ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ

Μέτρηση της διαφορικής ενεργού διατομή του ζεύγους Top Anti-Top quark με τον ανιχνευτή CMS στο LHC

Διδακτορική διατριβή
του
Ιωάννη Παπακριβόπουλου

Επιβλέπων:
Γεώργιος Τσιπολίτης
Καθηγητής, ΕΜΠ

Συνεπιβλέπων:
Κωνσταντίνος Κουσουρής
Αναπληρωτής Καθηγητής, ΕΜΠ

Αθήνα, Σεπτέμβριος 2023

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ

Μέτρηση της διαφορικής ενεργού διατομή του ζεύγους Top Anti-Top quark με τον ανιχνευτή CMS στο LHC

Διδακτορική διατριβή
του
Ιωάννη Παπακριβόπουλου

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΚΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:

1.....Γεώργιος Τσιπολίτης, Καθ. ΕΜΠ
2.....Κωσταντίνος Κουσουρής, Αν. Καθ. ΕΜΠ
3.....Κωσταντίνος Φουντάς, Καθ. Παν. Ιωαννίνων

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

1.....Γεώργιος Τσιπολίτης, Καθ. ΕΜΠ
2.....Κωσταντίνος Κουσουρής, Αν. Καθ. ΕΜΠ
3.....Κωσταντίνος Φουντάς, Καθ. Παν. Ιωαννίνων
4.....Γεωργία Καραποστόλη, Επικ. Καθ. ΕΜΠ
5.....Αλέξανδρος Κεχαγιάς, Καθ. ΕΜΠ
6.....Νικόλαος Μαυρόματος, Καθ. ΕΜΠ
7.....Χρήστος Μάρκου, Ερευνητής Α, ΕΚΕΦΕ ¨Δημοκριτος¨

Αθήνα, Σεπτέμβριος 2023

National Technical University of Athens

School of Applied Mathematics and Physical Sciences

Department of Physics

Measurement of the differential cross section for the top anti-top quark pair using the CMS detector at the LHC

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Ioannis Papakrivopoulos

Supervisor:
Georgios Tsipolitis
Professor, NTUA

Co-Supervisor:
Konstantinos Kousouris
Associate Professor, NTUA

Athens, September 2023

National Technical University of Athens

School of Applied Mathematics and Physical Sciences

Department of Physics

Measurement of the differential cross section for the top anti-top quark pair using the CMS detector at the LHC

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Ioannis Papakrivopoulos

ADVISORY COMMITTEE:

1.....Georgios Tsipolitis Prof. NTUA
2.....Konstantinos Kousouris, Assoc. Prof. NTUA
3.....Konstantinos Fountas, Prof, Univ. of Ioannina

EXAMINATION COMMITTEE:

1.....Georgios Tsipolitis, Prof. NTUA
2.....Konstantinos Kousouris, Associate Prof. NTUA
3.....Konstantinos Fountas, Prof. Univ. of Ioannina
4.....Georgia Karapostoli, Assis. Prof. NTUA
5.....Alexandros Kehagias, Prof. NTUA
6.....Nikolaos Mavromatos, Prof. NTUA
7.....Christos Markou, Researcher N.C.S.R 'Demokritos'

Athens, September 2023

Στην Κατερίνα και στη Λυδία

Abstract

This thesis presents the work done for the measurement of the differential cross section of a pair of high transverse momentum (boosted) top/anti-top quarks that decay hadronically. To conduct the measurement the data collected by CMS experiment during the "Run-II" data taking period of the LHC are used. They correspond to an integrated luminosity of $137.1 fb^{-1}$. The data correspond to collision at a center of mass energy of 13 TeV. This analysis is a precision measurement. The deliverable has already been measured before both by CMS and ATLAS. In this thesis by modifying parts of the measurement strategy and the tools used we aim to achieve better precision. This is also enhanced by the use of almost 4 times more data than the previous measurement performed by the CMS collaboration. Another nice addition to the study is in the section of the comparison with the theory. Previous measurements saw a difference between the measured values and the theoretical predictions. In this analysis we aim to better quantify this difference by including systematic uncertainties for the various theoretical models to better estimate the level of disagreement. In addition chi square values and p-values are calculated to better estimate it. During this thesis there was also extensive work done in the detector control system of the CMS detector as a service work to the collaboration. The work focused around the maintenance and the improvement of the system in order to ensure a better operation in the "Run-II" data taking period but also in the next ones to come. This work affected not only this particular analysis but the collaboration in general as it enabled a smooth operation and optimal performance of the detector enabling the acquisition of good quality of data.

Περίληψη

Η παρούσα διατριβή είναι αποτέλεσμα της δουλειάς που έγινε για την μέτρηση της διαφορικής ενεργού διατομή του ζεύγους τοπ/αντί-τοπ κουάρκ μεγάλης εγκάρσιας ορμής τα οποία διασπούνται αδρονικά. Για την πραγματοποίηση της χρησιμοποιήθηκαν δεδομένα τα οποία συλλέχθηκαν από τον ανιχνευτή CMS κατά τη διάρκεια του "Run-II" του LHC.Τα δεδομένα αυτά αντιστοιχούν σε φωτεινότητα $137.1 fb^{-1}$ και για να παραχθούν έγινα συγκρούσεις με ενέργεια κέντρου μάζας 13 TeV. Η παρούσα μέτρηση είναι μια μέτρηση ακρίβειας. Στο παρελθόν τα έχει ξαναπραγματοποιηθέι τόσο από το πείραμα CMS όσο και απλό το πείραμα ATLAS. Στην παρόυσα εργασία έγινε προσπάθεια μέτρησης με μεγαλύτερη ακρίβεια τροποποπιώντας τη μέθοδο που είχε χρησιμοποιήσει παλαιότερα το πείραμα CMS αλλά και χρησιμοποιώντας κάποια διαφορετικά εργαλέια σε συγκεκριμένα σημεία. Ένα ακόμα στοιχείο το οποί βοηθάει στην επίτευξη μεγαλύτερης ακρίβειας είναι οτι χρησιμοποιήθηκαν σχεδόν τετραπλάσια δεδομένα σε σχέση με την προηγούμεη μέτρηση του CMS. Μία ακόμα προσθήκη στη μελέτη αυτή είναι στο κομμάτι της σύγκρισης με τη θεωρία. Οι προηγούμενες μετρήσης και από τα δύο ξεχωριστά πειράμα είχαν παρατηρήσει μια ασυμφωνία μεταξύ των δεδομένων και των θεωρητικών προβλέψεων. Στη παρούσα ανάλυση γίνεται μια προσπάθεια ποσοτικοποίησης αυτής της ασυμφωνίας. Αυτό γίνεται σε πρώτο στάδιο με την χρήση συστηματικών σφαλμάτων στα διάφορα θεωρητικά μοντέλα. Στη συνέχεια για τον σκοπό αυτό, υπολογίζονται οι τιμές του χ^2 αλλά και τα p-values. Κατα τη διάρκεια της διατριβής αυτής, έγινε και εκτεταμένη δουλειά στο σύστημα αυτομάτου ελέγχου του ανιχνευτή CMS. Η δουλειά αυτή, έγινε σαν service work στο CMS collaboration. Η δουλειά αυτή περιστράφηκε κυρίως γύρω από την συντήρηση και βελτίωση του συστήματος αυτού με σκοπό την καλύτερη λειτουργία του ανιχνευτή κατά τη διάρκεια του "Run-II". Η δουλεία αυτή δεν επηρέασε μόνο την παρούσα διατριβή αλλά ολόκληρω το πείραμα καθώς επέτρεψε την ομαλότερη λειτουργία του ανιχνευτή. Σαν αποτέλεσμα ο ανιχνευτής λειτούργησε με τιν βέλτιστες δυνατές συνθήκες επιτρέποντας έτσι τη συλλογή υψηλής δεδομένων ποιότητας.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Physics aims to study and understand nature in its very core. It tries to find what it is made of how it behaves and what rules govern it. Particle physics is the field of physics that tries to explain the very essence of the world. It does that by studying particles, the building blocks of nature. It attempts to answer questions like which are these building blocks, how do they behave and why. The ultimate goal is to find a theory, a set of rules, that can explain everything that can be observed in the universe and everything that will be observed.

Every field in physics has two main "branches" that work together and against each other at the same time, theory and experiment. They work together in the sense that they try to explain the same thing. On the other hand, they work against each other because they challenge one another in order to progress the knowledge and understanding in that particular field. Advancement can happen in two ways. A theory is developed in order to explain a physical phenomena. Then it is the experiment's work to challenge this theory and try to break it. The other alternative is that an experiment finds a new phenomena or effect that is not explained by any theory. Then it is the theory's time to try and explain it.

In their endeavours to explain the universe, particle physicists have developed the Standard Model (SM) of particle physics. It is the most complete and accurate theory that has ever been developed in the field of particle physics. It predicts the existence of a set of particles that are called elementary. The term elementary comes from the fact that they have no substructure. At the same time, it predicts the ways that these particles interact with each other. Using them as a framework the theory tries to explain the "nature" of matter as well as its behavior. That is from the beginning of time till now and the future.

The standard model predicts the existence of many particles that can not be

found free in nature. These are fundamental for the explanation of the universe and the state it is now in the context of the theory. Their absence is true and in accordance with what the theory predicts. However the SM also describes how they can be produced. Accelerators, like the large hadron collider (LHC) have been built to test the theory and what it predicts. It has withstand extensive testing by many experiments including the LHC and it has predicted the results of the experiments with great accuracy. All the particles, their properties as well as their interactions have been detected. The final one was the Higgs boson, whose detection was announced by two of the experiments of the LHC completing the picture that the SM describes.

Despite its success we know that the SM is not perfect. There are phenomena in nature that are not contained in the theory and things whose existence is not predicted by it. One example is the existence of gravity. We know that gravity is a fundamental force of the universe but it is not included in the SM. The matter-antimatter asymmetry. According to the SM, matter and antimatter are produced in equal amounts but we know that in the universe matter has prevailed. How did this happen? Why do neutrinos have mass. Neutrinos are particles predicted by the SM and their existence has been proven. According to the theory they have no mass which we know is not correct. Why do they have mass and what is their mass? Also dark matter and dark energy are two entities that we know that they exists but are not predicted by the SM.

With the above in mind it is clear that the SM as a theory is true but it can not be the final theory that explains everything. Its great success can indicate that it is a subset of a bigger theory. Some scientists are trying to test new theories that are extensions of the SM. Others try to measure things that the SM predicts with bigger accuracy in order to find inconsistencies in the theory and prove it, partially, wrong. The story continues with a never-ending back and forth between theory and experiment until a general unified theory of everything is developed. An answer to the greatest question, who and what are we, what are we made of?

Chapter 2

Theory

Humans were always troubled with the structure of the world. Democritus (430 - 370 B.C.) in ancient Greece was the first one that formulated the atomic theory. His idea was that if you take a piece of matter and cut it in half then take one of the halves and repeat the process at some point you will reach to a piece that can not be cut. He named that part άτομο (atom) which stands for something that can not be cut. The modern concept of atoms was introduced by the chemist John Dalton [1]. He suggested that elements consist of really small particles that he called atoms. He also proposed that atoms of the same element are identical. We now know that this is not the case and that atoms can be "cut" into smaller pieces. But the same philosophy has led the search for the building blocks of the universe. This has resulted in the development and formulation of the SM of particle physics. This chapter will be a theoretical introduction to the SM and the top quark.

2.1 The Standard Model (SM) of particle physics

The SM was formulated in the late 60s and early 70s. It is a Quantum Field Theory (QFT) based on a gauge symmetry. It predicts the existence of a set of elementary particles, i.e. particles with no substructure, see Figure 2.1. Using these particles it describes formation of all visible matter as well as its behavior.

According to the SM the particles are split in two major categories, bosons and fermions. The differentiation is made based on their spin which is an internal property of all the particles. Fermions have a half integer spin, namely 1/2 while bosons have an integer value for spin.

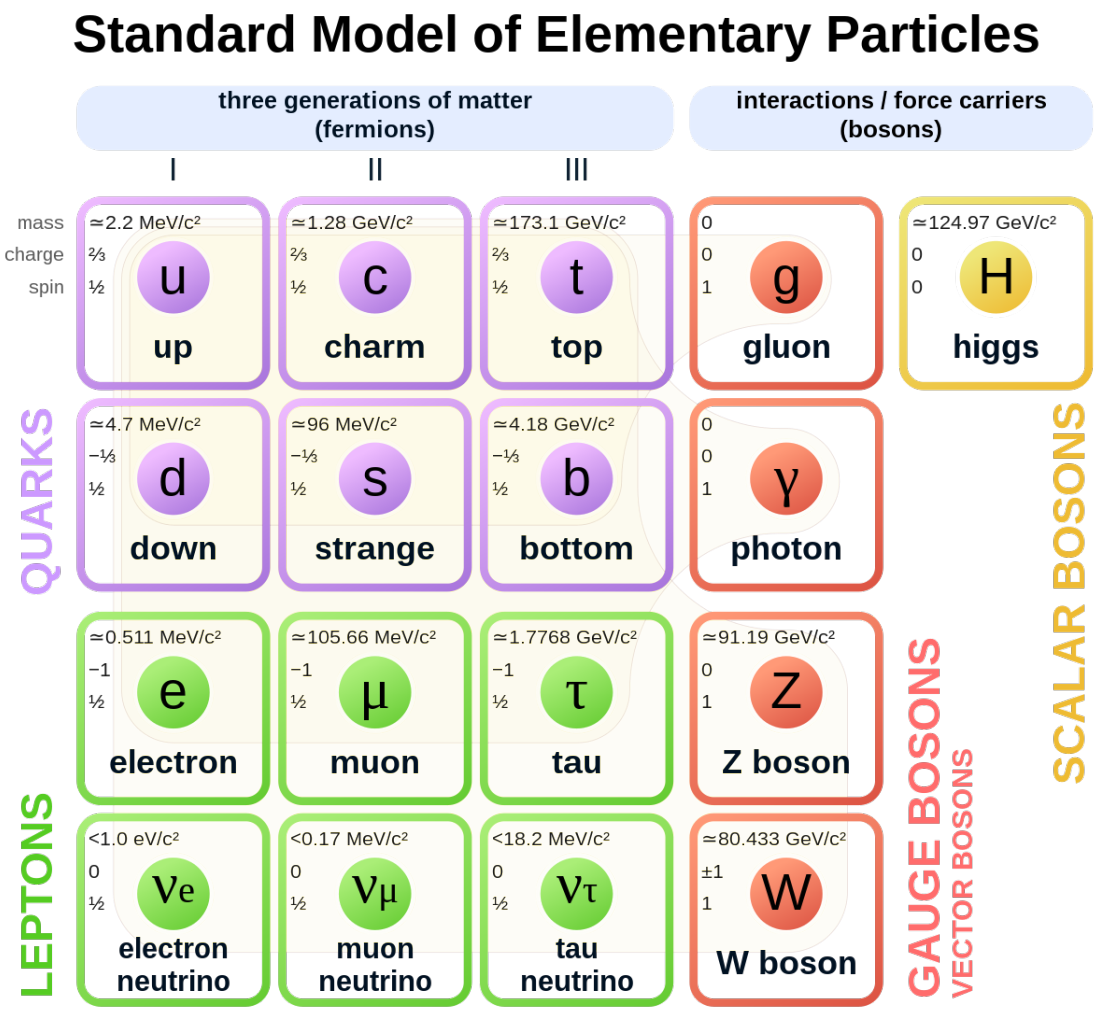

Figure 2.1: The particles included in the Standard Model of particle physics. They are split in two categories fermions and bosons. Fermions are the particles that make up matter, while bosons are the force carriers and are involved in interactions. Fermions are further split in two categories quarks and leptons. Leptons can be found free in nature while quakrs are combined to create more complex particles.

2.1.1 Fermions

Fermions are the particles that make up matter. They are splitted in two categories, leptons and quarks. There are in total 6 quarks (6 flavors), the up (u), down (d), charm (c), strange (s), top (t) and bottom (b). There are also 6 leptons. Three are electrically charged, the electron (e), the muon (μ) and the tau (τ). For every charged lepton there is a neutral charge one called a neutrino ($\nu_e,\ \nu_\mu,\ \nu_\tau$). Using as a measure the electric charge of the electron, we say that the electron has a charge of -1. The rest of the charged leptons have also a charge of -1, while neutrinos have 0. Quarks on the other hand can have a charge of 2/3 for the u, c , d and -1/3 for the d, s and b.

The mathematical formulation of fermions in the standard model (Dirac equation) predicts the existence of particles with negative energy. This has also been proven experimentally. So for every fermion there is also an antiparticle, i.e. a particle having exactly the same properties as the particle but opposite sign charges.

The fermions also carry another type of charge called color. It can have three possible values red green and blue. Leptons are colourless while quarks can have any of the possible colors. This property is associated to the ability of a particle to be free or in a bound state. In nature only colourless states can be found. This means that leptons can be found as free in the nature as they have no colour. On the other hand quarks are only found grouped together forming more complex particles. This happens so that they form colourless states. Quark combinations are usually quark anti-quark, called mesons, or combinations of more quarks called baryons. Until recently only 3 quark baryons had been found. Recently the LHCb exepriment at CERN has announced the discovery of more exotic particles having 4 quarks (tetra-quarks) and even ones with 5, called penta-quarks.

Fermions are also splitted in 3 other categories called generations. Each generation contains 2 quarks, a charge lepton and a neutral lepton, a neutrino. The particles of the same type have increased mass in each category. With the first generation containing the lightest ones and the third the heaviest ones.

2.1.2 Bosons

Bosons are the force carriers of the standard model. In the framework of the SM there are 4 fundamental forces in the universe. Each one of these forces is expresses through a particle, in particular a boson. This means that for every reaction to happen one of these force carrier particles must be involved. The 4 mentioned forces are the strong, the weak, the electromagnetic and gravity. The carrier for the strong force is the gluon (g) which is massless. For the weak

Force	Carrier	Strength
Strong	gluon (g)	1
Electromagnetic	photon (γ)	10^{-2}
Weak	$W^{\pm}$, Z	10^{-13}
Gravity	graviton (G)	10^{-38}

Table 2.1: The standard model forces

force there are three carriers, the W^+, the W^- and the Z^0. The first two are positively and negatively charge respectively while the third is neutral. All three particles for the weak force have a mass. The photon (γ) is the force carrier for the electromagnetic force which is also massless. Finally gravity is expressed via the graviton, which is a particle that has been theoretically predicted but has never been detected. In the context of the particle physics the effect of gravity is so small that can be neglected.

The strong interaction, as the name states, is the strongest of the three forces included in the SM. In Table 2.1 a summary of all the SM forces is presented. The part of the SM that models the strong force and its interactions is called Quantum Chromo Dynamics (QCD). It involves particles carrying a color charge, i.e. the six quarks and the gluon which is the force mediator. The gluon also carries a color charge which means that it can couple to itself. There are eight types of gluons one for each linear independent color state. Its strength is given by what is called the strong coupling constant α_s [2]. It can be seen (Figure 2.2) that as the energy decreases the strong force becomes stronger. This means that as the quarks become further and further apart they are pulled together even stronger. This is what keeps the particles like the proton together. It is also the reason why a bare quark has never been discovered and is called quark confinement. Providing enough energy should make it possible to extract a quark from a composite particle. But what actually happens is that the energy needed to separate the quark is enough so that a quark-antiquark pair would be created out of the vacuum. This would lead in the creation of a new hadron instead of a free quark. The process described is called hadronization and is responsible for the formation of composite particles. It can also be seen that the strong force decreases with distance. This is a property of the strong force that is called asymptotic freedom. This means that perturbative calculations [3] can be used to study the strong force.

