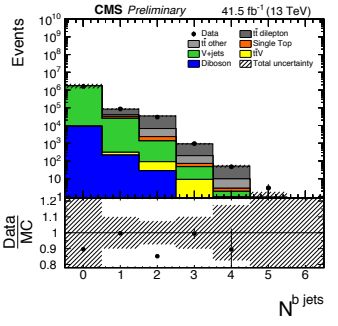
(f) Transverse momentum of jets



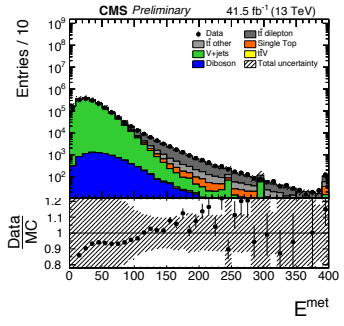
(g) Pseudorapidity of jets



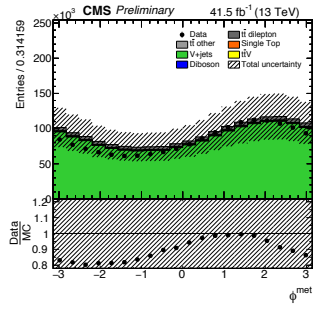
(h) Jet multiplicity



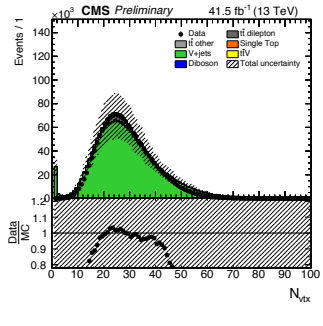
(i) B-tagged jet multiplicity



(j) Missing E_{T}

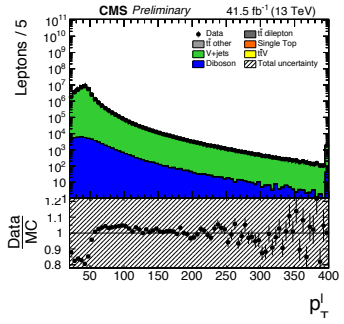


(k) Missing E_{T} azimuthal angle

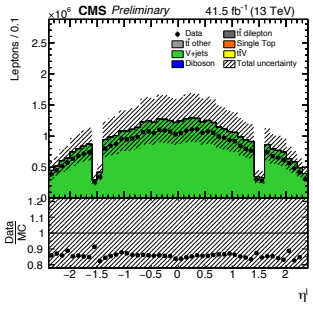


(l) Vertex multiplicity

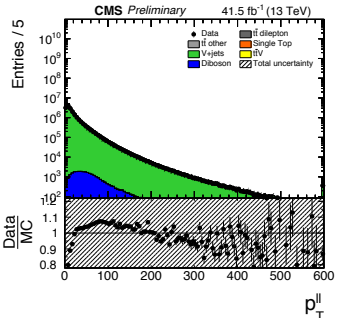
Figure A.8.: Control distributions for the *ee* channel with all cuts prior to jet selection applied