Particles with electric charge are involved in electromagnetic interactions. The photon (γ) is the force carrier for this force. The theory that studies this set of interactions is caled Quantum Electro Dynamics (QED). Since the strength is

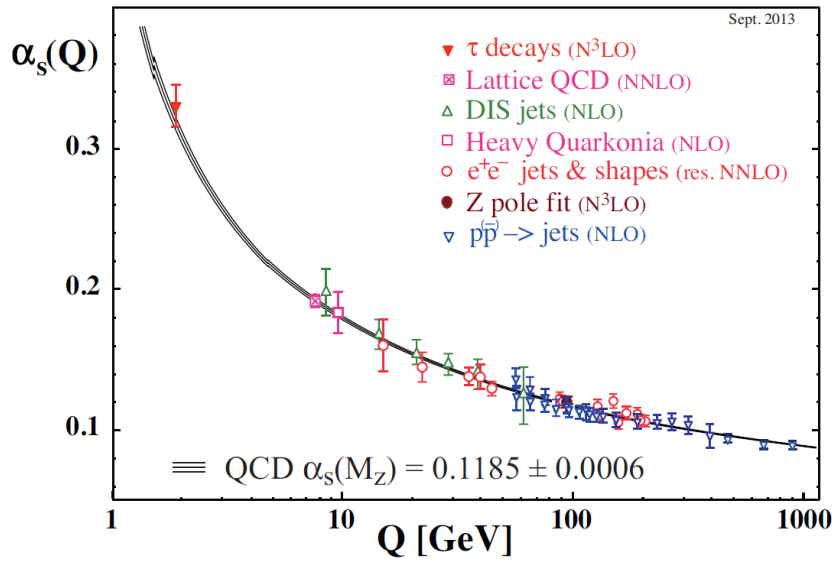

Figure 2.2: The strong coupling constant α_s, figure taken from [2]

already very small it can be studied as a perturbative theory. The force carrier of electromagnetic interactions has no charge, it is not involved in the interaction. This makes the calculations for this theory way less complex than QCD.

The weak force is the last force described in the SM. As its name suggests it is the weakest of the forces in the model. It is also the only one that involves all the fermions. The property called weak isospin (I or T) is the one that is associated to the weak force. This is done in a similar manner as color and charge are associated with the strong and electromagnetic force respectively. The weak force is able to change the quark flavor (flavor-mixing) or an electron to a muon. Due to this property, physical phenomena like radioactivity and muon decay are explained via the weak force. The theory for the weak force is a chiral theory. Chirality is an internal property of all particles that distinguishes particles in left and right handed [4]. Beeing a chiral theory means that the weak interaction treats particles differently based on the value of their chirality. Only left handed particles and right handed anti-particles carry the weak charge and hence can interact via the weak force.

The weak force mediators are the $W^{\pm}$ and the Z. Depending on the mediator and if it carries charge, we can have a charged or a neutral current. Only charged currents have been detected to change flavor. The probabilities of flavor changing

are described in the Cabbibo-Kobayashi-Maskawa (CKM) matrix [5], [6].

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.1}$$

Where $|W_{ij}|^2$ is the probability for the transition from flavor i to flavor j. Diagonal elements has significant bigger values than the off-diagonal ones. This means that it is more likely for a transitions of a quark to the quark of the same generation.

The feynman diagrams of the weak and electromagnetic force have many similarities. In fact in the late 60s a new theory was developed, the electroweak theory. It proved that in high energies it was possible to unify the two forces in to one. In the current state of the universe not enough energy is available which is why the force is now split in to two separate forces.

2.2 The Top quark

The top quark is the heaviest of the quarks in the standard model Figure 2.3 with a mass of 172.69 ± 0.30 [7]. It was discovered in 1995 from the CDF and D0 experiments in the Tevatron accelerator at Fermilab [8], [9]. It was the last of the quarks to be discovered and since then no more quarks have been detected. With such a high mass it has a very small lifetime, approximately 5×10^{-25} s. This value is smaller than the hadronization time 10^{-24} s meaning that it decays before it can form hadrons. Although its lifetime is still very small the fact that it decays before forming a hadron allows to study the properties of a bare quark even indirectly by studying the properties of the decay. Additionally due to its high mass it has a strong coupling to the Higgs boson which means that it plays an important role in the Higgs mechanism. Finally, the top quark is a background to many of the processes predicted by extensions of the standard model. This means that it is generated together with other new hypothesized particles in these processes. Having predicted its properties with very good precision will allow scientists to remove the top quark contributions in their measurements and be left with only the required signal if there is any.

2.2.1 Top quark production

A top quark can be produced via the electroweak interaction together with a b quark and a W boson. The leading order (LO) Feynman diagrams of the possible

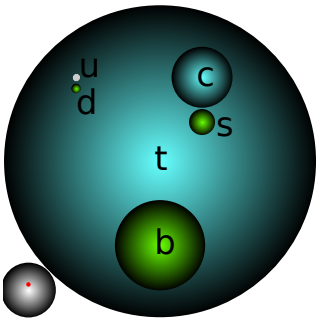

Figure 2.3: A schematic representation of the mass hierarchy for the quarks of the standard model.

interactions can be seen in Fig 2.4. Another possibility is the production of a top (t), anti-top ($\bar{t}$) pair ($t\bar{t}$) via the strong interaction. This can happen either with the interaction of two gluons known as gluon fusion or via the annihilation of two quarks. The leading order Feynman diagrams of the two processes can be seen in figures 2.5 and 2.6 respectively.

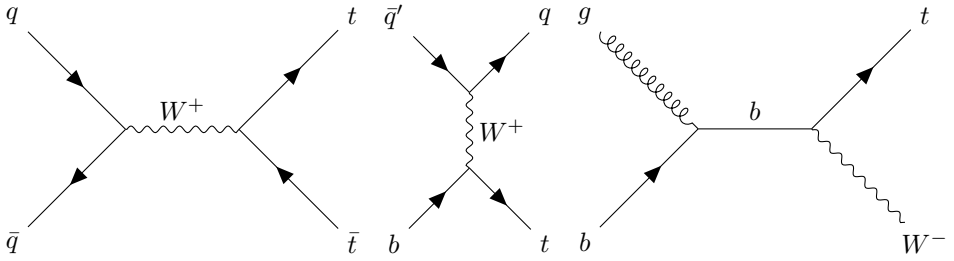

Figure 2.4: Leading order Feynman diagrams for the production of a single top quark.

In particle accelerators, the production of top quarks is done mainly with the $t\bar{t}$ process via the strong interaction. In the Tevatron accelerator where the top quark was discovered, the experiments were conducted with the collision of protons and anti-protons. The production was mainly done with the annihilation process of a quark and an anti-quark. In the LHC due to higher energies the production is done predominantly via the gluon fusion process at a percentage of around 90%.

The difference in the way of $t\bar{t}$ production in the two experiments can be explained from the diagram in Figure 2.7. The graphs show the parton distribution

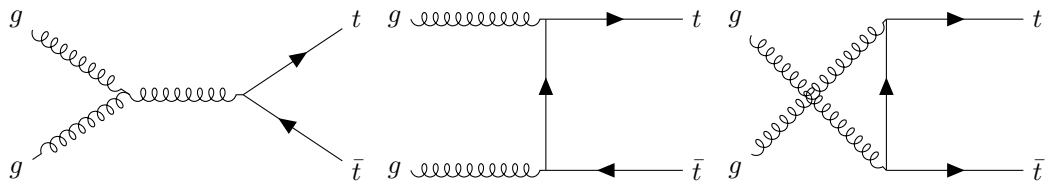

Figure 2.5: Leading order Feynman diagrams for the production of a $t\bar{t}$ pair via the gluon fusion process.

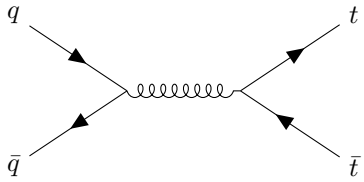

Figure 2.6: Leading order Feynman diagrams for the production of a $t\bar{t}$ pair via the quark annihilation process.

function (PDF) of a proton versus the momentum fraction χ. A PDF is the density of finding a parton carrying a momentum fraction (χ) at a certain energy scale Q^2 $(= -q^2)$. Where χ is the fraction of the nucleon longitudinal momentum carried by the parton. In order to produce a $t\bar{t}$ pair during a collision, the center of mass energy needs to be at least two times the mass of the top quark, $\chi \gtrsim \frac{2m_t}{\sqrt{s}}$. In Tevarton the value the value for χ was $\chi \approx 0.18$ while in the LHC it is $\chi \approx 0.027$. From Figure 2.7 it can be seen that the PDF values for a gluon are way bigger for the χ values at the LHC hence gluon fusion is he dominant process.

2.2.2 Top decay

The top quark as mentioned before decays very quickly due to its high mass. The decay happens via the electroweak force as it is the only one that can change the flavor of a quark. With the emission of a W boson, the top quark can decay into a d, s or b quark. The probability of decaying into each of the quarks is given by the square of the respective element of the CKM matrix (see Equation 2.1) $|V_{td}|^2$, $|V_{ts}|^2$ and $|V_{tb}|^2$. From the values of the corresponding we can see that it decays almost exclusively to a b quark with a probability of 99.8%.

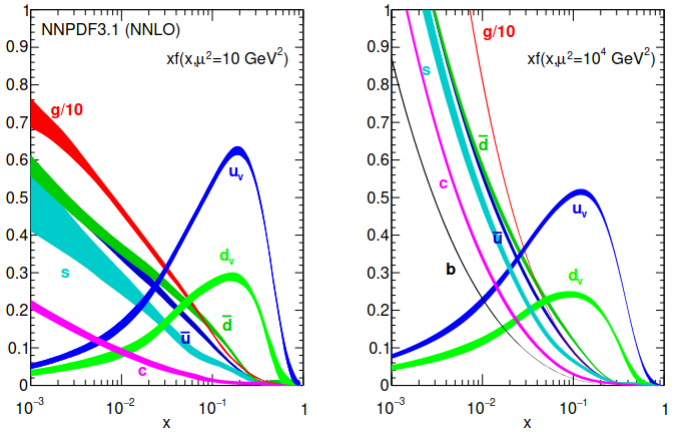

Figure 2.7: The two plots show the parton distribution function (PDF) for the proton versus the momentum fraction χ. On the left there is the plot for the energy present at Tevatron accelerator while on the right there is the one for the LHC.

The W boson produced in the top decay will also before particle detection. There are two decay categories based on the nature of the decay products, leptonically or hadronically. In the leptonic way, it decays into a lepton and its corresponding neutrino while in the hadronic way it decays into to a pair of a quark and an anti-quark. For the leptons all three are possible with equal percentages. For the hadronic way, all quarks are possible except of the top due to energy conservation. Again the probability is determined by the respective CKM elements. The percentages for all possible decay can be seen in Figure 2.8.

Taking into account that in the $t\bar{t}$ pair we have two tops decaying there will be two W bosons. In this case we can have any of the following scenarios:

- **Fully hadronic channel**[1]. In this channel both the W bosons decay into to a pair of quarks. This channel is the most probable out the three with a probability of 45.4%. This is the channel that is used in this physics analysis.

- **Lepton + jets channel**. In this process, one of the bosons decays hadronically while the other one decays leptonically. This process also has a high probability, 44.1%

[1]In particle physics terminology a channel is a process

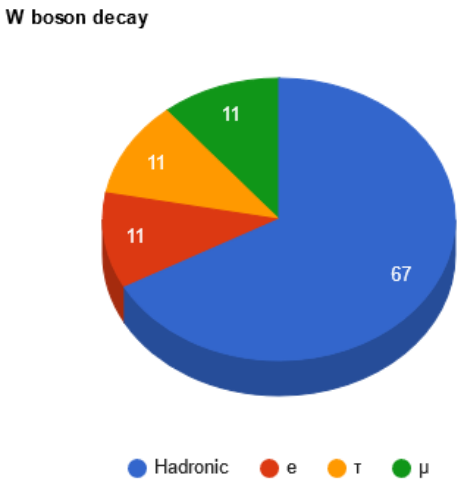

Figure 2.8: Possible decay modes of the W boson. In 67% of the times it decays into a pair of quarks. The rest 33% it decays in to a lepton and the corresponding neutrino.

- **Dilepton channel**. The final case is when both W bosons decay leptonically. It has a branching ratio of around 10.7%.

A picture with all the possible decay scenarios as well as their probabilities is given in Figure 2.9.

2.3 Cross Section

The cross section (σ) of a process is a measure of the probability that this process will take place. The cross section is calculated theoretically using the Feynman diagrams, e.g. like the ones in Figure 2.5. The calculation is done by summing all the possible diagrams. The diagrams in the Figure 2.5 are called leading order (LO) diagrams since they contain only the initial and final state particles plus the mediators. We know that particles can be emitted in any point so with this in mind one can create more complex diagrams like the ones in Figure 2.10. These diagrams are called next to leading order diagrams (NLO). The order of the diagram is specified depending on the number of extra particles emitted or what is called the number of vertices. A vertex is practically a point where an interaction happens.

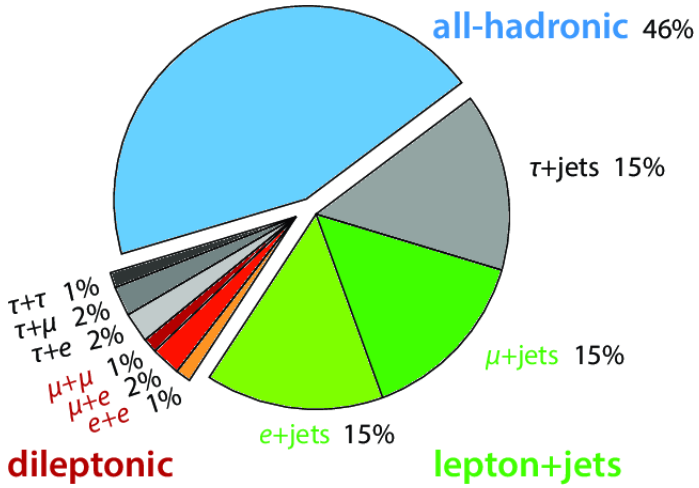

Figure 2.9: Pie chart with percentages of $t\bar{t}$ decay modes. The top quarks decay in to a W boson and a b quark. Subsequently the W boson also decays. It can decay in a quark pair or a lepton and a neutrino. Based on the way the two W bosons decay the possibilities of the total decay can be seen in the pie chart depicted in the picture. Taken from [10]

Increasing the complexity of the diagrams we can see that practically one can create and calculate an infinite number of diagrams. Then summing over all the diagrams, the cross section then can be calculated as:

$$\hat{\sigma}_{ij->t\bar{t}+X} = \underbrace{H^{(0)}}_{\text{LO}} + \underbrace{\alpha_S H^{(1)}}_{NLO} + \underbrace{\alpha_S^2 H^{(2)}}_{NNLO} + \dots \tag{2.2}$$

However since α_S is less than 1 we can see that as the diagrams become more and more complex they contribute less and less to the total cross section. To accompany this, the diagrams become more and more complex to calculate as their order increases. Usually a calculation up to the NNLO term is done.

Collider experiments on the other hand provide an easier way of calculating the cross section. It can be proven that the cross section is analogous to the number of measured events with the following relationship:

$$\sigma = \frac{N}{L_{int}} \tag{2.3}$$

Where N is the total number of measured events for this specific process and L_{int} is the integrated luminosity. Luminosity (L) is a measure of the number of

collisions that can be produced in a detector per cm^2 per second. The integrated luminosity is the luminosity "gathered" throughout the experiment. It is practically the integral of the luminosity for the duration of the experiment. We can see this way that collider experiments are a good way to calculate the cross section and thus verify the SM.

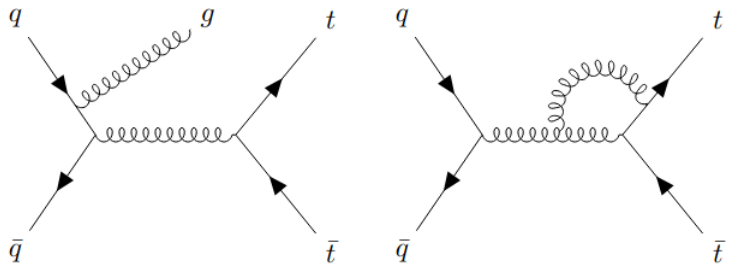

Figure 2.10: Next to leading order Feynman diagrams for the production of the $t\bar{t}$ via gluon fusion.

Chapter 3

The Compact Muon Solenoid Experiment

In this chapter the Compact Muon Solenoid (CMS) detector will be described. Its parts and principle of use will be explained.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the biggest particle accelerator ever built. It was build at the European Council for Nuclear Research (Conseil Européen pour la Recherche Nucléaire CERN), in Geneva, Switzerland and started operating at 2008. It is a circular accelerator with a circumference of 27 kilometers (km) sitting 100m underground in the France-Swiss borders between Geneva and the Jura mountains. Figure 3.1 shows an aerial photograph of the area where the LHC resides. It operates by accelerating two proton beams one clockwise and one anti-clockwise. The beams collide at 4 predefined interaction points. Particle detectors are built at each of the interaction points. The accelerator is designed in order to be able to accelerate proton beams to a center mass energy of 14 TeV making it the most powerful accelerator ever built. It aim to study the structure and behavior of elementary particles as well as their interactions.

3.2 The LHC accelerating complex

The LHC was constructed in the re-existing tunnel of the older accelerator Large Electron Positron collider (LEP). It consists of a set of superconducting magnets

Figure 3.1: Aerial photograph of the Large Hadron Collider (LHC) at the European Council for Nuclear Research (CERN). In the picture the tunnel that the two proton beams can be seen as well as the location of the four detectors. All structures are located 100 meteres underground.

used to curve the beam and guide it in the circular pipes but to also focus it in the spots where collisions are performed. Approximately 1200 magnets are used for the curvature of the beams and 400 for its focus. The magnets are cooled to a temperature of 1.9 Kelvin (K) and are kept at that temperature during the accelerator operation. To achieve such low temperatures, liquid helium is used for the cooling. At the center of each magnet there are two vacuum tubes where the proton beam lay Figure 3.2.

In order to accelerate the beams CERN uses a set of accelerators, not just the LHC. The beam goes through various accelerating stages before entering the LHC where they reach the maximum energy and collide. In Figure 3.3 the CERN accelerator complex can be seen. Protons used in the LHC originate from a bottle of hydrogen where under strong electric field electrons are removed from the atoms' nuclei resulting in free protons. After their production, protons are inserted in the first accelerator, LINAC 2 which is a linear accelerator. Inside LINAC2 they reach an energy of 50 MeV. After that, they enter the Proton Synchrotron Booster (PSB) and the Proton Synchrotron (PS) reaching an energy of 26 GeV. The next stage of acceleration is at the Super Proton Synchrotron (SPS) where protons are accelerated at an energy of 450 GeV. Finally the beams enter the LHC. It is worth mentioning that the beams are not continuous. They consist of packets of protons called bunches equally spaced between them.

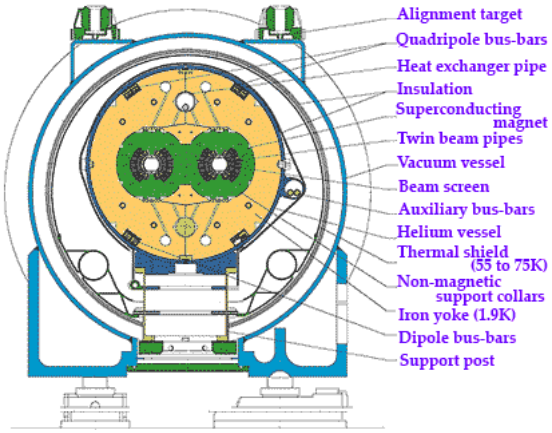

Figure 3.2: Cross section of one of the LHC magnets. In the center the two vacuum tubes can be seen where the beams travel as well as the various magnets used for guiding and focusing the beam.

The main physical quantities affecting the operation of the LHC are the energy of the two beams and the luminosity. The rate at which events occur during the accelerator operation is:

$$R = \sigma \mathcal{L} \tag{3.1}$$

where σ is the cross section of the physical process and $\mathcal{L}$ is the luminosity. To achieve a steady operation for a given energy, the luminosity must increase with a factor of E^2 because the cross section is inversely proportional to that quantity ($\sigma \propto 1/E^2$). The luminosity of an accelerator which collides 2 bunches of particles n_1, n_2 with a frequency f, is:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} \tag{3.2}$$

Where σ_x, σ_y is the transverse profile of the beam. The predicted luminosity for the LHC is $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$. It is worth mentioning that the LHC will also conduct heavy ion collisiont using lead with a luminosity of $\mathcal{L} = 10^{27} cm^{-2} s^{-1}$ at a center of mass energy of 1312 GeV.

To achieve such high luminosity, the frequency of collisions as well as the number of protons in each bunch must be very high. For that reason each proton beam consists of 2808 bunches closely spaced between them. Each bunch consists of approximately 10^{11} protons which collide very 25 ns. This results in a collision frequency of 40 MHz.

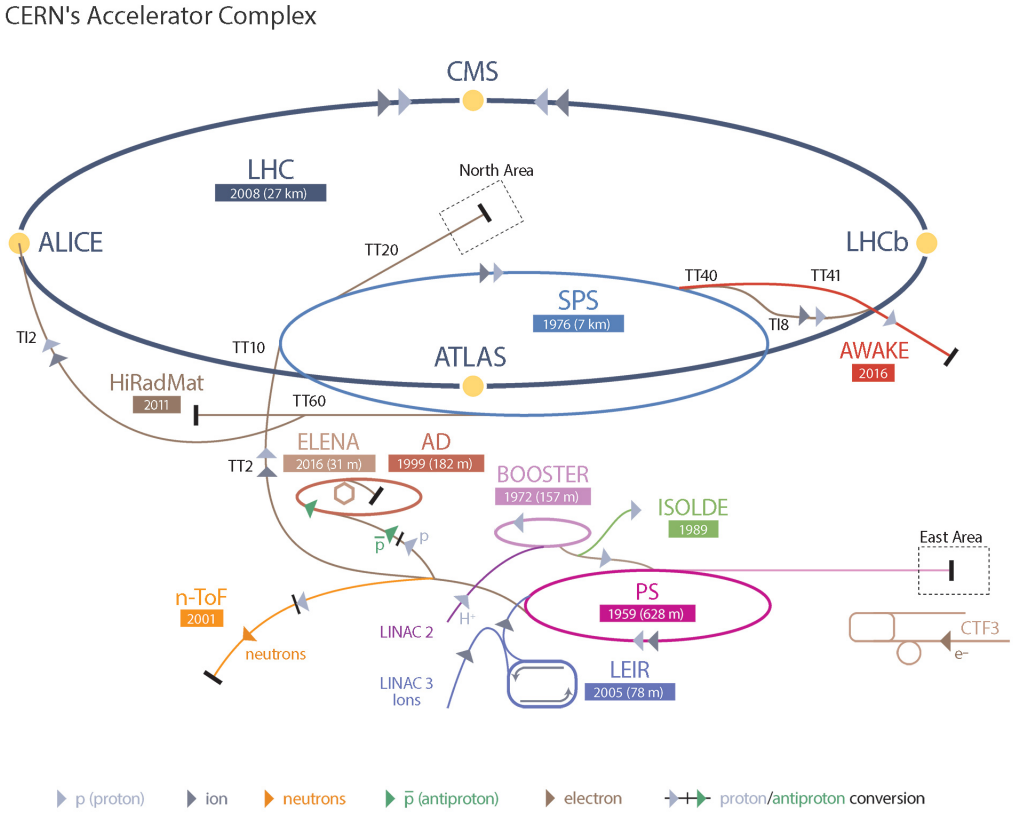

Figure 3.3: The CERN accelerating complex. Proton beams initially enter the linear accelerator LINAC2. After that the continuously go through the Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) before entering the LHC. Apart from the LHC the accelerator complex is used to provide beams to other experiments at CERN, like n-TOF, ISOLDE, ELENA and others.

3.2.1 The LHC detectors

The two accelerating proton beams collide with each other in order to study various physical processes. The collisions take place at 4 predefined points in the accelerator where detectors are placed Figure 3.3. In total the LHC has 8 access points. At these points it is possible to access the accelerator from the surface. At 4 of these points particle detectors are positioned and this is where the physics experiments take place. The 4 LHC particle detectors are:

- **A T**oroidal **L**HC **A**pparatu**S** (**ATLAS**), Point 1.
- **C**ompact **M**uon **S**olenoid (**CMS**), Point 5.
- **LHC b**eauty (**LHCb**), Point 2.
- **A L**arge **I**on **C**ollider **E**xperiment (**ALICE**), Point 8.

The first two detectors ATLAS and CMS, are general purpose detectors. On the other hand, LHCb and ALICE are built for studying specific physical processes. LHCb is designed to study the b (beauty) quark and the measurement of the violation of the charge parity (CP) symmetry. In short this is called CP-violation. ALICE is built to study the collisions of heavy ion beams and also the ones between heavy ions and protons. Apart from the 4 big experiments there are other smaller ones like TOTEM and CASTOR that share the shame interaction point with the CMS detector. Finally, in Figure 3.3 other smaller experiments are visible. These experiments use beams from smaller accelerator of the CERN accelerator complex. Some of these experiments are ISOLDE (**I**sotope **S**eparator **O**n **L**ine **DE**vice), nTOF (**n**eutron **T**ime **O**f **F**light), ELENA (**E**xtra **L**ow **EN**ergy **A**ntiproton) and others.

3.3 Compact Muon Solenoid

The CMS detector, is a general interest high energy physics experiment. It has a cylindrical shape with various coaxial layers around the beam axis like a barrel. In order to detect all the produced particles it needs to be hermetically sealed. To do so it uses 2 vertical disks perpendicular to the beam axis called endcaps. The detector has a length of 28.7 meters (m) and a radius of 15 m and weighs 14.000 tonnes. It is relative small for the amount of matter it contains, justifying the word compact.

One of the most interesting particles that CMS aims to detect are muons. Muons are indications that an "interesting" physical process has occurred during

the collisions. Considering the importance of muons, CMS is designed and built in a way to detect muons with very high precision. That is where the word muon comes from. Finally CMS uses a big solenoid magnet capable of producing a magnetic field of up to 4T parallel to the beam axis. The magnet is used to bend the tracks of the particles perpendicular to the axis of the beam. The presence of the soilenoid magnet justifies the word solenoid is the detector's name.

As the particles traverse the detector, the magnet is used to bend their tracks. By measuring the curvature of the tracks we can measurement the momentum of the corresponding particle. It can be proved that the particle's momentum can be reconstructed with a precision that is relative to the magnetic field and the length of the detector with the following relationship:

$$\frac{\sigma_{p_T}}{p_T} = \frac{\sigma_s}{s} = \sigma_s \frac{8 p_T}{0.3 B L^2} \tag{3.3}$$

Where p_T is the transverse momentum of the particle, s is the curvature of the track due to the magnetic field, B is the strength of the magnetic field and L is the length of the detector. CMS is built such that it has a big magnetic field and a small size while ATLAS is built the opposite way, having a bigger size and a relative smaller magnetic field.

To be able to achieve its goals, CMS must have:

- A high precision tracking system that is at the same time able to provide information for the momentum of the particles.
- An electromagnetic calorimeter of really high precision in order to accurately detect the energy of electrons and photons.
- A hadronic calorimeter capable of detecting all the neutrally charged particles produced by the collisions in the center of the detector
- A very efficient muon system for detecting muons but also measuring their properties such as the momentum etc

CMS is designed in a way that consists of multiple layers, each playing a different role in particle detection. Each of these layers is a separate detector system called a subdetector. They each provide a different piece of the total information needed to detect a particle. All the pieces are collected and combined in order to identify the particles produced during the proton collision. CMS consists of 4 subdetector systems Figure 3.4. The tracker, which is responsible for detecting the particle tracks, the Electromagnetic CALorimeter (ECAL), the Hadronic

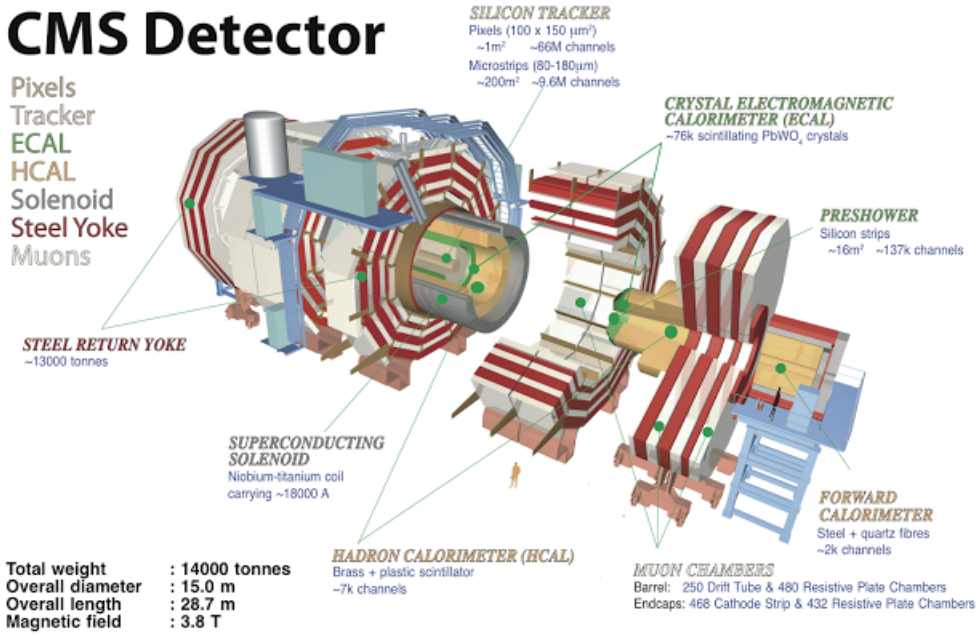

Figure 3.4: The Compact Muon Solenoid detector. It consistes of 4 subdetecrors. The tracker, the electromagnetic calorimeter, the hadronic calorimeter and the muon detectors. It also contains a solenoind magnet capable of producting a magnetic field up to 4 Tesla.

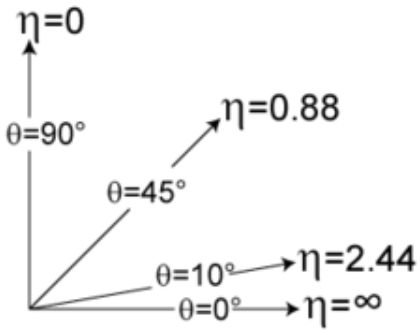

Figure 3.5: The value of pseudo-rapidity ($\eta = -\ln(\tan(\theta/2))$) as a function of the angle θ. Angle θ is the angle from the positive part of z-axis. As z-axis we define the axis parallel to the axis of the beam.

CALorimenter (HCAL) and the muon detectors. Apart for these systems, CMS also contains a solenoid magnet. The magnet is not used in detecting the particles but is a vital part in the operation of the detector.

The layout used for the subdetecting systems is the following. The Tracker, ECAL and HCAL are inside the solenoid magnet while the muon systems are outside the magnet. Each one of the subdetectors has a part in the main volume of the detector the barrel and a part in the endcaps.

In the next sections there will be a detailed description of each of the subdetecting systems. Before that it is important to do a quick introduction to the coordinate system used in the LHC experiments. For the description of the detector a Cartesian coordinate system can be used, where the x-axis points to the center of the LHC ring, y-axis points upwards perpendicular to the axis of the beam and z-axis is parallel to the axis of the beam. Because the shape of the detector is cylindrical it is more convenient to use a different coordinate system that consists of the three following quantities, r, ϕ, η, where r is the distance from the z-axis, ϕ is the azimuthial angle from the x-axis and η is a quantity called pseudo-rapidity. It is calculated from the following equation $\eta = -\ln(\tan(\theta/2))$ where θ is the angle from the positive part of the z-axis. In Figure 3.5 the value of pseudo-rapidity for various values of the angle θ.

3.3.1 The tracker

The tracker detector is the first detecting system that CMS has and is the one closer to the interacting point of the two proton beams. It is used to reconstruct the tracks

of the particles produced by the collisions and also measure their momentum. It can measure with really high precision the momentum of muons, electrons and also charged hadrons. It can also detects vertices, which are the points that particles produced by the collisions decay into other secondary particles. This detector has a length of 5.4 m and can cover an area of $r < 1.2$ m and $|\eta| < 2.5$.

Due to the high luminosity of the beams and of its position, the tracking detector needs to have very good spatial resolution but also to be radiation tolerant. It also needs to have a very good response time in order to record as many events as possible. The material used for this particular detector is silicon. For tracking detectors it is really important that they contain as less material as possible. The reason is that the particles that traverse the detector's volume interact with the material and leave energy during these interactions. If the detector contains too much material it is possible that the particle leaves all its energy inside the tracking detectors and does not move to the next subdetecting systems thus loosing important information for the detection of the particles. For the tracker in CMS two different technologies are used. The first stage is a detector consisting of silicon pixels while the second one consists of several layers of silicon strips. Spatially the detector is split in the following parts. the most inner part is the pixel detector. Then there is the strips parts which are split in the inner cylinder (Tracker Inner Barrel, TIB), then there is the outer part of barrel, (Trakcer Outer Barrel, TOB), the inner Disks (Tracker Inner Disks, TID), and the endcaps, (Tracker EndCaps, TEC). In Figure 3.6 the various parts of the tracker detector are visible as well as their spatial coverage.

3.3.2 The Electromagnetic CALorimeter (ECAL)

The second detecting system of CMS is the electromagnetic calorimeter (Figure 3.7). It is used to measure the energy of electrons and photons that are produced during the collisions. Ecal was designed so that it can detect with really high accuracy the energies of photons. That is because one of the most common ways for a Higgs boson to decay is via the interactions $H \rightarrow \gamma\gamma$ in to two really energetic photons. In addition the calorimeter was designed to have a very good resolution in the detection of two photons because two photons can also be produced from the decay of a neutral pion, $\pi^0 \rightarrow \gamma\gamma$. This means that the distinction of these two processes was of a vital importance for the detection of the Higgs boson.

Ecal was constructed using lead-tungstate ($PbWO_4$) crystals. This material

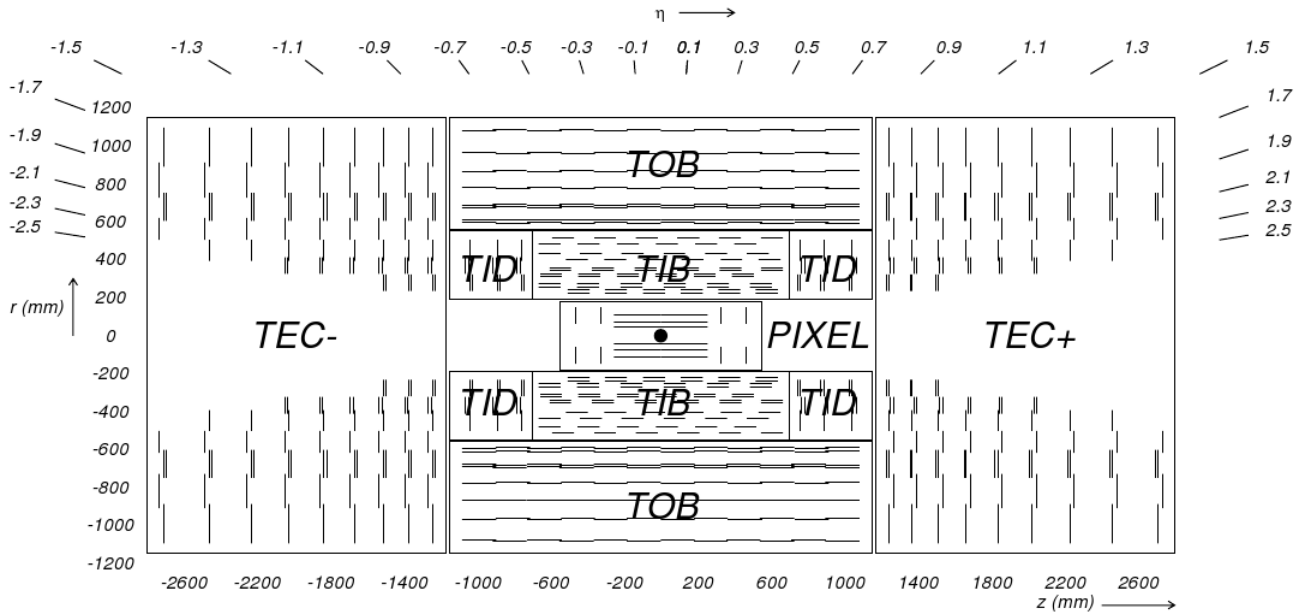

Figure 3.6: Graphical representation of the CMS tracker detector. It was built utilizing two different detector technologies. The first segment, at the inner part of the detector consists of silicon pixels. In the outer part there are several layers of silicon strips. The various segments are the: Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker Inner Disks (TID) and the Tracker EndCaps (TEC). In the picture the spatial coverage of the detector is also visible.

was chosen as it is very dense ($8.28 g/cm^3$) and has small radiation length[1] ($X_0 = 0.89\,cm$) and Moliere radius[2] ($R_m = 0.89\,cm$). In order be able to produce these kind of crystals that have the desired properties and can withstand the LHC conditions a lot of research had to be made. Each one of these crystals takes two days to produce.

The energy measurement of the particles is based on a principle called scintillation. When a particle that interacts electromagnetically traverses the volume of the detector it interacts with the material of the detector and produces light. The light produced can be detected and by measuring the amount of light produced, the energy of the particle can be measured. One more reason that this material was chosen for the ECAL crystals was that it has a very fast response. Approximately 80% of the produced light is generated in 25 ns which is the the same frequency with which the LHC beams are crossed. On the other hand, a disadvantage is that they emit little light. To compensate for that, photodiodes are used which collect the light and enhance it so it can be detected easily. Two types of photodiodes are used, the Avalance PhotoDiodes (APD) and the Vacuum PhotoTriodes (VPT).

Apart from the main part of the ECAL, CMS has another type of electromagnetic calorimeter. This detector is called Preshower and is located only in the endcaps. It is used in the following way. If the two photons produced in the decay of the neutral pion ($\pi^0 \rightarrow 2\gamma$) have a very small angle between them it is possible that they are detected as one very energetic photon and not two. The preshower is a calorimeter with very good spatial resolution, better than the one of the main ECAL. This way the two photons produced from the neutral pion decay can be identified correctly and assigned to the pion and not to a possible Higgs boson.

Like all subdetectors, ECAL is splitted in two parts. The one being in the barrel (ECAL Barrel, EB) and the one located in the endcaps (ECAL Endcap, EE). Ecal covers an area of 1.2 m $< r <$ 1.8 m and $|\eta| < 3$.

EB covers the area $|\eta| < 0.1479$ and consists of parts called supermodules each containing 4 modules. The first module of each supermodule contains 500 crystals while the rest contain 400. There is a total of 26 supermodules and 61200 crystals in the EB. The part of the detector located in the endcap covers the region $0.1479 < |\eta| < 3$. EE consists of 2 semicircles made of "supercrystals". Each "sypercrystal" contains 5×5 making a total of 7324 crystals for each one of the endcaps of the ECAL.

[1] Radiation length is defined as the distance that a particle needs to travel in order for its energy to be lowered by a factor of 1/e of its initial one, due to electromagnetic interactions.

[2] Moliere radius is a measure of the electromagnetic shower caused by a photon or an electron. It is defined as the radius of a cylinder that contains the 90% of the energy of the shower.

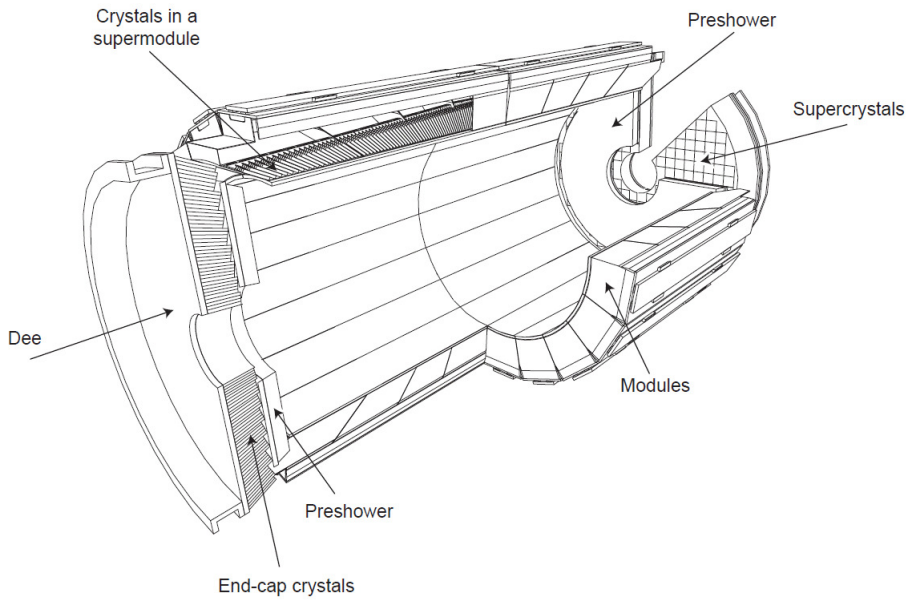

Figure 3.7: The electromagnetic calorimenter of the CMS detector. ECAL is built from lead tungstate crystals ($PbWO_4$). In total 61200 crystals were used for its construction. ECAL consists of the barrel part, ECAL Barrel (EB) and of the endcap part, ECAL Endcap (EE). Finally ECAL also includes the preshower detector.