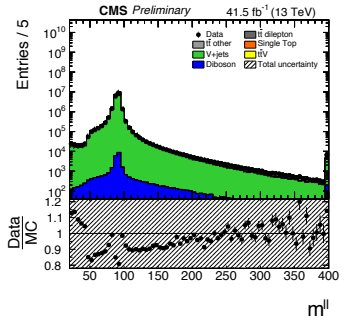
(a) Transverse momentum of leptons



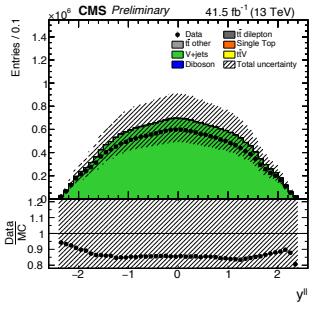
(b) Pseudorapidity of leptons



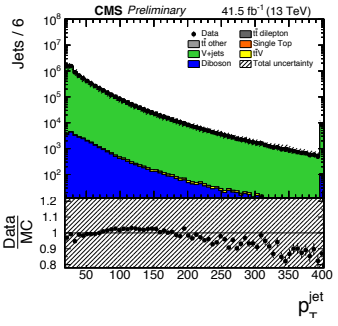
(c) Transverse momentum of dileptons



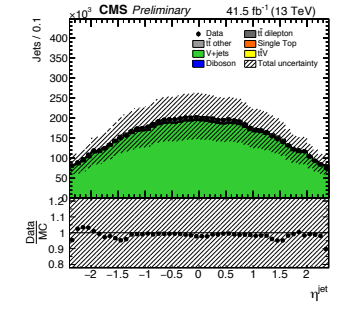
(d) Mass of dileptons



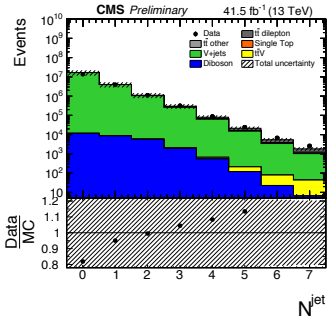
(e) Rapidity of dileptons



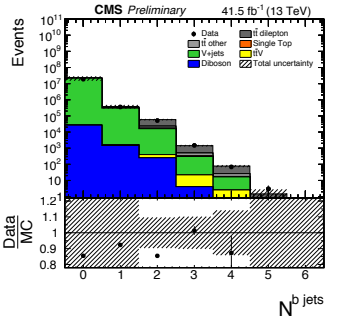
(f) Transverse momentum of jets



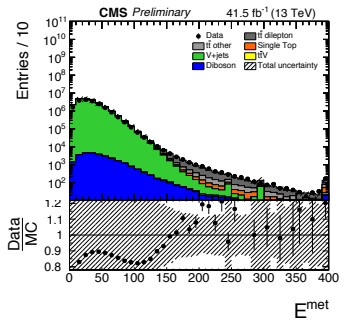
(g) Pseudorapidity of jets



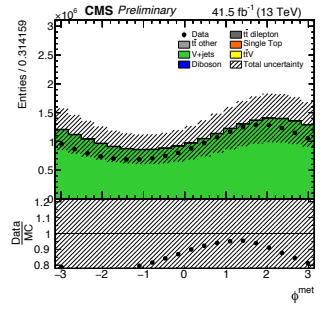
(h) Jet multiplicity



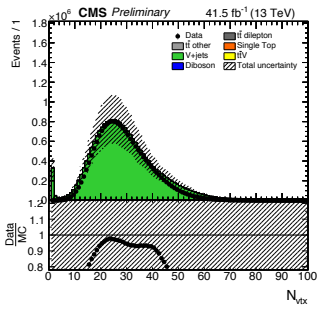
(i) B-tagged jet multiplicity



(j) Missing E_{T}

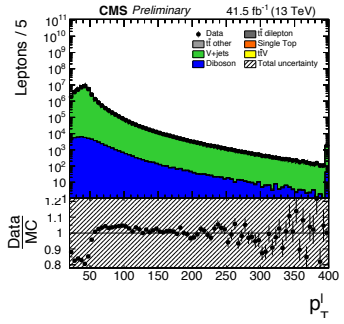


(k) Missing E_{T} azimuthal angle

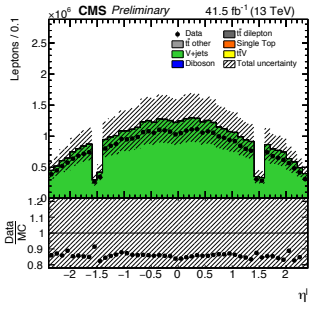


(l) Vertex multiplicity

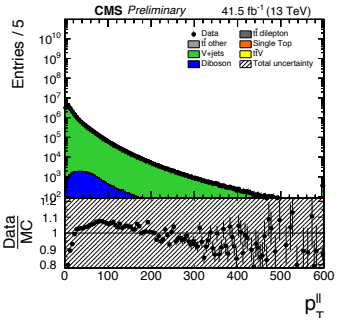
Figure A.9.: Control distributions for the *ee* channel with all cuts prior to Z-window applied



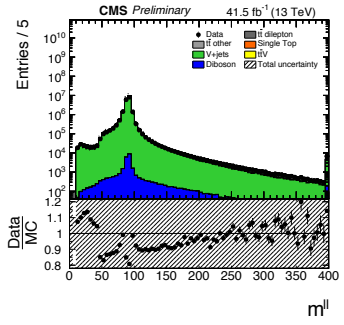
(a) Transverse momentum of leptons



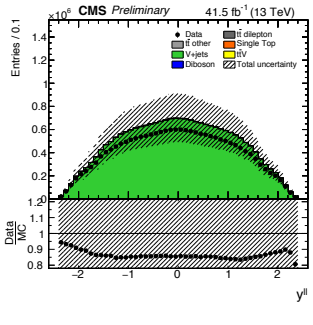
(b) Pseudorapidity of leptons



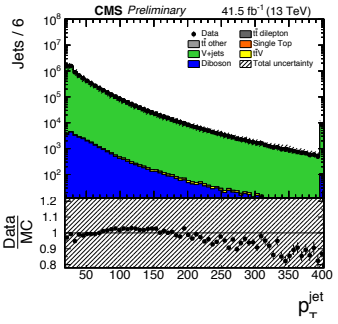
(c) Transverse momentum of dileptons



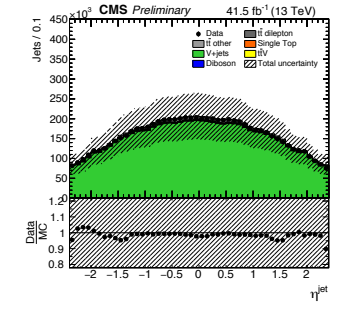
(d) Mass of dileptons



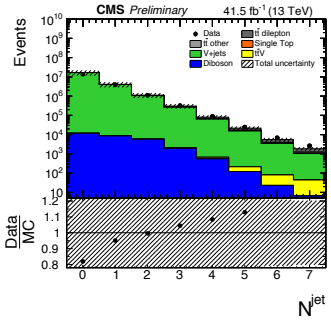
(e) Rapidity of dileptons



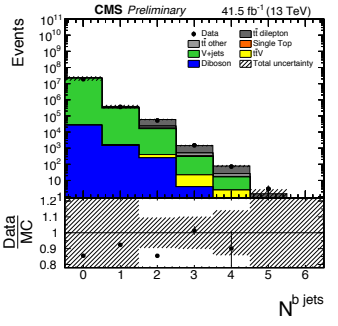
(f) Transverse momentum of jets



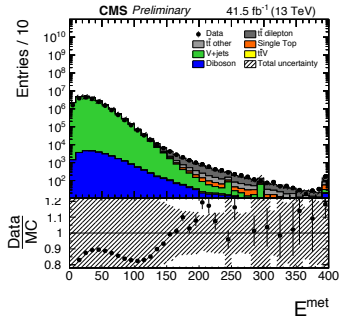
(g) Pseudorapidity of jets



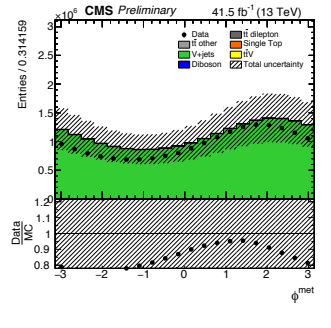
(h) Jet multiplicity



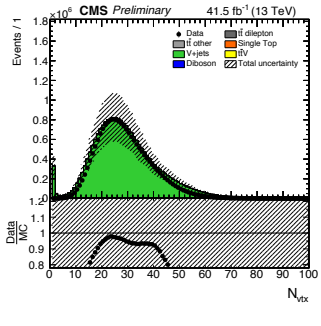
(i) B-tagged jet multiplicity



(j) Missing E_{T}

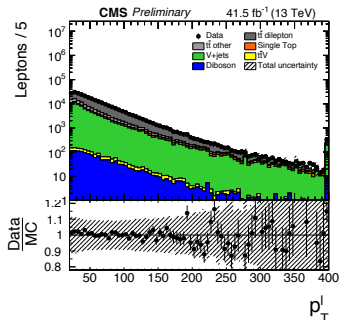


(k) Missing E_{T} azimuthal angle

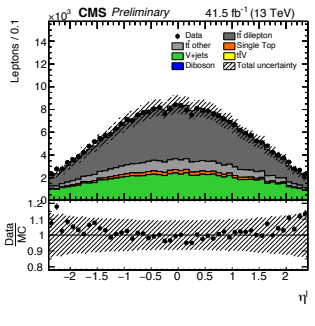


(l) Vertex multiplicity

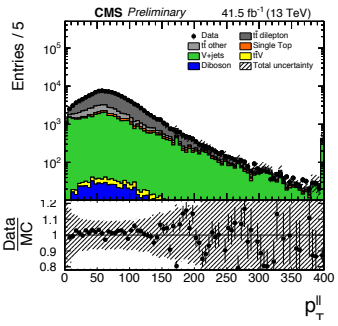
Figure A.10.: Control distributions for the *ee* channel with all cuts prior to dilepton mass applied