3.3.3 The Hadronic CALorimenter (HCAL)

The third layer of the CMS detector is the Hadronic CALorimenter (HCAL). This subsystem is responsible for measuring the energy of hadrons and their decay products. Particle detection is based on hadronic showers. They are the result of the interaction of the hadrons, traversing the detector's plane, with the nucleus of the atoms of the detector's material. The "active" material used in the HCAL is fluorescent plastic. The HCAL is a sampling calorimeter which means that it consists of layers of detecting material and absorber. As an absorber brass was used together with steel. The construction of this detector was a challenge as big amount of material had to be fit in between the ECAL and the magnet.

When a particle goes through the active material of the detector, light is produced. This light is of blue color and is collected by wavelength-shifting fibers which transform it in green. It is then transferred in special detectors called Hybrid Photodiodes (HPD) where it is amplified and transformed into electrical signal. From where it is finally transferred to the data collecting unit.

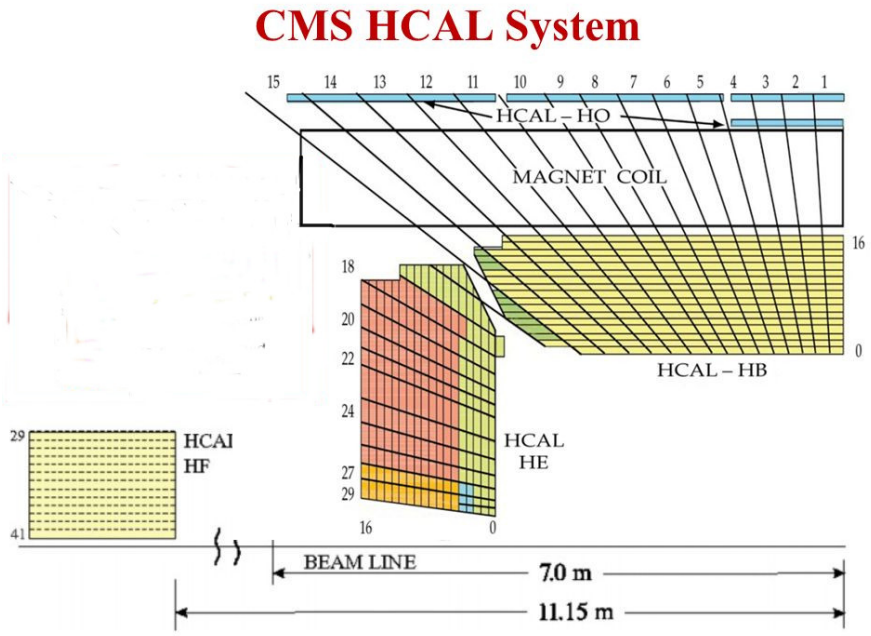

Figure 3.8: Graphical representation of the Hadronic CALorimenter (HCAL) of CMS. HCAL consists of two parts, the barrel part and the endcap. The part in the barrel is again split in to two more parts the inner and outer one. In the endcaps there is the HCAL endcap part and the HCAL forward detector.

HCAL (Figure 3.8) consists of two parts, the one at the barrel , HCAL Barrel (HB) and the other at the endcap, HCAL endcap (HE). These two parts provide a coverage of $|\eta| < 3$. HCAL also includes the HCAL forward detector (HF) which is used to detect hadrons that decay in very small angles compared to the beam axis. This part of the detector receives high amount of radiation because as the beams cross the majority of the protons do not collide but are scattered in small angles and hit the endcap part of the detector. For that reason the HF part uses optical fibers that emit Cherenkov radiation and are very tolerant to radiation. While HCAL resides inside the magnet hadrons that are emitted in large angles, almost perpendicular to the beam axis do not go through enough layers of the calorimeters so that their energy is measured well enough. To mitigate that, a part of the HCAL, called HCAL outer (HO) has been placed outside of the magnet, covering an area of $|\eta| < 1.26$.

Apart from detecting particles directly, ECAL and HCAL also measure particles indirectly. This is in the form of missing momentum. Some of the particles that

are of great interest in modern physics such as neutrinos interact very rarely or some times never with ordinary matter. This makes their detection impossible. From the missing momentum one can indirectly measure the existence of a particle that was produced and was not detected. For that reason ECAL and HCAL must be hermetically closed because there must be no possibility of ordinary particles escaping the detector undetected and them be miss-interpreted as missing momentum.

3.3.4 The magnet

One of the most difficult and important aspects in the constructions of CMS was the choice of the magnet. The difficulties included both the selection of the type as well as the strength of the field it will produce. CMS chose to use a superconducting solenoid. It has a length of 13 m parallel to the beam axis. The magnet is capable of producing a magnetic field of 4 T. It is big enough so that 3 out of the 4 subdetector systems of CMS reside inside the magnet's volume. It is used to curve the track of charge particles so that their momentum can be measured. The solenoid magnet covers and area of $|\eta| < 1.5$.

3.3.5 The muon detectors

The last layer of CMS is the muon detecting system. As mentioned above the detection of muons is very importance for the LHC experiments. Muons are very penetrating and can go through a long distance inside a material without interacting and thus without being detected. In order to detect them, a dedicated system of detectors is used. The muon system also plays a key role in the trigger system of CMS. The muon system consists of 4 stations covering a total area of 25000 m^2 of active area[3]. Due to the large area as well as the different radiation dosages in the various areas of the detector, three different types of detectors are used in the muon system:

- Drift Tubes (DT).
- Cathode Strip Chambers (CSC).
- Resistive Plate Chambers (RPC).

[3]Area in which a particle can be detected

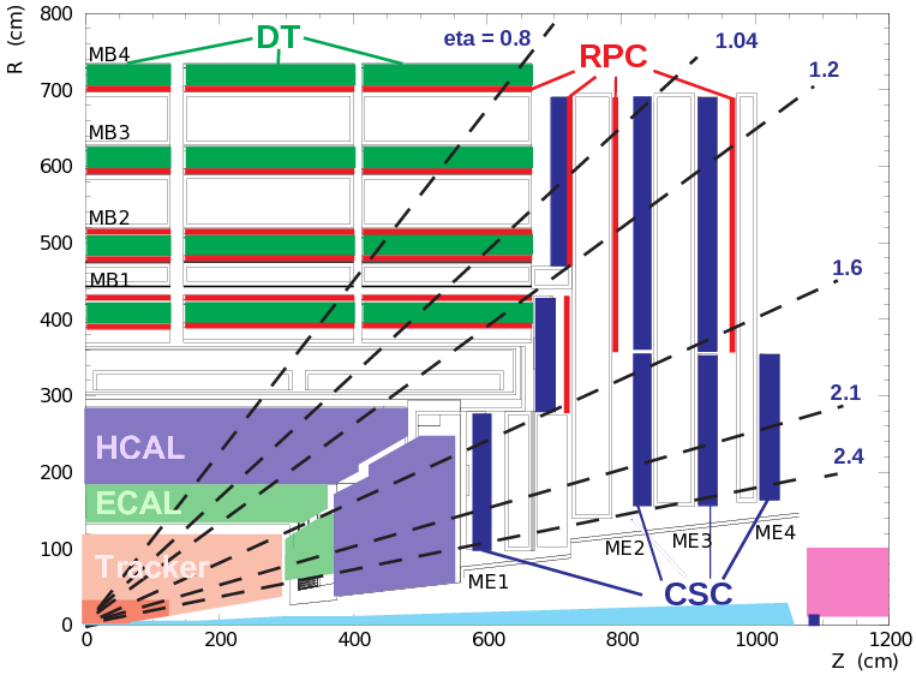

Figure 3.9: The muon system of CMS. In total, three different detector types are used for detecing muons in CMS. The Drift Tubes (DT) located in the barrel, the Cathode Strip Chambers (CSC) which are placed in the endcaps and the Resistive Plate Chambers (RPC) which are placed in both parts of CMS.

Each of these types is used for a different purpose. The first two have very good spatial resolution while the RPCs have very good time resolution. The muon system covers the area $4 < 7.4$, $|\eta| < 2.4$.

The **drift tubes (DT)** are placed in the main body of the CMS detector, the barrel. In the barrel the particle flow is relative slow and the magnetic field homogeneous. Drift tubes consist of 5 cylinders, coaxial with the center of the beam which cover all the barrel. Each disk has 4 stations of different radius (MB1 - MB4). Each tube has a diameter of 4 centimeters and a wire in the center. It is filled with a gas mixture of argon and carbon dioxide. When a muon passes through the gas it ionizes the gas molecules. The electrons created from the ionization move towards the wire due to the electric field in the center of the tube. By detecting the point at which the electrons hit the wire and the time it took for the electrons to drift to the wire the muons track can be calculated. Each DT detector has a size of $2 \times 2.5\ m$ and consists of 12 layers split in to 3 groups with

a total of 60 tubes. The middle group detects the coordinate of the track parallel to the beam axis while the other 2 groups detect the the coordinates in the plant perpendicular to the beam axis.

The **cathode strip chambers (CSC)**, are placed at the endcaps perpendicular to the beam direction where the flow is large and the magnetic field non uniform. The CSCs have very small response time and thus can be used in an environment with such a high amount of radiation. In addition they can provide information regarding the timing of the particles. Their operating principle is based on a mesh of wires. The have positive charged wires which act as anodes and negative charged wires which act as cathodes. The mesh resides inside a gas mixture. When a charge particle goes through the gas, the gas molecules become ionized. The electrons produced in the ionization move towards the cathodes producing an avalanche. The signal produced by the avalanche provides information regarding the trajectory of the particle that produced it .

The **resistive plate chambers (RPC)**, are placed both in the barrel and in the endcaps of CMS. They work complementary with the other two muon systems providing triggering information. They consist of two plates, one positively charged acting as an anode and one negatively acting as a cathode. In between the two plates, there is a gap (2 mm approximately) filled with gas. When a muon pass through the gas it ionizes it. Due to the charged plates there is an electric field inside the gas which makes the electrons accelerate causing an avalanche. The particles in the avalanche are guided to the anode where they are collected providing information for the muon that passed. The electric field inside the plates is causing the signal to be produced faster than the collision rate of the LHC. Due to this, the detector has a very good time response. This together with the good spacial resolution allows the RPC to be used for triggering.

3.3.6 The triggering system

During the operation of the LHC, the detectors record data at a rate of $10^9\ Hz$. With a rate like that each of the detectors will be producing 100 Terabytes of data per second. It is impossible to store all this information. Firstly due to the lack of storage space and secondly because it would not be possible to process this amount of data. In addition only a fraction of this data are interesting from a physics point of view. To solve this problem the detectors have a filtering system in order to reduce the amount of data recorded. In CMS this system is called **T**rigger and **D**ata **A**cquisition **S**ystem (**TRIDAS**). The process is separated in two stages. The first stage is what is called the Level-1 trigger and the second one the

High Level Trigger (HLT).

The Level-1 Trigger is the first step of the triggering system of CMS. At this stage the system has a total time of 3.2 μs to decide for each event whether it is worth storing or not. In reality due to delay problems the time is approximately 1 μs. The data are stored in buffers until the trigger decides whether they will be kept or not. This step consists entirely of hardware which is located as close as possible to the detector in order to minimize delay due to data transfer. Because of the very small amount of time that this system has in order to decide it uses information solely from the calorimeters and the muon system. After this stage the total amount of data is around 50 kHz which are then transferred to the HLT for further filtering.

The HLT consists of solely software. It uses algorithms similar to the ones used from physicists in offline analysis. The filtering is done in 3 stages. In the first stage data from the calorimeters and the muon system are used in a manner similar to the one in the Level-1 Trigger but with greater detail. After that, information from the tracker is incorporated. Finally all the information is used and the event is reconstructed. After the triggering process is finished, the data rate has dropped to around 100 Hz. The data selected by the the two levels of triggering system are then saved and used for offline analysis by the physicists.

Chapter 4

Object Reconstruction

One of the most important things in a high energy physics experiment is the proper reconstruction of the particles generated inside the detector. After protons collide particles are generated. As these particles traverse the detector volume they interact with the detector's material generating signals. Usually detectors consists of different types of subdetecting systems. The signals produced by the generated particles vary on type and characteristics based on the particle type as well as the detector type that they interact with. Object reconstruction refers to the use of the various signatures left behind by the particle in order to identify the type of particle as well as its kinematic properties e.g. momentum, energy etc. To do so CMS uses an approach called Particle Flow (PF). Instead of using the information from each sub-detecting system individually, CMS combines them to achieve better results. In the following chapter the various techniques and methods used for object reconstruction in CMS are explained.

4.1 Track and Vertex reconstruction

The first step of object reconstruction is the determination of the tracks of particles. This is done using information from the pixel and strips parts of the tracker detector. CMS uses an algorithm called combinatorial track finder (CTF). It is an iterative algorithm that runs in six iterations and has a very high efficiency and low fake rate. Hits on the tracker are combined into tracks using very tight criteria in the first iteration. This means that the algorithm has a small fake rate but not very good efficiency at the same time. Hits that are matched to a track are removed and a second iteration of the algorithm begins. At this step the requirement criteria are looser. This procedure is repeated again with loosening the criteria in each

iteration. In the final iterations more complex tracks are reconstructed.

The next step is the determination of the primary vertex. As a vertex we define a point of interaction. For example the point where a collision happened or where a particle decayed into other particles. As primary vertex (PV) we define the point where the hard scattering from the proton proton collision occurred. Tracks are selected and grouped based on their distance from the beam interaction point. A fit method is used utilizing the full tracking information to determine the likelihood that this track corresponds to the primary vertex. More than one vertices can be reconstructed in this way. After all vertices are determined, the one with the highest sum of transverse momentum squares ($\sum p_T^2$) is chosen as the PV.

4.2 Calorimeter clustering

The calorimenters are used to detect and measure neutral particles as they do not interact with the tracker. They separate the energy deposits that neutral and charged particles leave behind. They identify electrons and their accompanied photons of Bremsstrahlung radiation and finally give better accuracy in the measurement of particles with very high energy as the tracker can not measure their energy accurately enough.

Calorimeter clustering refers to the measurement of the energy of the particles by the various calorimetric layers. The calorimeters in CMS consist of many smaller parts. In ECAL they are called crystals while in HCAL modules. They are both called calorimeter cells. Usually a particle that goes through the calorimeter it leaves its energy in more than one of the calorimeters cells. In that sense, clustering refers to combining the information from the various cells in order to measure the whole energy of the particle.

It starts from a cell where the energy exceeds a certain threshold. Adjacent cells with an energy that exceeds twice the noise level of the cell are associated with the starting cell forming a cluster. Various methods have been developed for measuring and correcting the energy of the particles as its measurement is crucial for the experiment.

4.3 The linking algorithm

The various signals left behind in each subdetector system of CMS are called PF elements as they are used in the particle flow algorithm in order to identify particles. In Figure 4.1 the different signatures left by various particle types inside

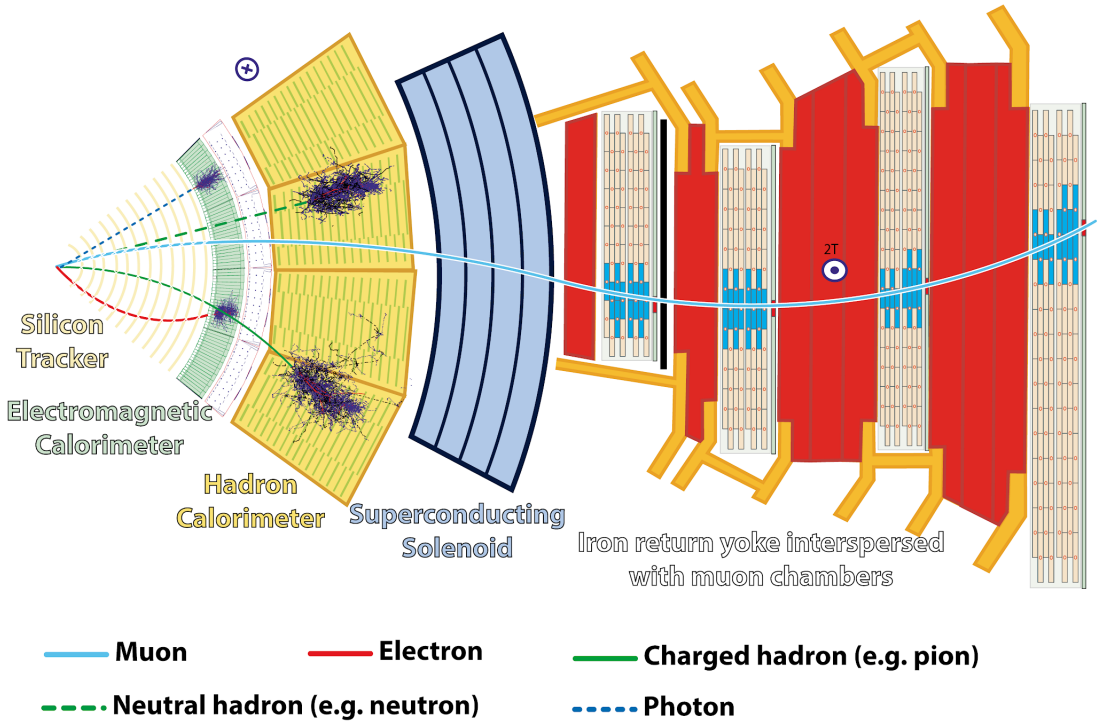

Figure 4.1: Cross section of the CMS detector. In the picture, the various signatures left by each particle type in the corresponding detector subsystem can be seen.

the detector can be seen. One of the most fundamental parts of the PF algorithm is the association of the separate elements of the same particle together at the same time avoiding double counting. The core of this procedure is the linking algorithm.

The linking algorithm links tracks and calorimeter clusters by extrapolating their trajectories. A link is also performed to associated charge particles with Bremsstrahlung radiation. The link is established by extrapolating tangents from the intersecting point of the tracks pointing to the ECAL. If the extrapolation falls within an ECAL cluster the cluster might be a Bremsstrahlung photon. ECAL, HCAL and preshower clusters are also linked. It is done by checking if the cluster in the more granular calorimenter is in the boundary of the other. Finally tracks from the tracker are linked to tracker in the muon system.

4.3.1 Muons

The particle flow algorithm reconstructs muons based on the information from the tracker and the muon system. A reconstructed muon from the muon system is called a standalone muon, while one generated using the tracker information is called a tracker muon. Combining the information from both detecting systems a global muon is reconstructed.

4.3.2 Electrons and Photons

Electrons and photons are reconstructed combining information from the tracker and the ECAL. While traversing the tracker material, electrons can emit photons via bremsstrahlung radiation. Photons can then be converted to electron/positron (e^+e^-) pairs which can also emit photons. For that reason the reconstruction of photons and electrons happens together. When a track is an electron candidate, an additional fit with a different GSF is done. This allows for the sudden drop of energy due to radiation. An electron candidate originates from a GSF track that is not linked to 3 or more additional tracks. On the other hand, a photon candidate comes form an energy cluster that has no link to a GSF track. The cluster must also have transverse energy (E_T) more than 10 GeV.

4.3.3 Jets

Quarks and gluons are not detected individually but form more composite particles through the hadronisation process that was described earlier. These composite particles then decay leaving the experimental signatures that we see in the detectors. Jets are the groups of particles that produced during this process. Due to the number of these particles it is not possible to reconstruct them individually so the structures called jets are formed. It is also not of much importance to detect the products of the particles that are produced by the decay but rather through the jet properties to study the initial particle from which the jet was produced. The main purpose it to measure the energy of the jet. Jets are reconstructed using algorithms that go through the various objects detected in the calorimeters in order to group them together. As jets are composite objects their definition is not unique and depends on the algorithm that was used in order to generate them.

Many jet clustering algorithms are used in particle physics. One of the most import feature is that need to be infrared and co-linear (IRC) safe. This means that they should provide the same result of clustering when a soft particle is included in the cluster or when a parton splits in to two other partons, see Figure 4.2.

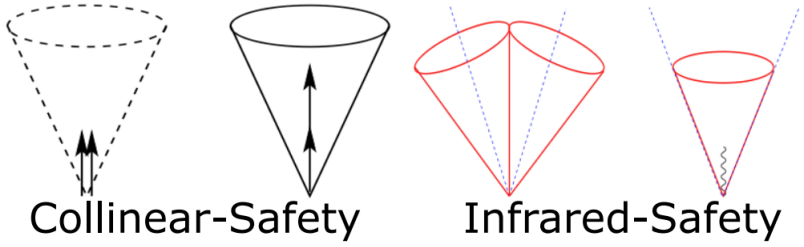

Figure 4.2: An infared and colinear safe jet reconstruction algorithm. The algorithm should give the same result when a soft particle is included in the particle or when a parton splits in two other partons.

The most used algorithm in CMS is the anti-k_t algorithm [11] and the jets are called AK jets. In the process of reconstruction with the anti-kt algorigth, for each PF candidate i the following two parameters are calculated:

$$d_{i,B} = p_{T,i}^2 \tag{4.1}$$

$$d_{i,j} = min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.2}$$

where $d_{i,B}$ is the distance between entity i and the beam line. $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, is defined as the distance between the two entities i,j in the $\eta - \phi$ plane, where y is the rapidity and ϕ is the azimuthial angle from the x-axis. After calculating there quantities, the algorithm proceeds as follows. It calculates the minimum of these two quantities. If $d_{i,B}$ is the minimum, then i is considered a jet and is removed from the collection and the algorithm continues. If $d_{i,j}$ is the minimum, then the two entities are merged. The distances are then re-calculated and the process is repeated until no entities are left.

The parameter R is called the jet radius and defines how big is the cone that is used to group particles. The default one in CMS is 0.4 and the jets produced are called AK4 jets. Other distance parameters can be used. In this particular thesis, AK8 jets were used where the jet radius is 0.8. Bigger radii are usually used to reconstruct jets that originate from particles with high transverse momentum and hence large Lorentz-boost. Figure 4.3 shows the reconstruction of jets using the anti-kt algorithm with a radius of 1.

When reconstructing a jet it is possible that extra energy is accounted to it coming usually from pileup. Pileup is defined as the phenomenon under which two or more particles collisions happen during the same beam crossing and hit the detector at the same time. This can lead in significant loss of the detector's

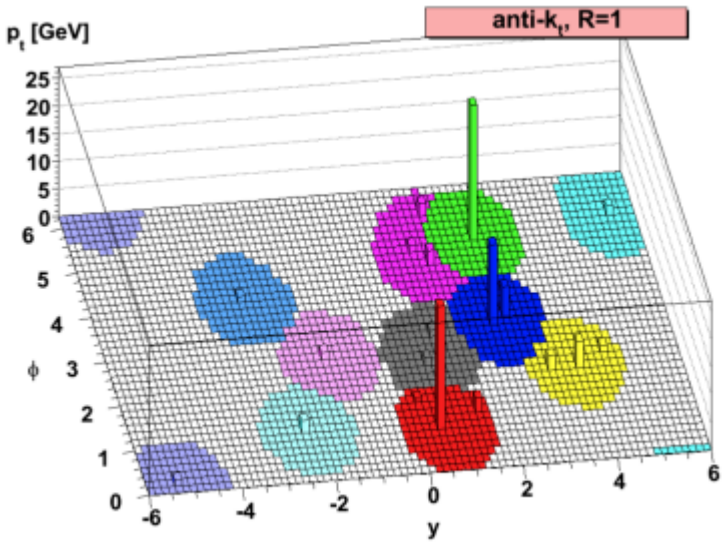

Figure 4.3: Example of the jets reconstructed using the anti-kt jet reconstruction algorithm with a radius 1.

resolution. There are three kinds of pileup depending on the time that the particles hit the detector. The in-time pileup (IT) refers to particles from separate events hitting the detectors. This type of pileup leads to associating particles from different events to the same event. The out-of-time pileup (OOT) can be separated in to two categories. The early out-of-time refers to energy left in the calorimeters from previous events and the late out-of-time pileup refers to energy from later events. This phenomenon, leads to wrong energy measurements for each particle.

In order to account for this, CMS uses special algorithms developed for that purpose [12]. The one that was widely used in CMS during Run-I was the charged hadron subtraction (CHS). It relies on the knowledge of the position of the primary vertex. It removes all the charged particles that have a track that is not pointing to the primary vertex. Its main advantage is that it removes the pileup particles before the reconstruction of the jets. On the other hand since it relies on the knowledge of the position of the various vertices it means that it can only be applied in the area of the detector which is covered by the tracker.

Another widely used algorithm is the pileup per particle identification (PUPPI) algorithm [13]. This was also widely used during Run-II. It works on an event by event basis. It calculates the probability that each particle originated from the primary vertex and assigns a weight to each particle based on it. The weight is

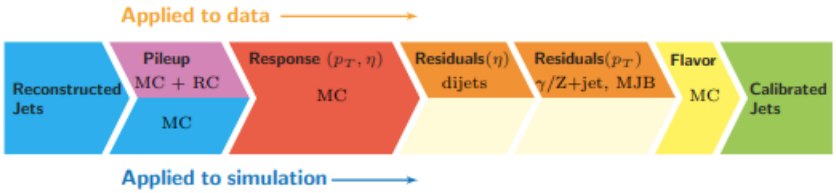

Figure 4.4: Consecutive stages of Jet Energy Corrections (JEC), for data and MC simulation. All corrections marked with MC are derived from simulation studies, RC stands for random cone, and MJB refers to the analysis of multijet events [14].

calculated using the energy distribution around the particle as well as information from the tracker. Using this weight the four-momentum of each particle is scaled before clustering the jets.

The jet energy measured with the CMS detector does not match the true value of the jet's energy. Different effects account for this difference in the measurement and they need to be corrected. To do so, CMS applies a set of corrections [14] to the energy measured. These corrections are applied in a set of steps. Each step of corrections is independent of each other and can be applied separately. But they are derived in a consecutive manner and need to be applied in the correct order. The first level (L1) of corrections, called offset corrections, is used to remove energy that was accounted to the jet that is not part of the jet and usually comes from pileup. The next two levels, (L2) relative and (L3) absolute correction,s are applied as a function of p_t and η and take into account for the difference in response between the various subdetector systems of CMS. Finally it has been observed that this procedure yields different results in data and simulation. An extra set of corrections is applied only in data in order to make sure that the result of the corrections is the same. In Figure 4.4 there is a diagram explaining the various corrections explained in the previous paragraph.

Another important step in the object reconstruction process if the determination of the flavour of the jets, mainly the flavour of hadron jets. The algorithms used in reconstruction are able to identify:

1. b-jets: jets coming from the hadronization of b-quarks
2. c-jets: jets coming from the hadronization of c-quarks
3. light jets: jets that do not contain b or c quarks are called light flavor (LF) jets. This means the are from u,d or s quarks
4. pile up jets: the rest of the jets

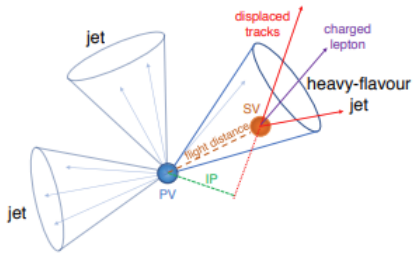

Figure 4.5: Graphical representation of the impact parameter and the identification of a secondary vertex [15]. The IP is the distance of closest approach between a particle track and the primary vertex. It is used to detect displaced track that do not come from the primary vertex.

The correct identification of the jet flavour is important as many analysis including this one heavily rely on the identification of final states that contain jets of one of these flavours.

Due to the importance of the identification of these jets, CMS uses specifically developed algorithms for this task. During the hadronization of the b or c quarks, short-lived hadrons are produced that live for around 1 ps. Depending on their p_T they travel a distance of a few mm to 1 cm. This means that when they decay they produce particles whose tracks do not point to the primary vertex but form another one that is called a secondary vertex. The tracks for these new particles are "displaced" compared to the PV. This displacement can be measured using the impact parameter (IP) which is the distance of closest approach between a particle track and the primary vertex. Another source of information that can lead in the distinction of a b or c jet from a LF one is the relative higher mass of the first ones. In Figure 4.5 the various properties discussed in the previous paragraph can be seen.

The algorithms used for the identification of HF jets are called taggers. In particular the ones used for the identification of b jets are called b-taggers. They usually rely on the use of the principles and variables discussed before combining them with machine learning in a multivariate analysis approach. The classifiers provide a set of working points (WP) that have a particular efficiency and miss-tagging probability.

Regarding b-tagging in CMS during the Run-I the Jet Probability (JP) and Combined Secondary Vertex (CSV) algorithms were used [16]. During Run-II the JP was used only for calibration and CSV was improved in the the CSVv2 algorithm. It combines more information in a different way in order to overcome limitations of

the version of the earlier version of the algorithm. In 2017 and while this analysis was under going the CSVv2 algorithm was enhanced even more using deep learning techniques. This new algorithm was named DeepCSV. It uses the same information with CSVv2 but combines the information in a deep neural network with even more layers. The outcome is four independent probabilities P(b/bb/c/udsg). The first two P(b) and P(bb) are defined as the probability that the jet contains one or two b quarks. P(c) is defined as the probability that the jet contains one c quark and finally P(udsg) is defined as the probability that the jet con taints none of the above. When the sum of P(b) and P(bb) are above a given value, the working point, the jet is characterized as a b-jet.

4.3.4 Missing Transverse Energy (MET)

CMS is designed to be hermetically closed as mentioned before. This is done because it needs to detect all particles produced by the pp collisions. The are however particles like neutrinos and other theoretical particles that do not interact with the detector. These particles are in-directly detected by measuring the imbalance in transverse energy. This imbalance is accounted to the production of particles of one of these types. The missing transverse energy (MET, Missing E_T) is defined as the negative vectorial sum of the transverse momentum (p_T) of all the particles reconstructed with the particle flow algorithm.

$$\vec{E}_T^{miss} = -\sum_{PF} \vec{p}_T$$

The term $\vec{p}_T$ includes all the reconstructed objects but also the energy of the candidate particles that were not clustered in a physics object. Analyses that include neutrinos or that search for new hypothetical non-interacting particles are very sensitive to the measurement of MET. The measurement of p_T is very sensitive and delicate process since it combines information from various subdetector systems. It can be affected by noisy sensors or electronics and dead areas in the detector as well as from the efficiency of the object reconstruction. CMS has performed extensive studies [17] to ensure the correct performance in the calculation of MET. To mitigate for any effect caused by what was described above the MET is actually calculation using the following formula

$$\vec{E}_T = \vec{p}_T^{miss,uncorrected} - \sum_{jets} (\vec{p}_T^{corrected} - \vec{p}_T^{uncorrected})$$

where $\vec{p}_T^{corrected}$ and $\vec{p}_T^{uncorrected}$ are the transverse momentum before and after applying the jet energy corrections respectively. Finally in order to avoid effects from

pile up only jets with a p_T of more than 15 GeV are used when calculating the sum.

Chapter 5

Control Systems

This chapter focuses on the description of the concepts used in the design and implementations of control systems such as the Detector Control System (DCS) used in CMS as well as the tools developed and used at CERN for creating control systems.

5.1 SCADA systems

Systems like the DCS used in CMS belong to a wider category called Supervisory Control And Data Acquisition (SCADA) systems. SCADA as a term is widely used for describing systems that monitor and control remote processes. The term data acquisition in the context of a SCADA should not be confused with the one used in a physics experiment (see Section 3.3.6), which refers to the data taking for physical purposes. It implies the reading of values regarding the monitoring of the system, which in most cases is done in a very low rate compared to a data acquisition (DAQ) system. That is why in many cases SCADA systems are also called Slow Control Systems. These systems are used extensively in the industry as well as many other fields apart from physics experiments. In general a control system consists of three layers:

- The **supervisory layer** which is responsible for visualizing and controlling. This is the part of the system that the user interacts with.

- The **front end layer** that is directly connected to the hardware and reads signals from it sending them to the supervisory layer.

- The **communication layer** that is intermediate layer between the supervisory and the front end layer and is responsible for the communication between the other two. It needs to convert signals coming from the hardware to human-readable form and signals coming from the user in to machine-readable form.

SCADA systems are usually not complete control systems but are software toolkits used in order to develop the supervisory layer of a control system. Most of the systems come with built-in ways to communicate with many types of hardware. In that sense they provide the ability to create and implement the communication layer. As the name states, a SCADA system should be able to perform three individual but complementary tasks.

- Supervise the process. To do that it needs to be aware of its exact state at all times. This is achieved by communicating with the hardware and constantly exchanging information with it.
- Control the systems. This is done by sending commands in order to manipulate the system's state.
- Acquire data. This is done for determining the system's state but also for examining its behavior over time.

5.2 The supervisory layer

A SCADA system should meet the following requirements:

- Collect data in order to be able to determine the state of the system at any time.
- Monitor the system at all times and store the information gathered from the monitoring procedure to a database.
- Flexibility in the sense of programming and configuring supervising procedures.
- Having a user friendly Graphical User Interface (GUI) so that the system can be easily interfaced by the user/operator.
- Alert and reporting mechanism so that the user can be informed of emergency situations

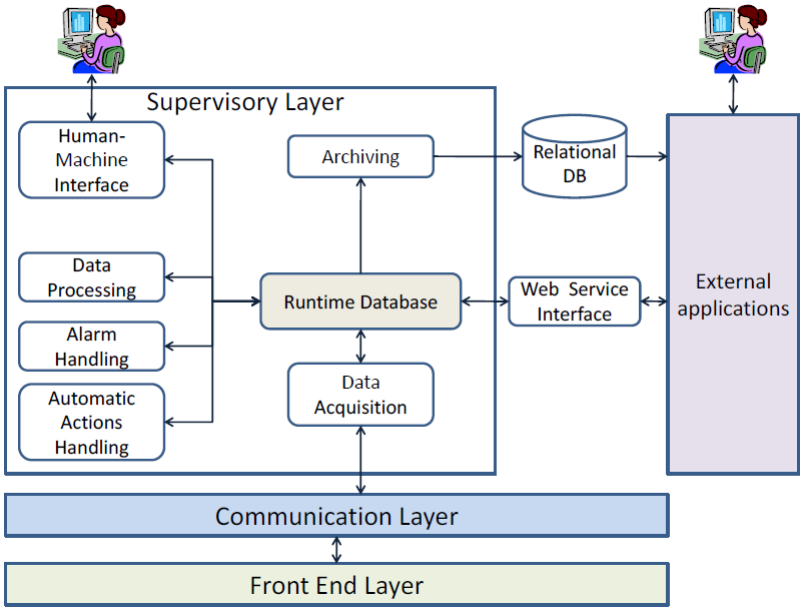

Figure 5.1: A schematic representation of a control system. The various components of the Supervisory layer can be seen in more detail.

- Being able to interact with external applications

A more detailed representation of the structure of a control system can be seen in Figure 5.1. The data collection needs of the system are met by what is called the runtime database. It is responsible for holding all the information about the system structure, updating it constantly and having it accessible to the rest of the system's processes. Each element of the system that can hold data is represented as a point. The database must provide an identifier, a timestamp and a quality flag for each one of these data points as well as holding the last value of this point.

The database is updated in an event driven way. The value of a point is updated only when an event (i.e. a value change) occurs. On top of that there is a smoothing mechanism which provides the event definition for this particular point. Its most simple form is old/new comparison but has the possibility of more complex filtering by defining a dead band or a time interval.

In addition to the run time database, the system must have a way to store all the collected data in order to determine the evolution of the process because the run time database can only store the current value. This is usually done through a relational database. The contents of this database are used for offline analysis, so access speed is not a issue. The features of the run time database

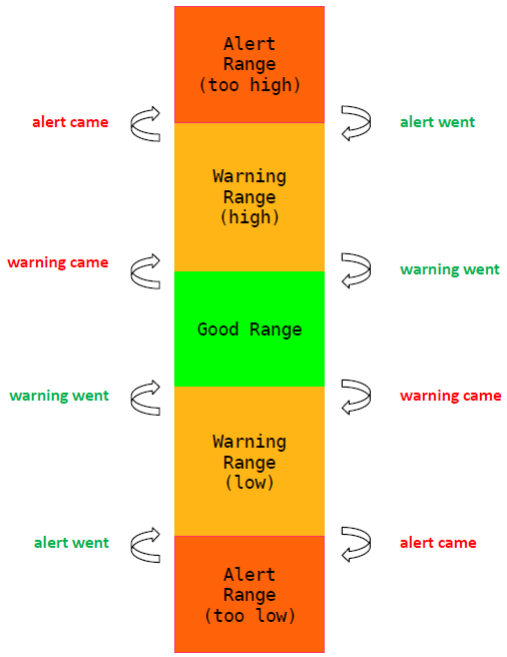

Figure 5.2: Example of the alarm ranges defined for a data point. Alerts are defined by specifying valid ranges for the values of data points. Each of the ranges is assigned with a priority, with the normal range having priority 0. The alarm mechanism is implemented with transitions between the defined ranges.

(events/smoothing) are also used here in order to control the amount of stored data.

Another part of the supervisory layer is the so called Human-Machine Interface (HMI). It exposes the state of the system and allows the user to interact with it by sending commands acting as an middleman between the operator and the process. The interface should be simple, responsive and user-friendly. It must be intuitive and provide confidence to the operators that human errors are avoided. The available actions need to be straight forward and feedback from the performed actions needs to be clear.

Following these concepts, schematic representations are preferred over text e.g. a coloured LED is better than a text field in indicating the state of a machine. Color conventions should also be used so that the information displayed to the operator is transparent. The use of trends displaying the evolution of the process is also useful in order to allow the operator to predict and identify evolution of the process over time.

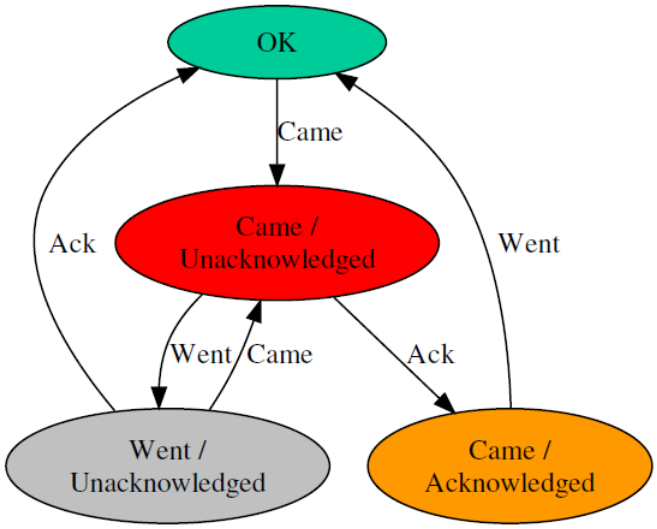

Figure 5.3: A schematic representation of all the possible states and transitions of the alarm mechanism of a control system. Two types of transitions exist the CAME and the WENT. CAME indicates a transition from a good state or a state with low priority to a state with higher priority while WENT indicate the opposite. Most alerts have to be acknowledged. If an alert WENT before the operator acknowledged it, it is indicated as a WENT/Unacknowledged alert.

A role distinction system is also needed. Experts and normal users/operators should have access to different kind of information. An expert will be able to distinguish the state of the system in a complex interface or identify a pattern in a complex trend allowing him to determine the current status of the process. In addition the available commands should be restricted for a normal user while an expert should have more freedom to perform actions that are more invasive in the procedure of the monitored process, allowing him to identify and/or fix issues.