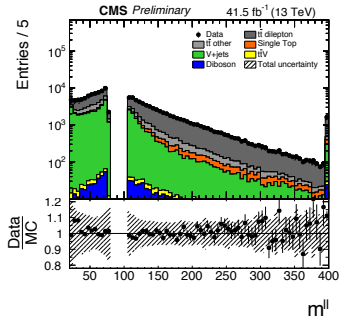
(a) Transverse momentum of leptons



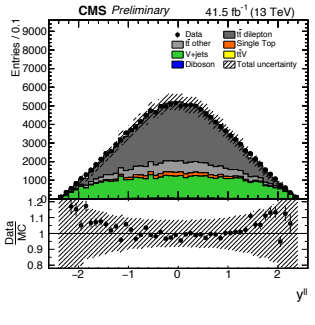
(b) Pseudorapidity of leptons



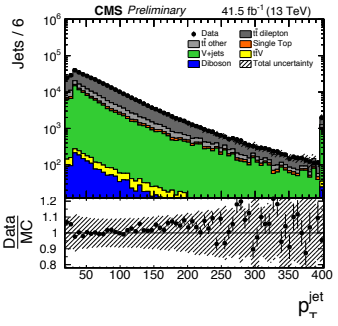
(c) Transverse momentum of dileptons



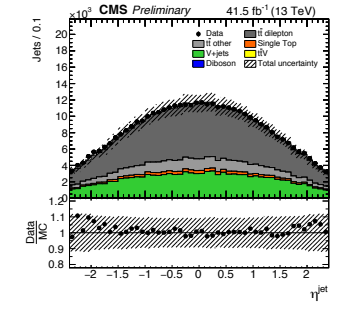
(d) Mass of dileptons



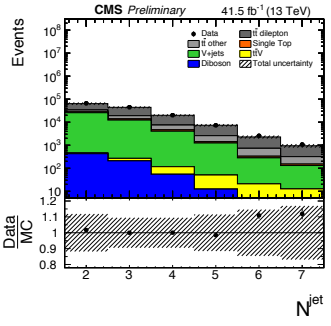
(e) Rapidity of dileptons



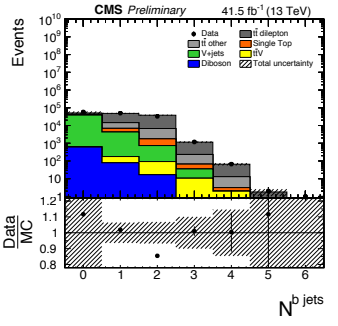
(f) Transverse momentum of jets



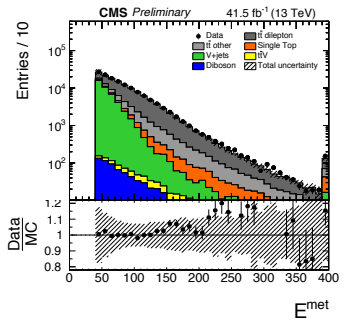
(g) Pseudorapidity of jets



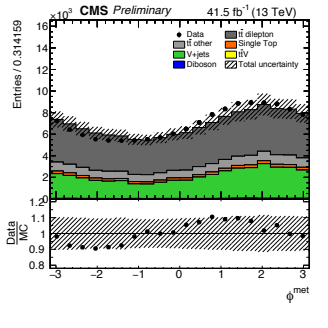
(h) Jet multiplicity



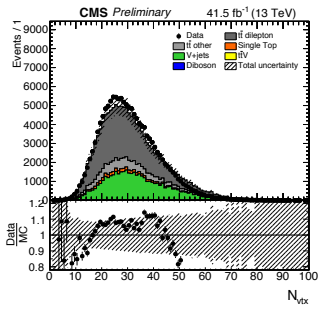
(i) B-tagged jet multiplicity



(j) Missing E_T

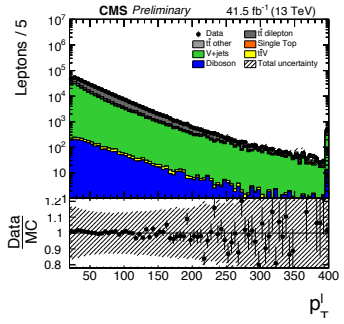


(k) Missing E_T azimuthal angle

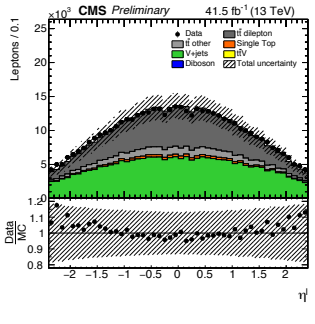


(l) Vertex multiplicity

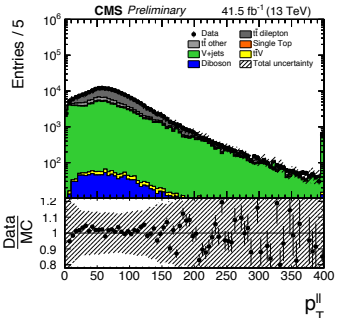
Figure A.11.: Control distributions for the $\mu\mu$ channel with all cuts prior to b-tagging applied