Another key feature for a SCADA system is the alarm handling mechanism. Alarms are notifications that are used to inform the user for an abnormal behavior in the monitored process so that he can intervene and restore the activity. Alerts are defined by specifying valid ranges for the values of the system's data points (See Figure 5.2). Each one of the ranges is assigned with a priority number indicating the severity of the event occurred. The normal range of the value has a priority 0. The alarm mechanism is implemented with transitions. Two types of transitions exist, the CAME and the WENT transition. The CAME transition indicates that the value of a data point shifted from a good state or a state with

low priority to a state with a higher one. In contrary the WENT one indicates the that the value changed to a state with a lower priority number. Most of the alarms must be acknowledged by the operator. This implies that if an alarm CAME and WENT before the operator acknowledged it, the system stays in alarm state but is indicated as a WENT/Unacknowledged so that the information about the alarm is not lost. A schematic representation of all the possible states and transitions can be seen in Figure 5.3.

Finally the SCADA system should be able to communicate with external applications. It should be able to expose its data to other pieces of software. This can happen whe another piece of software will be responsible for manipulating the data and the SCADA will be used for for monitoring and controlling the process. One way of achieving this is by storing the data in a relational database which can then be accesses by other applications.

5.3 The front end layer

The front end layer is the part of the control system that is directly connected to the hardware. The hardware can consist of general-purpose devices such as sensors (temperature, pressure, humidity etc.) as well as application specific devices like power supplies, pipes, pumps, fans, motors and many others. One of the most common device type used in control systems are Programmable Logic Controllers (PLCs). They are very robust, low level computers. They work by making calculations based on input signals coming from sensors and outputting the result of the calculation to the rest of the hardware. Because of their robustness they are widely used for safety purposes, significantly decreasing the response time of a safety action. In very complex systems during an emergency the appropriate command might not be straight forward for the operator and the use of PLCs helps avoiding human error.

5.4 The communication layer

As it has already been mentioned, this part of the control system is responsible for the communication between the hardware (front-end layer) and the supervisory layer. In the past the supervisory layer was built using software spesific for each hardware type. This was due to the need of drivers in order to interact with the hardware. This made the procedure of creating a generic control system software a very hard task. In the past 20 years there has been an effort to standard-

ize the communication procedure by making it hardware independent and using widespread protocols such as TCP/IP or UDP/IP. This has led to the creation of communication standards for control systems such as OLE for Process Control (OPC) protocol, Modbus and S7.

The OPC protocol is a set of standards created together by the hardware and software vendors and acts as a middle man between the hardware and the software. It is based in the client server logic and is actually a translator between two protocols. The hardware vendor is responsible for providing a server. The server translates information in to a form that the hardware can understand. On the other hand it converts information coming from the hardware to an OPC format. The software vendors have to embed an OPC client in their product. The client is responsible for translating information coming from the software to an OPC format that the server understands and thus send it to the hardware. In the opposite direction it needs to translate OPC formatted information, coming from the server (initially from the hardware) to a format that the software understands. In this way there is no need for the development of drivers and the SCADA developer can communicate with the hardware by simply configuring the OPC server and client.

Modbus and S7 are communication protocols used mainly with PLCs. They are vendor specific. Siemens PLCs use the S7 protocol while Schneider ones use Modbus. At CERN also CAN buses are used. This is due to the extreme conditions under which the hardware has to operate. In strong magnetic field and high rate radiation environment the use of Ethernet based digital protocols is not possible. CAN is an analog protocol and offers a good alternative. In many cases control systems are made of many smaller individual components that need to communicate with each other. One example is the communication between the LHC and the detectors where the accelerator needs to be aware of the state of each detector and vise versa. A protocol called DIP (Data Interchange Protocol), developed at CERN is used for this purpose. It is based on the public subscribe pattern and is used widely in CERN built applications in control systems but also in others.

5.5 WinCC_OA SCADA Toolkit

In order to select the product that will be used for building control systems CERN conducted a market survey in the end of the 90's. As a result PVSS-II[1] was

[1] PVSS is an acronym for Prozess Visualisierungs und Steuerungs System, which means Process Visualization and Control System

selected. It is a SCADA toolkit developed by the Austrian company ETM. Later, ETM was purchased by Siemens which renamed PVSS into WinCC_OA[2]. Siemens is now the vendor of the toolkit. In the following sections, the structure of WinCC as well as a set of factors that played a major role in the selection procedure will be listed.

5.5.1 Main Features

One of the main features on WinCC that made CERN select it as the tool to develop its SCADA system was its ability to scale. Many of the other available SCADA solutions have a limit in the number of items they are able to control. WinCC has no limit since it offers the ability to develop a distributed system. It natively offers the ability to connect different systems. This way a large system can be splitted into several smaller ones that together form the control system. This is very important as it allows for the distribution of the computational load making it possible to design and develop really large systems like the oned needed for the operation of the LHC and its detectors. Extensive tests were made at CERN in that direction in order to test the limit of WinCC. They found that it was possible to connect more than 100 systems in order to form one very large distributed system. The way distribution is implemented in WinCC is that all connected systems are automatically updated when something changes in one of them. This allows the user to modify and develop further tools for one of the systems since it is ensured that if one component is deployed in one of the them it will not interfere with the structure of the whole system.

Another advantage of WinCC is its run time database. It comes with an internal database which consists of structures called data points (DPs) (See Figure 5.4). A data point is a structure made of individual elements, called data point elements (DPE) which can hold values. Each data point is an instance of a certain type, called data point type (DPT) (See Figure 5.5). The DPT defines the general structure of the data point as well as the name and type of its elements. Data points can be used to simulate devices that are connected to WinCC but can also be used for modeling more abstract types. The concept is similar to the one used in Object Oriented Programming (OOP). A DPT is a class, a DP is an object of that class and a DPE is a field of that class.

The advantage of using this philosophy for the run time database is that data that are associated with a device can be grouped together and are not held in separate variables as it is done in other SCADA systems. This way the re usability

[2]Short name for WinCC Open Architecture

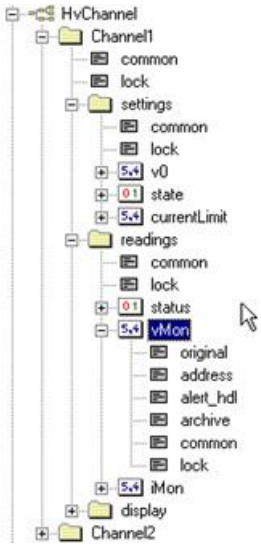

Figure 5.4: A schematic representation of a data point. A data point is an abstract structure holding data. It consists of data point elements that are of a certain type and have a value. Each data point is an instance of data point type which specifies its structure.

of code and components which are developed for this type of device is enabled. Although WinCC is not object-oriented (OO), the structure of its run time database can be thought as this. A type that represents to the structure of the data is created, similar to a class. Then instances of this type which correspond to actual devices are being created like objects of a particular class are created. Finally data point elements are part of the data point like are class fields in OO. WinCC also provides the ability to create generic user interface components (reference panels). This way, graphics can be designed for a data point type and instantiated for each data point.

Another advantage of WinCC is that it can be easily extended through the native language that the toolkit provides. This programming language is called ConTRoL (CTRL) and it has been used at CERN to develop a set of libraries (See section 5.6) that can be used as a basis for the design and development of control systems. In addition, there is a custom WinCC Application Programming Interface (API) which can be used to write WinCC managers and drivers[3] extending it so that it can meet the needs of the user.

CTRL is a C-like language and is used for creating scripts and user interfaces

[3]The concept of managers and drivers in WinCC will be explained in the next section.

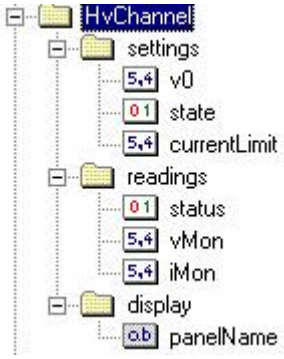

Figure 5.5: A schematic representation of a data point type. Data point types specify the structure of each data point that is created and is of that type. They are like blueprints and are used to model devices in an abstract data point of view. For example a power supply can be specified as a data point with two values, representing the voltage and the current it supplies to the connected devices.

(panels). The advantage of this language is that it is user friendly and easy to learn as it follows the general rules of a programming language. This way it is easy for new users to get familiar with the toolkit being able to contribute in large projects quickly. Additionally, people that are not that familiar with programming as software engineers are, can develop control systems easier than in other SCADA toolkits.

Another major advantage of WinCC, in contrast to most SCADA products, is that it is multi-platform. It can offer the same functionality in Windows as well as Linux. This is a very important feature, since Linux is very popular in the scientific community and a lot of software, like DAQ, are usually developed and operated under Linux. This allows running the DCS (or some part of it) and the DAQ software under the same machine. Although this is not entirely true, since some of the communication protocols like OPC or in general external application force the use of Windows, all the native WinCC functionality can be used in both operating systems.

5.5.2 Structure

WinCC is not a control system on its own but rather a set of tools that can be used to build one. It provides most of the features needed to create a robust SCADA system, like a well structured run time database, modularity and scalability as it allows for the connection of many individual projects to create a bigger

one. The main philosophy behind its architecture is modularity. Each function of the system is implemented by an independent process which is called a manager. Managers work together by communicating with each other forming the whole system. A figure of the structure of a WinCC project with all the individual managers can be seen in Figure 5.6.

The core of the system is the Event Manager (EV). Only one event manager is allowed per system and it is responsible for the communication between all the other managers as they can not directly interface each other. It operates by receiving and evaluating inputs (events) from other managers and distributes these events to the rest managers. It contains the current image of the system (run time database) in memory and it is the way all the other managers have in order to access the system's image. Managers can get the values of data points by interacting with the event manager. They can also update the values of data points. Finally they can subscribe to a specific data point and be notified by the event manager whenever there is a change to the subscribed data point.

Another very important manager is the database manager (DM or DB). This manager reflects the current status of the monitored process by keeping in memory the current image of the database. While the EV holds the structure of the database, this manager keeps the values of the particular elements in the database which at the end determines the status of the project. This manager also keeps in memory the latest alarms so that a complete picture of the system is readily available to the other managers. Additionally it is responsible for archiving the data acquired from the monitored process. WinCC uses a RAIMA (file based) database as the internal run time database. It is highly optimised for read and writes allowing quick access and update to the systems's image. For value archiving, WinCC provides two possibilities. The first one is a local file storage. The second one is the use of a Relational DataBase (RDB). At the moment only Oracle databases are supported but in the future more database types will be included.

User Interface (UI) managers are responsible for running the HMI processes. They work by executing panels and projecting them to the screen so that the user can have a graphical representation of the monitored process. A special panel that is called Graphics Editor (GEDI) is also available. This panel is used as an Integrated Development Environment (IDE) for the development of other panels, various graphical objects and also scripts and libraries. Control managers (CTRL) work together with UI managers by running scripts in the background written in WinCC's native language CTRL.

A big advantage of this toolkit is that it comes with an API allowing users to develop their own custom managers in C++ or CSharp. The API manager runs

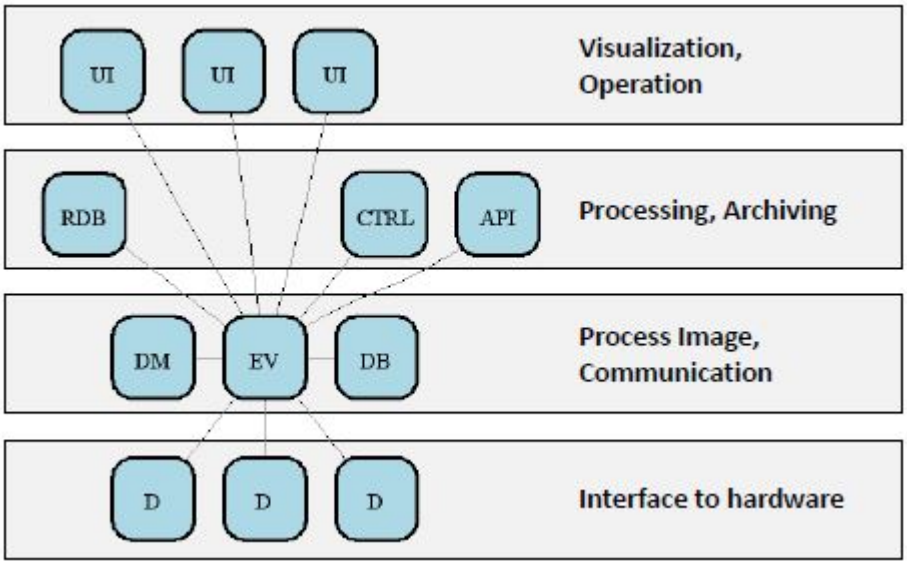

Figure 5.6: The structure of a system built with WinCC_OA. Each part of the system is an individual process. All the processes communicate with each other to form the entire system providing great modularity and scalability.

custom code developed byt the user implementing new functionality and extending the standard WinCC one.

In order to interface the hardware WinCC provides a special type of managers, called drivers (D). They are responsible for handing the different communication protocols in WinCC. The standard ones support protocols such as TCP/IP, Modbus, ProfiBUs, CANbus, OPC, S7 and others. On top of that the user has the ability to write its own drivers and extend the poll of hardware that can be integrated in the system. At CERN drivers have been developed for the DIP and DIM protocols.

Through WinCC different kind of systems can be created. The two most important ones are distributed and scattered. They provide WinCC with the ability to scale. A scattered system is one that has no event or data manager. This can be done as WinCC allows a manager to connect to a remote event/data. Scattered systems are a collection of graphics and control managers with no system image (runtime database). They connect to remote data/event managers for getting the image. A distributed system (See Figure 5.7) can be connected with other control systems so that together they form a bigger one. For this to be possible WinCC provides a special type of manager, the Distributed Manager (Dist). The data of

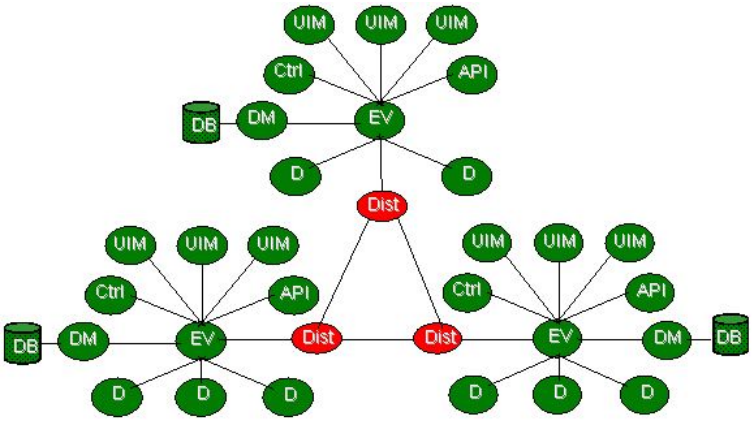

Figure 5.7: Picture of a distributed control system. Each WinCC system, features a distribution managers. This manager can be used to connect WinCC systems together to create a large control system.

both systems are available at each time and can be accessed by adding the system name before the data point name so that the EV knows where to look for the data. This way, division of concerns can be achieved by splitting the responsibilities between systems.

5.6 Joint COntrol Project Framework (JCOP)

The Joint COntrols Project, was created in order to provide the tools for all the experiments at the LHC to develop control systems using a common basis. During the LEP experiments it was made clear that there was lack of standardization in terms of hardware as well as in terms of software. This made the integration and the maintenance of the various projects a very hard task. As a result it was decided that CERN would rely as much as possible in commercial products for hardware and software. The goal was to create a set of tools so that the effort for developing and maintaining and integrating projects would be reduced to minimum. As a result a framework was created. It contains a set of libraries, panels as well as other components that can be extended by the developers. All the basic tools to create a control system are already available and since all systems are built using the same basis they can easily be integrated. The framework is built as a set of components each one having a different functionality allowing the developer to use the ones that he needs in order to develop his system. It provides a higher

layer of abstraction as the developer interacts with the framework and not with WinCC directly. Although interaction with WinCC is also available since not all functionality is covered by the framework.

5.6.1 Hardware and Logical view

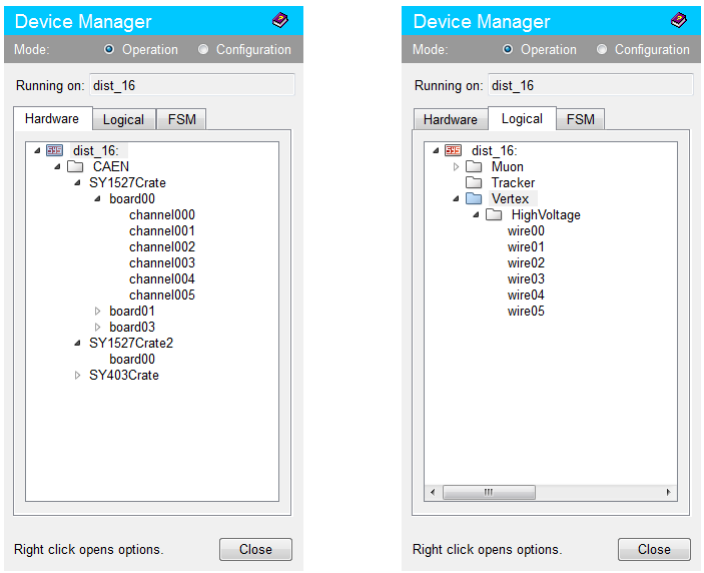

Figure 5.8: The JCOP framework Device Editor Navigator (DEN)

The framework comes with a built in way to organize the connected devices called Device Editor Navigator (DEN). The organization is composed of two parts, the hardware and the logical view (See Figure 5.8). The hardware view is mandatory while the logical is not. The hardware view corresponds to the actual configuration of the hardware (power supplies, crates, mainframes etc) and the devices connected to them. It is designed in such a way that corresponds to the actual hierarchical connection of the hardware (mainframes contain board that contain channels). The framework also provides built in structures (DPT) to represent the most common hardware used in the experiments so that the development of different structures (DPT) for the same hardware is avoided because it could result in conflicts while integrating various systems. The logical view is implemented by providing a more descriptive name (alias) to the data points contained in the hardware view. It is used in order to organize the detector in a more user friendly way. A mainframe name may not be very meaningful to the user, but the part of

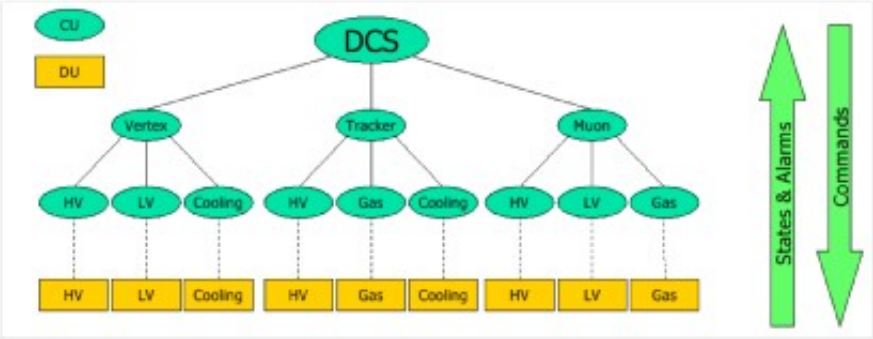

Figure 5.9: A sample FSM diagram created using the JCOP framework component. Bottom nodes represent actual hardware while the intermediate ones are used to represent the process in a hierarchical way. In the FSM approach, states propagate from bottom up while commands the other way around.

the detector that this piece of hardware controls is much more straightforward for the user in order to operate the detector.

5.6.2 Finite State Machine

The best way for implementing large control systems is through the Finite State Machine approach. As the name states, through this approach the whole system is represented as a machine which has finite number of states in which it can be. This model is used in many fields and its definition and implementation may differ from field to field. The JCOP framework comes with a native tool which allows the use of the finite state machine model for building a control system. It acts as a bridge between the SCADA system and the State Management Interface (SMI++) which is a framework for building FSMs and was developed at CERN for the DELPHI experiment at LEP.

The SMI++ framework provides the ability to create the hierarchical representation of the experiment in order to control it. The structure is created as a tree of nodes (parent-children). Each object of the hierarchy (node) is a domain which is a separate process responsible for managing its sub-nodes. The lower nodes represent the actual hardware and are not a separate process but operate within the parent node to which they belong. The inner nodes are not coupled to actual hardware but provide an abstract way of representing the structure of the experiment. In the SCADA part, the same structure is implemented through data points. The FSM data points that correspond to devices are directly connected to the SMI++

and are used to calculate the state of the FSM. This two processes interact via the WinCC API. All the SMI[4] domains are connected and form the whole experiment which can be controlled by giving commands to the top (supervisor) node. During the operation of the FSM, states are propagated from the lower part upwards while commands move the other way around (See Figure 5.9).

The JCOP tool for the FSM implementation offers three types of nodes. The type specifies among others the place that each of the nodes has in the FSM hierarchy.

- **Control Units (CU)** are internal nodes, used to represent the hierarchical tree. Each CU runs as a separate SMI domain.
- **Logical Units (LU)** are also internal nodes, but they run within the process of the CU they are included.
- **Device Units (DU)** they are the "leaves" of the tree, they represent a device and they are connected to a WinCC data point that is directly connected to a device.

Additionally, each node has another kind of "type" called FSM type. It defines the possible states that every node can be in and the actions that can be performed in each of the states.

As mentioned, DUs are directly connected to the SMI and are used in order to compute the state of the FSM and at the same time send commands from the FSM to the actual hardware. For each one of the types an initial state is declared. A set of rules is used in order to compute the new state in case of a change and transition to it. The actions are implemented as a sequence of commands and can be sent to all children of a node or to all children of a specific type. Each child executes the action and then reports back its state forcing the parent process to re-compute its state.

A large experiment, like CMS, is composed of different sub-detectors and partitions. In order to operate smoothly, a flexible partitioning mechanism is needed. This means that different parts of the experiment need to be operated separately or even part of the experiment should be excluded at a given moment.

To cover this need, the JCOP framework uses the concept of ownership in the FSM. Each part (domain) of the tree has an owner. The operator can own the whole tree or part of it. A shared mode is also available in which the operator takes control of the whole tree and can then share part of it with one or more

[4]SMI++ referes to the framework used to build the FSM, while SMI referes to processes that are written in the SMI language provided by the SMI++ framework.

users. While this mode is active, he is still the owner but someone else can also send commands to this part of the tree. Nodes can also be excluded from the tree meaning that they are not taken into account in the computation of the FSM state and commands are not send to them. Only the owner can exclude or share a part of the tree. In addition, LUs and DUs can be disabled which has the same effect as excluding a node. The difference is that since they are not a separate domain a different user can not take control of them. They can only be owned by the owner of the CU that is above them in the hierarchy.

Different parts of the tree can be controlled by different users, a feature that is very useful in times of testing or failure. If a part of the experiment is in error state the operator can exclude this part of the tree and an expert can take control of it is being fixed. After the expert is done fixing the error, the operator can re-include this part of the tree in the main FSM.

5.6.3 Archiving Database

A very important role of the DCS is to be able to archive data regarding the state of the system e.g. environmental conditions etc. These data are used not only to determine the behaviour of the detector but are also used in order to determine the quality of the data collected from the experiment. If a part of the experiment was malfunctioning at a certain period maybe the data collected can not be used for physics analysis and have to be discarded.

For this reason, WinCC provides a way to store data in a file structure locally in the computer that the system is running. This may be enough for industrial purposes or smaller systems, but is not for an experiment as big as the ones that take place at CERN. Instead it was decided to rely on Oracle databases for the archiving procedure. Although WinCC comes with an native RDB manager which allows a system to connect with an Oracle database for the archiving purpose this was not enough for CERN as the amount of data collected from the detectors is huge.

The archiving policy of the CMS experiment is to archive data on change. This means that whenever a change is made values are archived. The other option would be to archive data using a polling mechanism. This means that a value would be archived at a certain time no matter weather its value is changed or not. The ability to apply some time interval and/or a deadband[5] can be also defined.

[5]A deadband is defined an interval (band) in which no action occurs. It can be an absolute value or a percentage of the value and if the change is lower than it then no change is registered in the system.

These conditions can be applied independently on the data, meaning that if one of them is true values are archived, or all together and archive data when all of them are true.

5.6.4 Configuration Database

In large control systems, such as the one controlling the CMS detector, the settings of the detector have to be changed corresponding to the mode of operation. That is, changing the values of the power supplies, the accepted temperatures of a PLC and many more. More over in the event of a failure concerning the DCS PCs, the system has to be up and running within a reasonable time.

To address these issues, JCOP provides a tool which is called the configuration database. The configuration database is actually an Oracle database that is used to store the system's image. This database is able to store two types of data.

- **Static Data** which correspond to the image of the system. This means data points, addresses, aliases archiving settings etc.
- **Dynamic Data** which consist of values and alert limits of data points that are connected to the hardware.

In the event of a failure or a change in the system's image the first type of data is used. This allows to restore the system in to a specific state, since the system image is stored in a database. Also under change (removal/addition) of hardware the database is used again in order to update the current image of the system.

Dynamic data are used to configure parameters of the hardware and are called "recipes" in the JCOP terminology. Recipes are used to change the detector's state from one mode of operation to another (Physics, Cosmics etc), as well as to different states in a particular mode (On, Off, Standby etc.). Apart from the Oracle DB, for the recipes mechanism an internal cache is also provided where recipes are stored in the computer's memory. This ensures that the detector settings are set as fast as possible which is crucial in times when a quick operation is required.

5.6.5 Access Control Mechanism

Another really important feature of the JCOP framework is access control. It is implemented in the UI level. It is not meant to protect the system from external threats but rather to distinguish the roles of the various users by allowing different actions to each one of them depending on their expertise. This way, mistakes can be avoided by actions performed from unauthorized and not experienced users.

It defines domains which corresponding to a part of the system and each one of them has levels of access (e.g. operate, modify etc.). A domain can represent for example a subdetector. Inside each domain, roles are defined (e.g. operator, expert etc.) allowing a set of actions for each role of this particular domain. Users that need to perform specific operation are assigned to different roles (e.g. ECAL Expert, HCAL Operator etc.). Users can be assigned to several roles at a time.

Chapter 6

The CMS Detector Control System

The Detector control system of CMS is the main interface between the operators and the actual detector. It is used to monitor and control the detector and other peripheral systems of the CMS experiment. This chapter discusses the design and implementation of this system. During this thesis I contributed greatly in the maintenance and improvement of this system. I was also involved in day to day operations regarding the running of the detector but also in the process of solving problems that were encountered during operation.

6.1 Introduction

The main purpose of the detector control system is to bring the CMS detector in a state where it is able to record physics data, in a safe and controlled manner. To achieve this, million of parameters are constantly monitored. These parameters are also recorded and used offline. A first use is as a rough estimate of the quality of physics data. If a part of the detector was not working or was misbehaving at a particular instance in time then the data recorded at that time may not be suitable for physics analysis. These data however are always valuable as they can be used to determine cause of the problem but also for calibration and performance measurements of other parts of the detector. Additionally the data recorder by the DCS are also used in the actual reconstruction of the physics objects. Each sub-detector system behaves differently depending on the environmental conditions. These conditions are monitored by the dcs and their recorded values are used to determine and calibrate the detector's response.

In this part it is important to clarify that the DCS is not a safety system. Although the word safe was mentioned before its role is not to protect the detector

and its underlying hardware. Although it provides information that can be used to take preventive and safety actions by the operators and some actions are taken by the system itself it is not a safety system. In other LHC experiments the Detector Safety System (DSS) and the DCS are integrated together. This is not the case in the CMS experiment. A dedicated safety system exists. It is entirely hardware based and its sole responsibility is to ensure the well-being of the detector. This means that it can take automated actions starting/stopping parts of the detector in order to ensure its integrity.

6.2 Architecture

Conceptually the DCS is implemented as a distributed system. It consists of around 30 sub-systems each having a specific role. Some are generic ones and serve experiment wide functionality for example controlling the racks or being responsible for the access control mechanism. Others are dedicated to a certain task. An example of this can be the control of the high voltage of the ECAL. Each of these systems is implemented as a redundant system meaning that two copies exist for every one of them. These two systems are identical and communicate with each other at all times. One is the active one while the other is in standby mode. In case the active system has a failure the passive, standby one, takes its position in the system. To realise the system 60 blade servers are used, two for each system. One running the active peer and one the passive peer. Apart from running the two systems in physically different machines, the servers are also geographically located in different places. This ensures that there is fail-over even in case of a natural disaster were physically the servers are damaged. The CMS has two server rooms each hosting many servers. One is underground while the other is overground. The 30 DCS servers are in the underground server room while the other 30 are overground.

Utilising the JCOP paradigm the control system in CMS is not a single application. It is a set of small distinct installable software entities each responsible for providing some functionality. These are called components. Each of the 30 systems mentioned above start from the exact same point/configuration i.e. a plain WinCC OA project. Then depending on the role of the project in the system a set of components are installed giving it its final form. The DCS of CMS consists of around 200 components. Some are general and installed everywhere while others provide specific functionality and are installed in a single system.

Following the structure of the DCS itself the DCS community is structured in a similar manner. There is a central team that is responsible for all the generic

DCS related things in the collaboration and there is usually one team for each subsystem in charge of developing and maintaining the sub-detector related system. The maintenance of the servers and the wincc installations also falls under the scope of the central team. It is also responsible for developing and providing components with generic functionality used throughout the experiment's control system. The central team provides in a sense a centralized batch-like service where the various sub-detector groups can run their control systems. Only the central team members have access to the control system infrastructure but for maintenance purposes. The various subsystems take their form by installing components. The configuration of each system is decided by the team responsible for this system. The central team creates the system and is responsible for its maintenance but which components are installed is decided by the corresponding sub-detector team.

The DCS of CMS is designed, structured and operated as an FSM using a hierarchical organisation. It uses the SMI++ toolkit via the JCOP framework as discussed in chapter 5. Through the FSM the detector is modeled in a more physics related view and thus can be operated by non-technical personnel. This design was done in a way were the system complexity would be hidden from the operator. The main requirement was that it would be possible for non-experts to be able to operate the detector as the CMS collaboration does not use dedicated operators but any member can volunteer and do a shift during data taking. Even if this was not the case the system is far too complex for a single person to be able to take into account all the parameters in order to take the correct decision. For this reason only the top level parts of the detector are exposed to the operator and the system takes automated actions designed by experts of the various subsystems in order to self operate with minimal supervision be the operator.

The FSM is designed in an abstract way so that it is easy to integrate the several subdetector systems independent of their structural and/or operational details. In order to be integrated, each subsystem should expose one or more top nodes that have a specific set of states and can understand a specific set of commands. The states are:

ON: ready for data taking

STANDBY: on or partly on but not ready for data taking

OFF: Off

ERROR: Some problem occurred. Usually requires some intervention to fix the problem.

Each state provide a set of commands in particular ON, OFF, STANDBY that are used to bring the detector to the corresponding state. A schematic representation of the FSM structure with the central system and one connected sub-system can be seen in Figure 6.1.

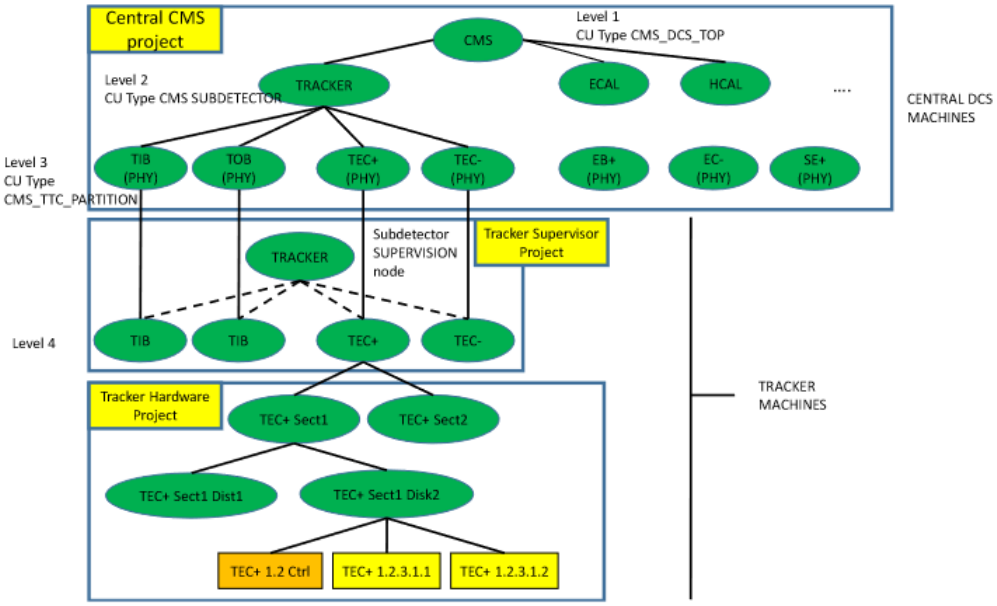

Figure 6.1: Sample schematic representation of the Finite State Machine (FSM) of the CMS experiment.

6.3 Main Features

A consequence of the previously mentioned architecture is that the systems relies a lot in the installation of components. Additionally since the production systems are only accessible by the central team the installation becomes more complex. This means that there is a need for a component installation procedure with specific guidelines and rules. This procedure should be able to be executed somehow remotely by the various users of the systems. The JCOP framework comes with a built in installation tool for its components. CMS utilises this tool extensively. On top of the installation tool JCOP offers the ability for remote installation. A specific script is used for this purpose. The script connects to a database that holds all the project information. It contains configuration options like ports and hostnames,

paths, components that need to be installed in this projects and more. The script periodically checks for inconsistencies between the projects configuration and the database and updates the project based on what is stored in the database. In addition it stores back to the database the status of the project for monitoring and administration purposes. This tool is greatly utilised in CMS.

The project setup procedure starts with the members of the central team installing some generic components that are used throughout the experiment. Then in coordination with the team of the subdetector experts they propose other additional components that fit their needs or the development of new ones is done by the subdetector group in coordination with the central team. Finally when the list of needed component for a project is complete they are targeted to the system. When newer versions of an existing component are released the corresponding database entry is updated and the project is updated automatically. In addition to this tool CMS has developed a set of scripts that allow for project creation from the contents of this database. While this tool is very useful it is mainly used in production environments since it is hard to setup in a development environment since it requires a database setup etc. The central team of CMS has performed a set of extension to the existing tool in order for it to be used by the CMS community in development environments. This is very useful as it allows for fast and easy recreation of production like projects. This way the users can create a project that is exactly like one of the production projects. This allows for easier and fast problem solving as the problem can be reproduced in a project that is exactly like the one in production and faster development.

Another widely used tool is the configuration database. It is another database that provides the ability to hold datapoints, datapoint types but also configs. It is in a sense a "copy" of the internal runtime database. This is done for backup purposes but also for deployment. One of the main steps of installing a component is modifying the runtime database, either by creating datapoint types and datapoints but also by setting values and altering datapoint configs. While wincc offers a native tool for this operation in the form of files, the mechanism is not very easy to use and has limitations. The team that developed the JCOP framework has developed a tool for this purpose. The tool is based on the configuration database that was described before. It allows saving and loading from the database. This tool is very useful and easy to use but still it was not intended for heavy use and still has some limitations. In CMS the way they system is designed, installation and interaction with the configuration database can be a daily task. The CMS developed component CMSfwInstallUtils makes this easier. It provides an easy way with filters and patterns in order to select the datapoints to be saved in the

database. It also provides a checking mechanism that can identify inconsistencies between the configuration database and the projects runtime database. This way the user can easily see what will be changed if he uploads to the database or download from it.

Another key feature of the CMS DCS is the alert system. Wincc comes with an integrated configuration for alerts. It enables the user to define alerts for existing datapoints. Moreover there is a native alert screen. That it a panel that shows all the active alerts at the current system. For distributed systems the alert screen can be configured to show all the alerts in all the connected systems. This is a very nice feature as CMS is a big distributed system and through this screen the operator can see alerts from all the systems in a single screen. In extension to this, CMS has a tool than can send extra notifications. Firstly it can generate sounds in the control room that will notify the operator in case of an alarm. This means that the operator does not have to constantly look at the screen. In addition a notification system has been created. For already defined alarms, a list of people can be defined that will be notified in case of an alarm. The notifications can be sent via email and sms. This is very useful as in cases of serious alarms experts can be notified immediately by the system so that they can intervene when necessary.

In big systems like the one described here, a lot of people are involved. Some are users or operators, some need to just browse the system to monitor values that affect some other subsystem that is of interest to them. Others are experts that need to ensure that things are running and monitor the conditions. In addition they need to understand what went wrong in cases of failure. For a system like this to work smoothly some kind of access control needs to be defined. Again Wincc comes with a native solution to this. There is the possibility to create users and groups and give access rights and privileges to them. The JCOP team has created a very nice extension to this mechanism [18]. This extensions allows Wincc to connect to a central catalog service through the LDAP protocol [19]. This allows Wincc to connect and download all users and groups that exist in the central CERN user directory. Additionally when a user can log in to Wincc his credential are authenticated against the central CERN authentication servers. Another nice feature of this tool is that one Wincc system can be registered as the master. So instead of all the systems connecting to the central CERN servers only one does. This system periodically synchronizes its contents with the CERN servers. Then it pushes all user/group information to the rest of the systems. The authentication requests are forwarded from the requesting system through the central system to the central CERN servers.

Another important feature of Wincc is redundancy. In general in IT high availability and redundancy is an important topic. It is the concept of having more than one active versions/copies of some software or hardware in order to ensure no downtime. When a component of the system fails there is another one ready in standby mode in order to take its place. Although this is a standard feature in Wincc it is not used everywhere. At CERN CMS is one of the few that use it. Since the DCS needs to operate at all times even when all the other systems are off, redundancy has proven to be very effective. The DCS of CMS has had very small downtime since this feature was used. In addition to the standard Wincc redundancy, CMS has extended it to meet its needs [20]. The tool integrates redundancy in the JCOP framework as it is not handled natively. It allows component installation in redundant systems. It also allows the handling of redundancy for communication protocols were only one connection is allowed between the hardware and the software. This is done by constantly monitoring the state of the system and starting/stopping processes in the active/passive peers.

Another really key feature of the control system in CMS are the automatic actions. They allow the user to define a set of reactive actions based on an input condition. For example switching on the cooling when the temperature exceeds a certain point. In the earlier days of the control system many actions had to be done manually either by the operator or an expert. Since the development of this tool the system's response time and operation has become way smoother. At the present moment the operator is more supervising the automated operation of the system rather than having to act on it. He mostly has to notify experts in case of a failure or perform non-standard operation when the system is in an unusual state. Automatic actions are not to be considered safety actions. They can be used as a reactive and preventive measure is case of a problem or a failure in order to bring the detector in a safer state but they should not be considered a safety mechanism. For this reason a separate system exists that is low level and entirely hardware based that acts directly on the hardware avoiding any latency or complexity that the software layers imposes.

In order for the experiments to take place the LHC needs to produce proton beams. The two beams are then collided at the center of the four detectors as discussed in detail in chapter 1. This is also true for the CMS detector. During the process of beam production and after that while the beam is accelerated to reach the desired energy the accelerator goes through various stages. Even when not colliding the beams still go through the center of the detector. During this process the conditions present inside the detector area may harm some of the subsystems. This is mainly true for the ones closer to the beam line, the Tracker and ECAL.

During times when no physics data are produced the LHC will need to do some testing or development. In order to do this it will need to circulate beams inside the beam pipes. This again maybe be harmful for some sub-detecting systems. To avoid problems during any of the situations described above, the LHC needs to take approval from all the 4 experiments in order to use the beam pipes. The LHC publishes the type of operations that it is going to do and at every stage of this operation publishes the exact step that it is in. For each of these states/steps the experts of every subsystem have studied the underlying conditions and have decided what is the safest state for their subsystem. This list has been handed to the central DCS team which has established some automatic commands that bring the particular subsystem to the desired state. For this purpose the automatic action mechanism described above is used. The full list of these actions can be found in [21].

A special use case of the previously described mechanism are the beam inject, dump and adjust operations performed by the LHC. Because these are the most potentially dangerous for the detector instead of just announcing the state change, the LHC waits for feedback from the experiments before performing the action. This mechanism is called the LHC handshake. In the case of CMS the detector is brought to a safe state and then a response is sent back to the LHC. All this mechanism is automatic without the operator having to do any manual operation.