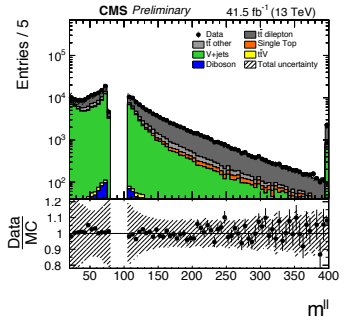
(a) Transverse momentum of leptons



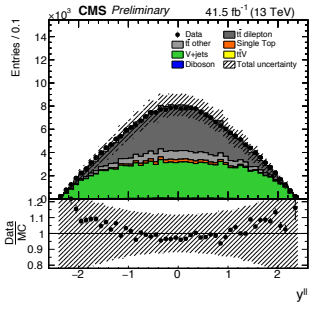
(b) Pseudorapidity of leptons



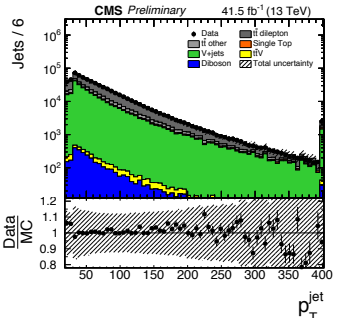
(c) Transverse momentum of dileptons



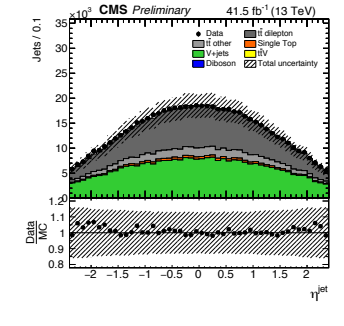
(d) Mass of dileptons



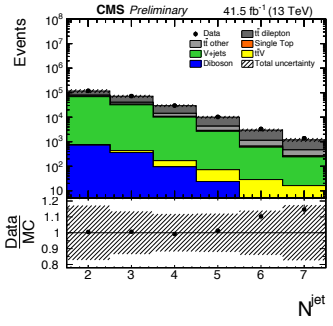
(e) Rapidity of dileptons



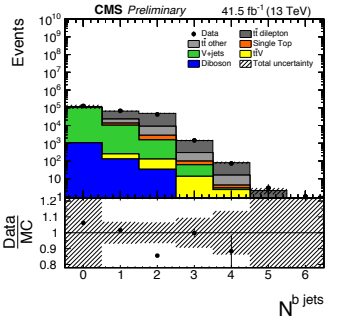
(f) Transverse momentum of jets



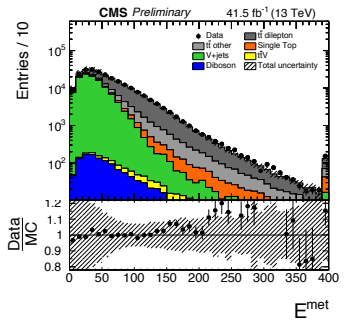
(g) Pseudorapidity of jets



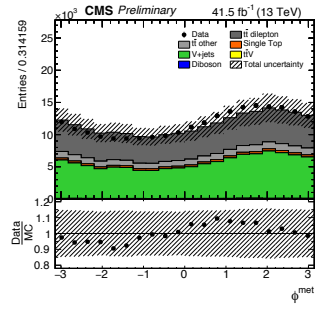
(h) Jet multiplicity



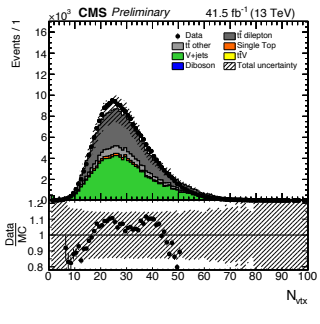
(i) B-tagged jet multiplicity



(j) Missing E_{T}

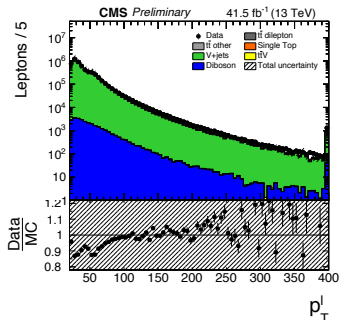


(k) Missing E_{T} azimuthal angle

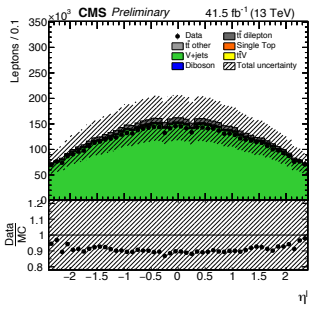


(l) Vertex multiplicity

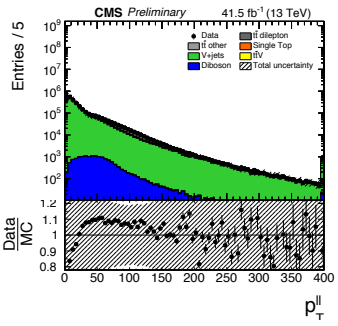
Figure A.12.: Control distributions for the $\mu\mu$ channel with all cuts prior to MET applied



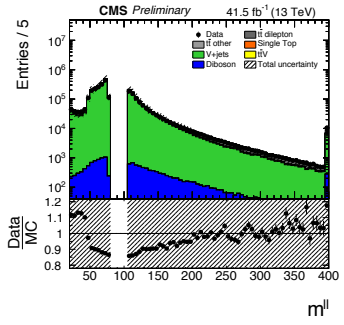
(a) Transverse momentum of leptons



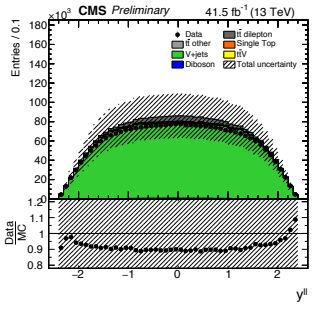
(b) Pseudorapidity of leptons



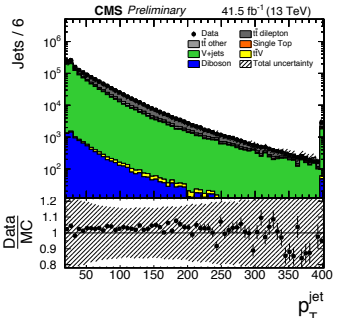
(c) Transverse momentum of dileptons



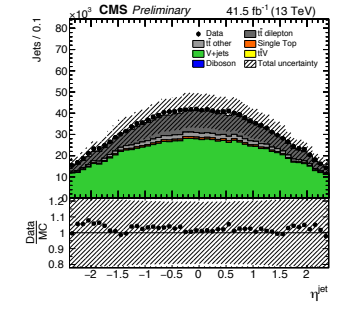
(d) Mass of dileptons



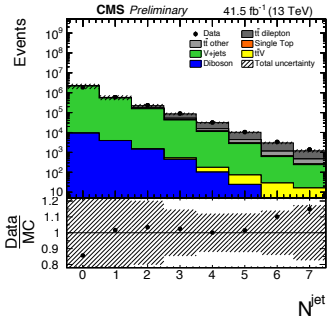
(e) Rapidity of dileptons



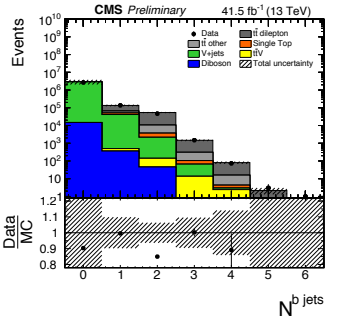
(f) Transverse momentum of jets



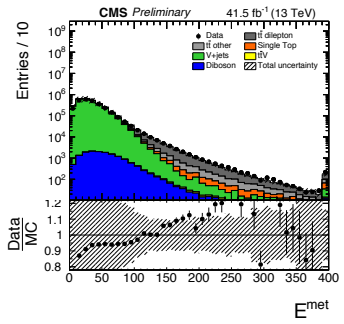
(g) Pseudorapidity of jets



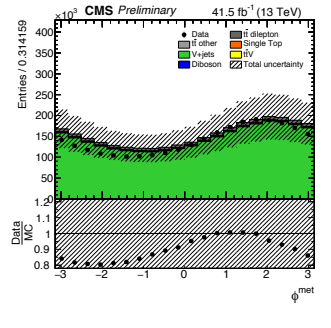
(h) Jet multiplicity



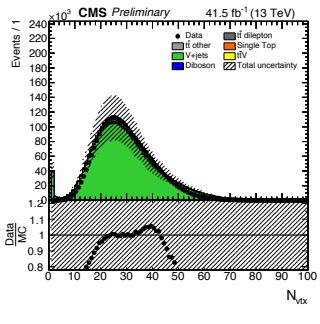
(i) B-tagged jet multiplicity



(j) Missing E_{T}

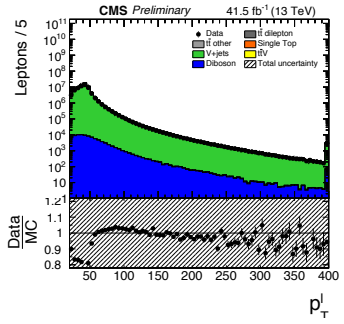


(k) Missing E_{T} azimuthal angle

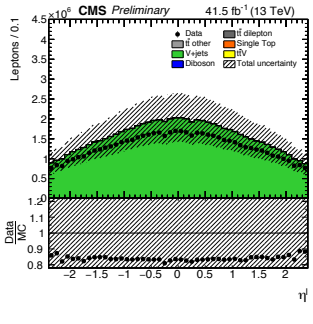


(l) Vertex multiplicity

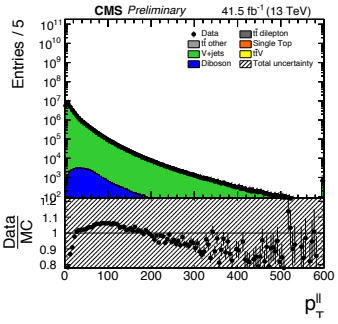
Figure A.13.: Control distributions for the $\mu\mu$ channel with all cuts prior to jet selection applied



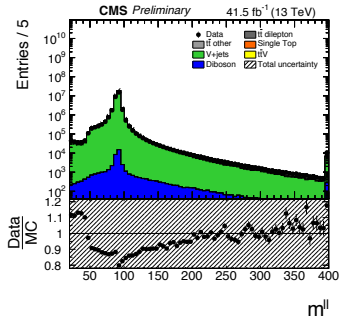
(a) Transverse momentum of leptons



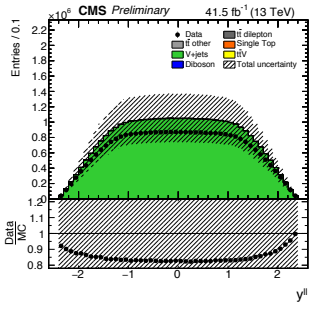
(b) Pseudorapidity of leptons



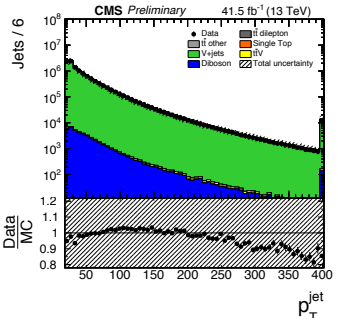
(c) Transverse momentum of dileptons



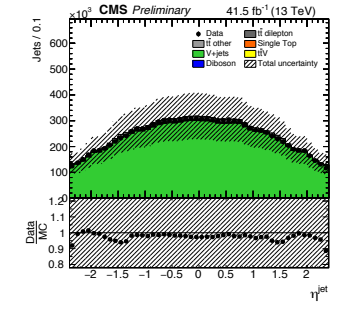
(d) Mass of dileptons



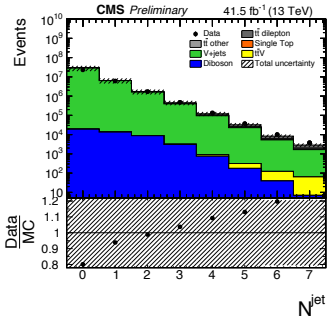
(e) Rapidity of dileptons



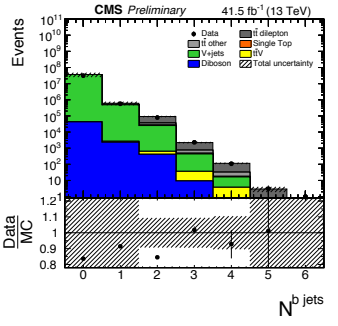
(f) Transverse momentum of jets



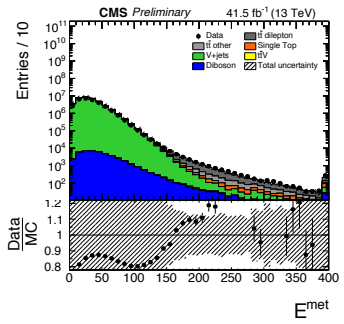
(g) Pseudorapidity of jets



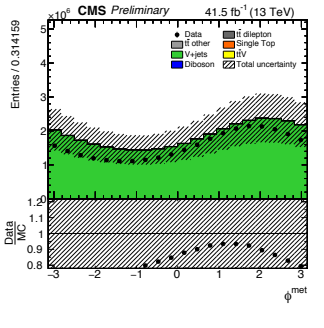
(h) Jet multiplicity



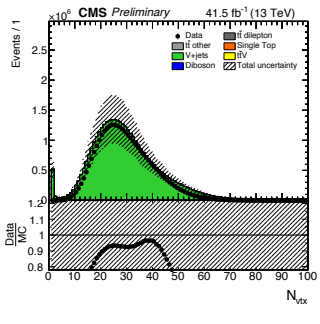
(i) B-tagged jet multiplicity



(j) Missing E_{T}

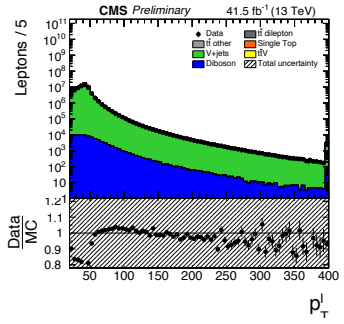


(k) Missing E_{T} azimuthal angle

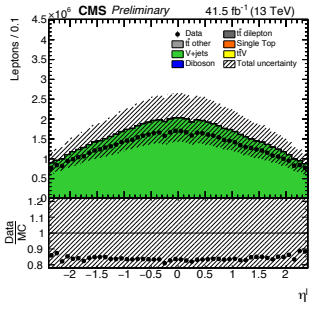


(l) Vertex multiplicity

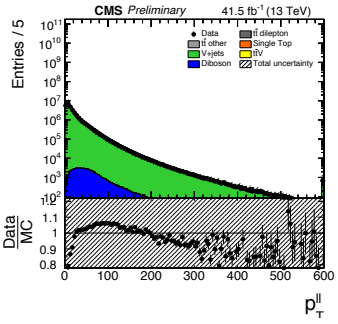
Figure A.14.: Control distributions for the $\mu\mu$ channel with all cuts prior to Z-window applied



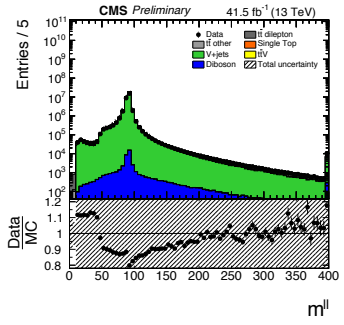
(a) Transverse momentum of leptons



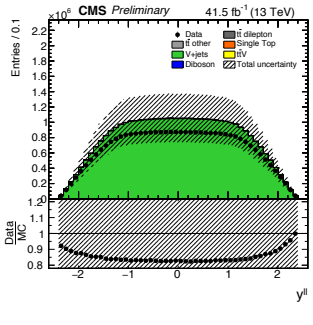
(b) Pseudorapidity of leptons



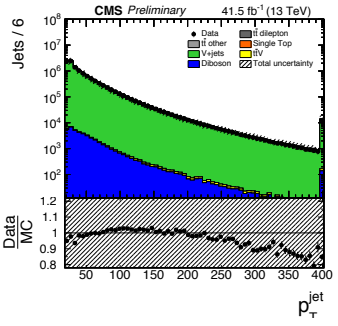
(c) Transverse momentum of dileptons



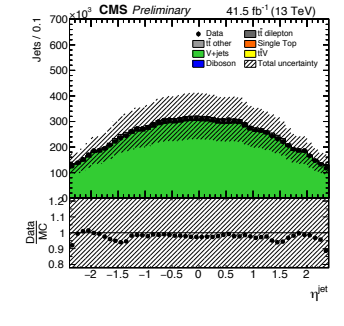
(d) Mass of dileptons



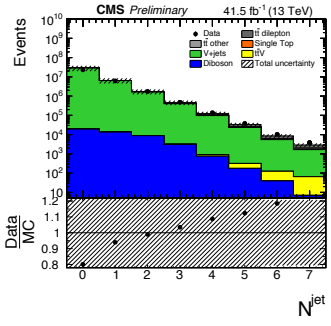
(e) Rapidity of dileptons



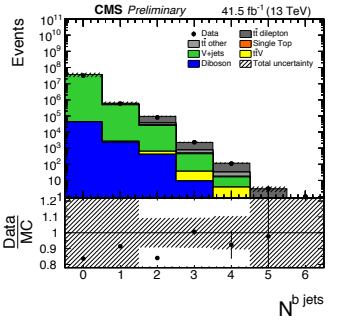
(f) Transverse momentum of jets



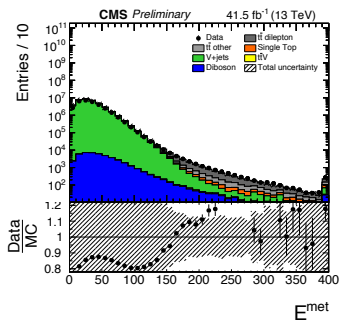
(g) Pseudorapidity of jets



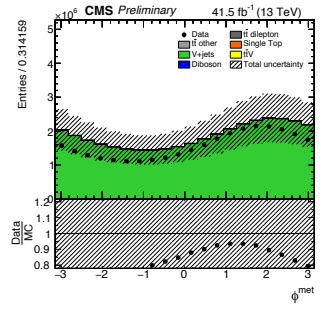
(h) Jet multiplicity



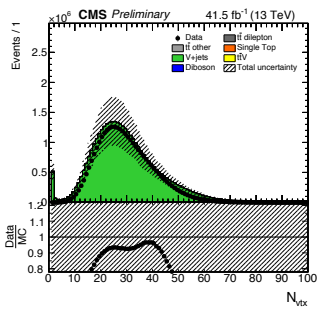
(i) B-tagged jet multiplicity



(j) Missing E_{T}

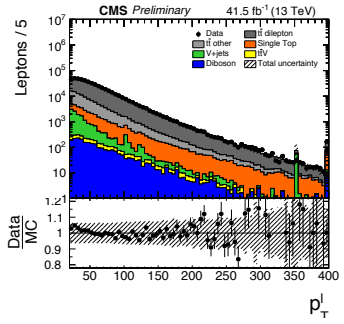


(k) Missing E_{T} azimuthal angle

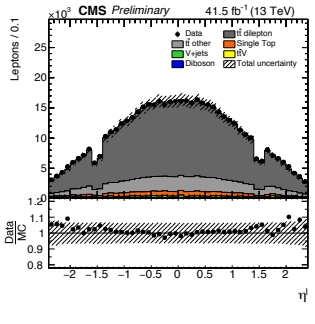


(l) Vertex multiplicity

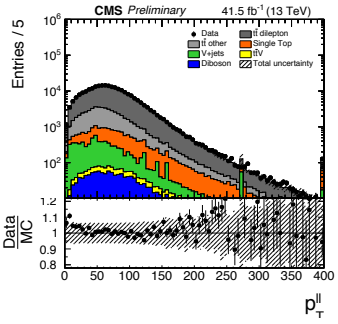
Figure A.15.: Control distributions for the $\mu\mu$ channel with all cuts prior to dilepton mass applied



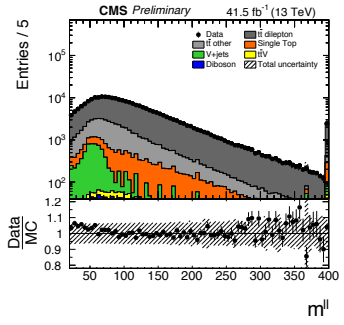
(a) Transverse momentum of leptons



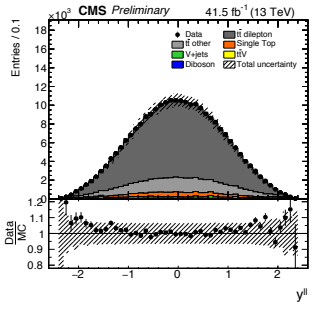
(b) Pseudorapidity of leptons



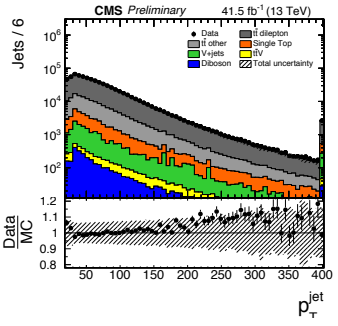
(c) Transverse momentum of dileptons



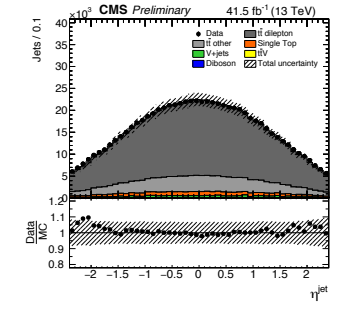
(d) Mass of dileptons



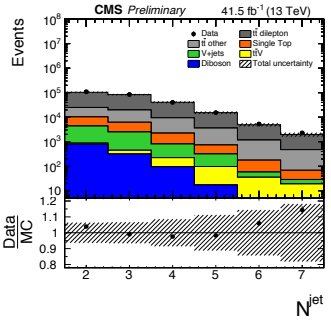
(e) Rapidity of dileptons



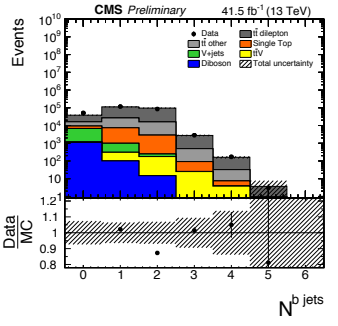
(f) Transverse momentum of jets



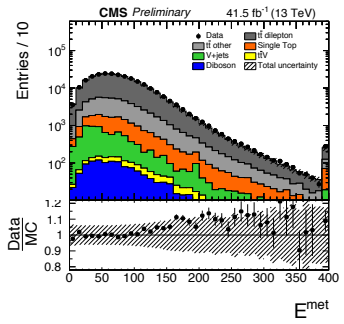
(g) Pseudorapidity of jets



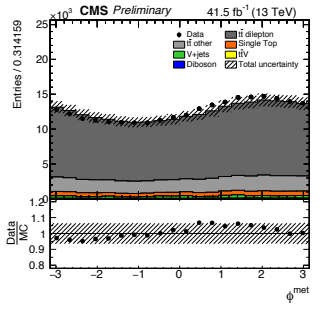
(h) Jet multiplicity



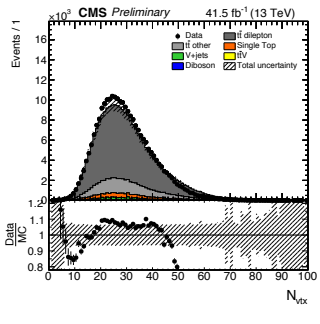
(i) B-tagged jet multiplicity



(j) Missing E_T

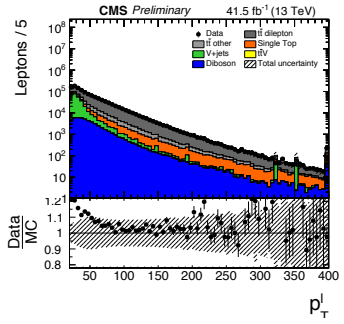


(k) Missing E_T azimuthal angle

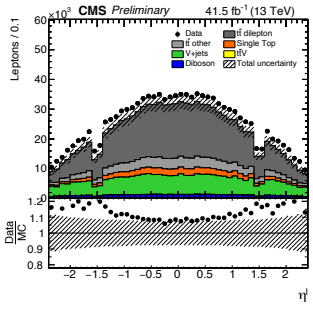


(l) Vertex multiplicity

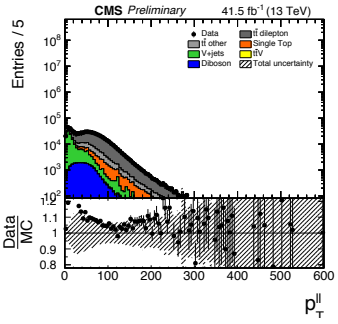
Figure A.16.: Control distributions for the $e\mu$ channel with all cuts prior to b-tagging applied



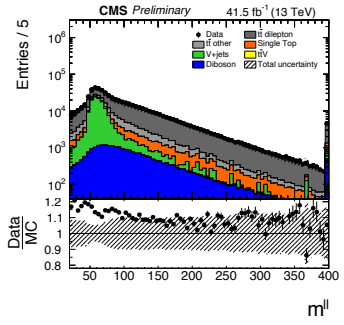
(a) Transverse momentum of leptons



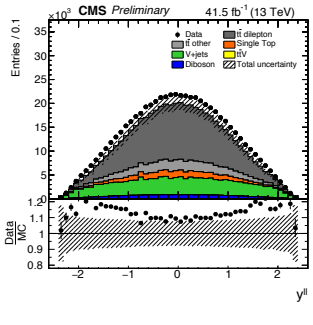
(b) Pseudorapidity of leptons



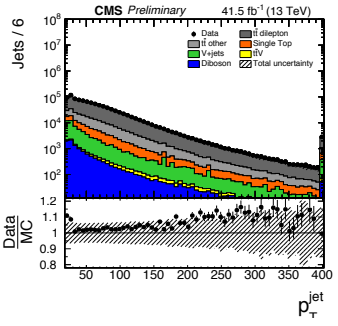
(c) Transverse momentum of dileptons



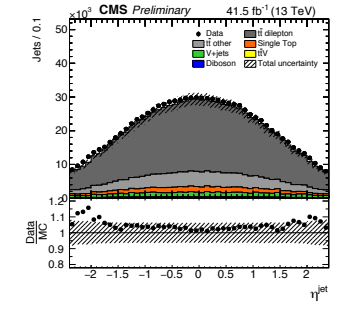
(d) Mass of dileptons



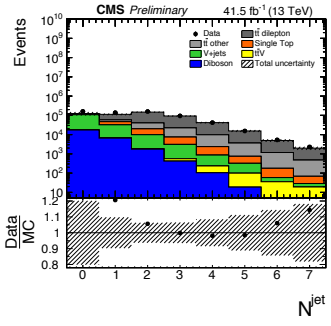
(e) Rapidity of dileptons



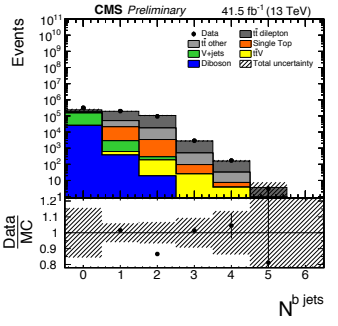
(f) Transverse momentum of jets



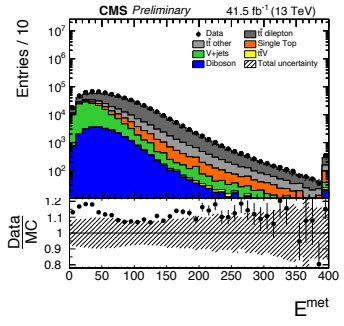
(g) Pseudorapidity of jets



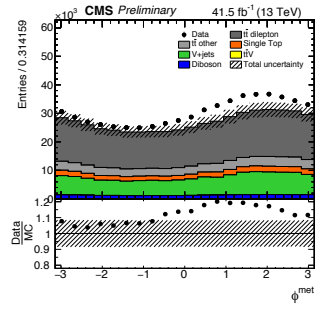
(h) Jet multiplicity



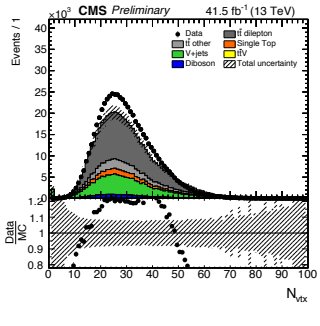
(i) B-tagged jet multiplicity



(j) Missing E_T

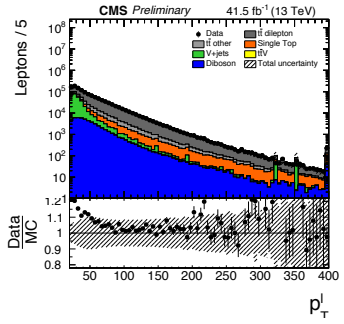


(k) Missing E_T azimuthal angle

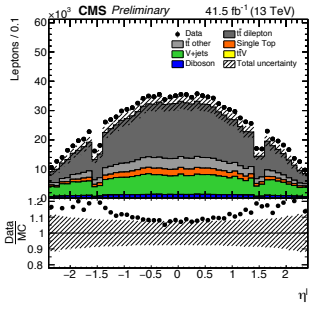


(l) Vertex multiplicity

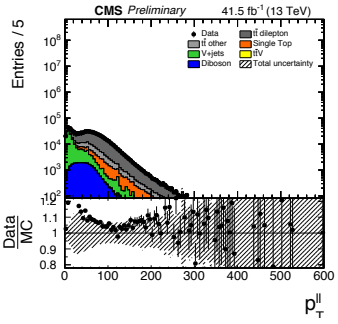
Figure A.17.: Control distributions for the $e\mu$ channel with all cuts prior to jet selection applied



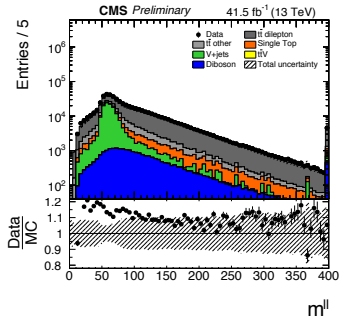
(a) Transverse momentum of leptons



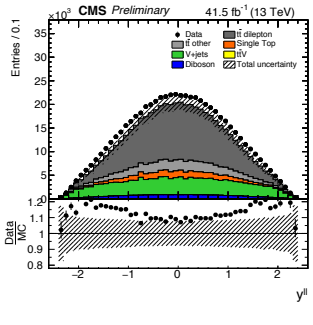
(b) Pseudorapidity of leptons



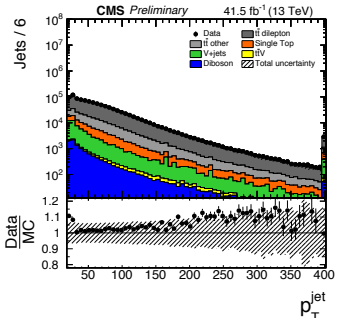
(c) Transverse momentum of dileptons



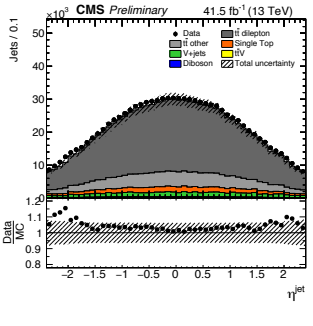
(d) Mass of dileptons



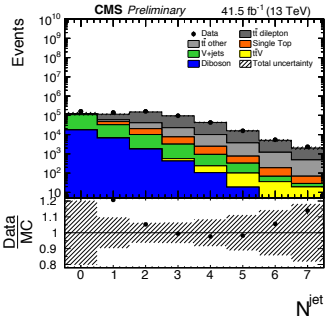
(e) Rapidity of dileptons



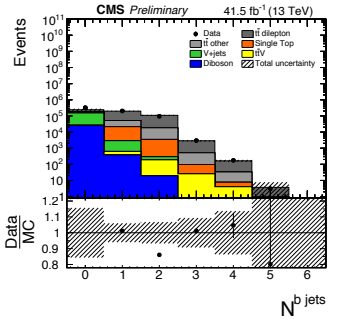
(f) Transverse momentum of jets



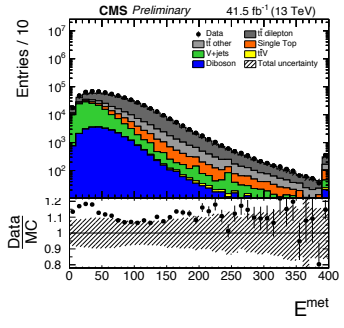
(g) Pseudorapidity of jets



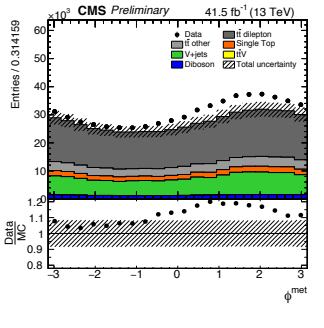
(h) Jet multiplicity



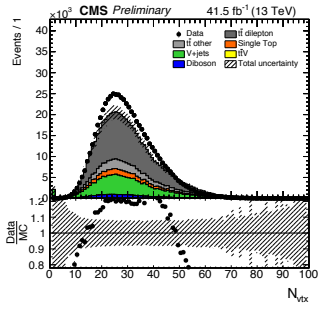
(i) B-tagged jet multiplicity



(j) Missing E_{T}



(k) Missing E_{T} azimuthal angle



(l) Vertex multiplicity

Figure A.18.: Control distributions for the $e\mu$ channel with all cuts prior to dilepton mass applied