Another important feature and functionality of the CMS DCS is its ability to communicate with external systems. This allows for the control system to receive information from other external systems and expose them for example to the operator so that everything is part of the control system. At the same time commands can be send. Also an external system can take information from the DCS in order to use it in any way desired. Some examples of this mechanism are what was described before with the handshake between the LHC and the CMS detector. Another example is the communication between the CMS DCS and the systems that are used in CMS for cooling but also for the magnet. These are external systems independent of the actual control system. Nevertheless information from these systems is displayed and used inside the control system. Another example is the communication with the data acquisition system. The DAQ uses the information from the DCS to identify when the detector is ready for data taking in order to start collecting them. It also uses the information from the DCS to determine what state was the detector during the data taking to determine the quality of the data recorder.

As discussed again above the direct access to the control system machines is prohibited. It is only available to the central dcs team members and only for main-

tenance purposes. Though access is not allowed it is desired for users operators or experts of the various systems to be able to interact with them in some way. This is done with the use of terminal servers. They are a set of windows machines that have the ability to host many sessions at the same time. This means that multiple users can connect to the at the same time. More than one machine is used first of all for redundancy purposes but also for load balancing. The round robin mechanism is used for load balancing. The machines that run the DCS application are in a separate network that connects all the CMS experiment machines. On the other hand users at CERN are connected to the main CERN network. The terminal server bridges these two network allowing users from one to connect to the other via these machines. When connected to one of the terminal servers, the user has a list of applications that he can run. The list is determined using the egroup mechanism of CERN. This way computer accounts are grouped together and get the permissions. Most available applications are Wincc oa related ones but there are also applications that give direct access to some equipment for maintenance purposes.

Another way of remote interaction with the CMS DCS is the webserver. As all communication between the various Wincc processes is done via networking and in particular the TCP/IP protocol, it is possible to use this in order to connect them natively with external applications that communicate via networking. The central DCS team in CMS utilizes this functionality and exposes the control system to the users via a web server. In particular a website has been create, `https://cmsonline.cern.ch` that hosts all this information. The web server runs on two servers for redundancy and load balancing. The servers are accessible from the CERN network and from the outside world using two gateway machines.

The server hosts around 70 web applications that allow the user to interact with the control system. Some are just exposing system information to the user while others allow him to interact with it by sending for example commands. Through the webserver it is also possible to deploy code by interacting with the installation database that was mentioned before or deploy new versions of components. The user can also browse historical data that are archived by the control system or even browser the run time database of the DCS. The applications are developed as portlets using the java enterprise edition (JEE) standards. They are hosted in a portal server using the java webcenter which runs on weblogic server. For the gateways the oracle http server (OHS) is used. The OHS is an extension to the apache http server made by Oracle. The whole system is connected to the single sign on mechanism of CERN so users can connect to the webserver using the same account for the rest of the CERN infrastructure.

Chapter 7

Differential cross section measurement

This chapter contains information regarding the physics analysis done in the context of this thesis. The analysis was the measurement of the differential cross section of the boosted $t\bar{t}$ pair where both top quarks decay hadronically. It introduces and explains the strategy used and finally presents and discusses the results. This analysis is a precision measurement allowing us to test the predictions of the QCD theory and potentially further constrain its parameters. As discussed before since this particular process is a background to many extensions of the SM accurately calculating it is essential for testing these models. This measurement has been conducted before by both Atlas [22, 23, 24, 25, 26, 27, 28] and CMS [29, 30, 31, 32, 33, 34, 35, 36, 37, 38] at 7,8 and 13 TeV center of mass energy. The difference with this analysis is that the final state is boosted, meaning that the decay products have high transverse momentum (p_T). The conditions under which the collisions at the LHC performed allow for such a measurement since the energy is high enough to be able to produce particles in the boosted regime. The really high collision rate makes the LHC a "top factory" allowing for enough events to have kinematic characteristics in this far end of the spectrum so that this analysis is feasible.

The big challenge of this analysis is the reconstruction of the decay products of the top quark. As mentioned in chapter 2 the top quarks are not detected directly, but what we see is their decay products. In the previous analysis mentioned above, the study was conducted to a phase space up to 500 GeV in which the decay products can be reconstructed separately. At higher p_T, $(p_T/m \approx 1)$ the decay products are collimated and can not be reconstructed separately.The particles of

interest are rather reconstructed as jets of large radius. The fully hadronic final state is selected because it is rather easier the reconstruct the decay products in this channel than the other channels.

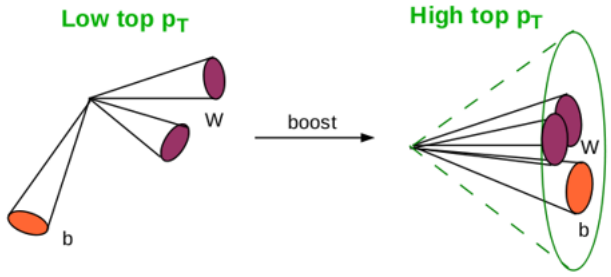

Figure 7.1: Schematic representation of the decay of a boosted (right) and a non-boosted (left) top quark. It can be seen that for a boosted quark the decay products are collimated and need to be reconstructed together. On the other hand for a non-boosted top quark the decay products are separate entities and can be detected separately.

This analysis is an extension of a previous analysis performed by the CMS experiment. The exact same measurement was performed using the data collected in the run period of 2016 with an integrated luminosity of 35.9 fb^{-1} at a center of mass energy of $\sqrt{13}$ TeV [38]. This analysis uses the full Run-II dataset that include data collected during 2017 and 2018 for an integrated luminosity of 41.5 fb^{-1} and 59.7 fb^{-1} respectively. The whole dataset used adds up to a total luminosity of 137.1fb^{-1}. This means that 4 times more data are used since the last measurement, allowing for greater precision. Additionally, the method used has been tuned in order to better optimise the background subtraction methods leading to even more accuracy in the measurement. Finally uncertainties have been taken into account in the comparison with the various theoretical models enabling a more realistic comparison between theory and data.

7.1 Signal and Backgrounds

As discussed again in chapter 2 top quarks decay in a W boson and a b quark. Since this analysis considers only the hadronic decay of the W boson, the events of interest we will have another two jets originating from the W decay. The final state will have at least 6 partons. Even more maybe present since additional particles can be radiated during the process. Figure 7.2 shows the possible ways of the $t\bar{t}$

pair with one decaying hadronically and one leptonically. In our case both t quarks should decay hadronically. The partons in the final state will be reconstructed as large radius jets. The signal definition for this analysis is the presence of two of these large radius jets.

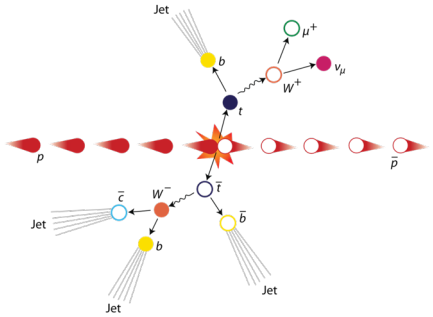

Figure 7.2: Schematic representation of the decay of the possible decays of the $t\bar{t}$. One top quark (bottom) decays hadronically, while the other one (top) decays leptonically

As background we identify processes that can have similar signature in the detector as our process of interest. These events can pass our selection criteria and be wrongly identified as a $t\bar{t}$ bar. In our analysis the by far main background process is the production of QCD multijets. This means that we can have events that via the strong interaction produce other highly energetic jets that pass our selection criteria and be reconstructed as boosted jets. These jets do not originate from a specific interaction but can be a result of any process that can radiate highly energetic gluons. This background process is modeled via a data-driven method. This means that we use information coming from the data to identify the total contribution of these types of events to our selection in order to remove it.

Other important contributions but of lower magnitude are:

- Single top quark: There is a minor contribution of events of this type but it is possible. Since we require that the events have two highly energetic jets most events of this type will fail the selection criteria. But it can happen that events of this type are selected. The contribution is small.

- W + jets: Events that include the production of a W boson accompanied with other jets can also be miss identified as signal.

- Z + jets: A Z boson is less likely to be produced that a W boson (lower cross section) though it is possible and in the presence of high p_T jets it can

happen, although rare, that it is selected.

All these processes that we call subdominant background in the context of this analysis are modeled with dedicated simulation (Monte Carlo (MC) samples.

7.2 Data and Simulation Samples

This section lists the data sets and Monte carlo simulation samples used in this analysis. During the 2016 data taking period it was noticed that a specific setting in the strips of the tracker was wrong. In particular the preamplier feedback voltage bias (VFP) had a wrong value which led to non optimal performance for the strips detector. This was realised and fixed but still the data were recorded. This lead to a splitting in the 2016 data taking period in to two sub-periods 2016_preVFP and 2016_postVFP. For all the samples mentioned later the MINIAOD data format is used [39].

7.2.1 Data

- 2016_preVFP: The total integrated luminosity of the analyzed data is $19.5 fb^{-1}$. `Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt` was used for luminosity and valid runs.
- 2016_postVFP: The total integrated luminosity of the analyzed data is $16.8 fb^{-1}$. `Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt` was used for luminosity and valid runs.
- 2017: The total integrated luminosity of the analyzed data is $41.5 fb^{-1}$. `Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt` was used for luminosity and valid runs.
- 2018: The total integrated luminosity of the analyzed data is $59.7 fb^{-1}$. `Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt` was used for luminosity and valid runs.

Table 7.1: Data samples used for the 2016 data taking period.

Sample	Run range	Lumi (pb^{-1})
/JetHT/Run2016B-21Feb2020_ver2_UL2016_HIPM-v1	273150-275376	5750
/JetHT/Run2016C-21Feb2020_UL2016_HIPM-v1	275656-276283	2573
/JetHT/Run2016D-21Feb2020_UL2016_HIPM-v1	276315-276811	4242
/JetHT/Run2016E-21Feb2020_UL2016_HIPM-v1	276947-277420	4025
/JetHT/Run2016F-21Feb2020_UL2016_HIPM-v1	277932-278807	3105
/JetHT/Run2016F-21Feb2020_UL2016-v1	278769-278808	7576
/JetHT/Run2016G-21Feb2020_UL2016-v1	278820-284035	8435
/JetHT/Run2016H-21Feb2020_UL2016-v1	281613-284044	216

Table 7.2: Data samples used for the 2017 data taking period.

Sample	Run range	Luminosity (pb^{-1})
/JetHT/Run2017B-31Mar2018-v1	297047-299329	4793
/JetHT/Run2017C-31Mar2018-v1	299368-302029	9755
/JetHT/Run2017D-31Mar2018-v1	302031-302663	4320
/JetHT/Run2017E-31Mar2018-v1	303824-304797	9422
/JetHT/Run2017F-31Mar2018-v1	305040-306460	13568

Table 7.3: Data samples used for the 2018 data taking period.

Sample	Run range	Luminosity (pb^{-1})
/JetHT/Run2018A-17Sep2018-v1	315257-316995	13530
/JetHT/Run2018B-17Sep2018-v1	317080-319310	6788
/JetHT/Run2018C-17Sep2018-v1	319337-320065	6612
/JetHT/Run2018D-PromptReco-v2	320497-325175	31947

7.2.2 Simulation

Monte Carlo (MC) simulation is used to simulate the $t\bar{t}$ pair production but also the various background processes. Regarding $t\bar{t}$ the POWHEG event generator was used to a next to leading order (NLO) precision with a top quark mass at 172.5 GeV. Single top quark production in the t channel or in the presence of a W boson is also simulated via POWHEG at NLO precision. The production W or Z bosons with jets and the QCD multijet production simulkation is done with the madgraph event generator MG5_AMC@NLO at leading order(LO) precision. For these samples the MLM algorithm is also used for matching.

For modeling of the parton shower, hadronization and the undelying event PYTHIA is used. In particular version 8.240. For PDF the NNPDF 3.1 is used. For all the simulation samples the CP5 tune is used. For the detector simulation the GEANT4 software package is used.

The measurement of the cross section is compared with three theoretical models. These are POWHEG combined with PYTHIA, MADGRAPH combined with PYTHIA and POWHEG combined with HERWIG. For the herwig sample the tune CH3 was used.

Regarding simulation, the following list contains the simulation era that was used for the simulation samples of each year.

- 2016 preVFP: `RunIISummer20UL16MiniAODAPV-106X_mcRun2_asymptotic_preVFP`
- 2016 postVFP: `RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic`
- 2017: `RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic`
- 2018: `RunIIAutumn18MiniAOD-102X_upgrade2018_realistic`

The tables below contain the list of simulation samples samples for each year.

Table 7.4: List of Monte Carlo samples used for the 2016 preVFP.

Sample	Events ($\times 10^6$)	σ (pb)
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	98.1	377.96
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	138.6	365.34
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	41.6	88.29
TTToHadronic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	40.0	377.96
TTToSemiLeptonic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	56.8	365.34
TTTo2L2Nu_mtop166p5_TuneCP5_13TeV-powheg-pythia8	16.7	88.29
TTToHadronic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	39.6	377.96
TTToSemiLeptonic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	57.7	365.34
TTTo2L2Nu_mtop169p5_TuneCP5_13TeV-powheg-pythia8	16.4	88.29
TTToHadronic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	38.9	377.96
TTToSemiLeptonic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	57.9	365.34
TTTo2L2Nu_mtop171p5_TuneCP5_13TeV-powheg-pythia8	16.9	88.29
TTToHadronic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	39.5	377.96
TTToSemiLeptonic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	57.5	365.34
TTTo2L2Nu_mtop173p5_TuneCP5_13TeV-powheg-pythia8	16.9	88.29
TTToHadronic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	39.8	377.96
TTToSemiLeptonic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	55.5	365.34
TTTo2L2Nu_mtop175p5_TuneCP5_13TeV-powheg-pythia8	16.9	88.29
TTToHadronic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	38.3	377.96
TTToSemiLeptonic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	57.9	365.34
TTTo2L2Nu_mtop178p5_TuneCP5_13TeV-powheg-pythia8	16.9	88.29
TTToHadronic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	40.0	377.96
TTToSemiLeptonic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	56.9	365.34
TTTo2L2Nu_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	17.0	88.29
TTToHadronic_hdampUP_TuneCP5_13TeV-powheg-pythia8	40.0	377.96
TTToSemiLeptonic_hdampUP_TuneCP5_13TeV-powheg-pythia8	57.0	365.34
TTTo2L2Nu_hdampUP_TuneCP5_13TeV-powheg-pythia8	17.0	88.29
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	99.3	832
TT_TuneCH3_13TeV-powheg-herwig7	71.5	832
QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	52.6	315400
QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	58.5	32260
QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	45.5	6830
QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	14.1	1207
QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	10.3	119.1
QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	5.1	25.16
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	3.4	38.09
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	3.3	38.09
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	31.0	35.6
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	56.0	35.6
ST_t-channel_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	31.0	119.7
ST_t-channel_top_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	57.1	82.52
WJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	8	2549.0
WJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.14	276.5
WJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	7.62	59.25
WJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8	7.71	28.75
TT_TuneCH3_13TeV-powheg-herwig7	72	832

Table 7.5: List of Monte Carlo samples used for the 2016 postVFP.

Sample	Events ($\times 10^6$)	σ (pb)
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	112.6	377.96
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	158.8	365.34
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	48.3	88.29
TTToHadronic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	43.6	377.96
TTToSemiLeptonic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	64.0	365.34
TTTo2L2Nu_mtop166p5_TuneCP5_13TeV-powheg-pythia8	18.7	88.29
TTToHadronic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	44.8	377.96
TTToSemiLeptonic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	55.1	365.34
TTTo2L2Nu_mtop169p5_TuneCP5_13TeV-powheg-pythia8	18.3	88.29
TTToHadronic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	45.0	377.96
TTToSemiLeptonic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	63.9	365.34
TTTo2L2Nu_mtop171p5_TuneCP5_13TeV-powheg-pythia8	18.9	88.29
TTToHadronic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	44.8	377.96
TTToSemiLeptonic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	64.0	365.34
TTTo2L2Nu_mtop173p5_TuneCP5_13TeV-powheg-pythia8	18.8	88.29
TTToHadronic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	43.5	377.96
TTToSemiLeptonic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	64.0	365.34
TTTo2L2Nu_mtop175p5_TuneCP5_13TeV-powheg-pythia8	19.0	88.29
TTToHadronic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	44.6	377.96
TTToSemiLeptonic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	64.0	365.34
TTTo2L2Nu_mtop178p5_TuneCP5_13TeV-powheg-pythia8	19.0	88.29
TTToHadronic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	45.0	377.96
TTToSemiLeptonic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	63.7	365.34
TTTo2L2Nu_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	18.3	88.29
TTToHadronic_hdampUP_TuneCP5_13TeV-powheg-pythia8	45.0	377.96
TTToSemiLeptonic_hdampUP_TuneCP5_13TeV-powheg-pythia8	63.0	365.34
TTTo2L2Nu_hdampUP_TuneCP5_13TeV-powheg-pythia8	19.0	88.29
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	110.1	832
TT_TuneCH3_13TeV-powheg-herwig7	176.5	832
QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	57.6	315400
QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	59.7	32260
QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	47.4	6830
QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	15.3	1207
QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	10.4	119.1
QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	5.3	25.16
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	10.3	38.09
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	3.8	38.09
ST_t-channel_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	4.0	35.6
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	22.8	35.6
ST_t-channel_top_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	6.0	119.7
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	46.5	82.52
WJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	7.07	2549.0
WJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	4.49	276.5
WJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	6.76	59.25
WJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8	6.85	28.75

Table 7.6: List of Monte Carlo samples used for the 2017.

Sample	Events ($\times 10^6$)	σ (pb)
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	249.6	377.96
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	355.9	365.34
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	107.0	88.29
TTToHadronic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	97.8	377.96
TTToSemiLeptonic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	135.7	365.34
TTTo2L2Nu_mtop166p5_TuneCP5_13TeV-powheg-pythia8	42.5	88.29
TTToHadronic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	98.3	377.96
TTToSemiLeptonic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	138.6	365.34
TTTo2L2Nu_mtop169p5_TuneCP5_13TeV-powheg-pythia8	41.0	88.29
TTToHadronic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	97.9	377.96
TTToSemiLeptonic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	141.0	365.34
TTTo2L2Nu_mtop171p5_TuneCP5_13TeV-powheg-pythia8	43.0	88.29
TTToHadronic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	99.9	377.96
TTToSemiLeptonic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	141.7	365.34
TTTo2L2Nu_mtop173p5_TuneCP5_13TeV-powheg-pythia8	42.7	88.29
TTToHadronic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	100.0	377.96
TTToSemiLeptonic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	141.6	365.34
TTTo2L2Nu_mtop175p5_TuneCP5_13TeV-powheg-pythia8	42.9	88.29
TTToHadronic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	99.6	377.96
TTToSemiLeptonic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	141.3	365.34
TTTo2L2Nu_mtop178p5_TuneCP5_13TeV-powheg-pythia8	42.7	88.29
TTToHadronic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	99.6	377.96
TTToSemiLeptonic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	141.6	365.34
TTTo2L2Nu_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	41.1	88.29
TTToHadronic_hdampUP_TuneCP5_13TeV-powheg-pythia8	99.7	377.96
TTToSemiLeptonic_hdampUP_TuneCP5_13TeV-powheg-pythia8	140.9	365.34
TTTo2L2Nu_hdampUP_TuneCP5_13TeV-powheg-pythia8	42.2	88.29
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	249.1	832
QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	57.2	315400
QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	57.9	32260
QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	45.8	6830
QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	15.3	1207
QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	10.6	119.1
QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	5.5	25.16
ST_t-channel_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	3.7	35.6
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	24.3	35.6
ST_t-channel_top_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	5.5	119.7
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	46.2	82.52
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	10.0	38.09
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	9.2	38.09
WJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	15.42	2549.0
WJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	9.91	276.5
WJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	14.4	59.25
WJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8	14.75	28.75

Table 7.7: List of Monte Carlo samples used for the 2018.

Sample	Events ($\times 10^6$)	σ (pb)
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	347.4	377.96
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	495.0	365.34
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	148.5	88.29
TTToSemiLeptonic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	191.1	365.34
TTToHadronic_mtop166p5_TuneCP5_13TeV-powheg-pythia8	137.8	377.96
TTTo2L2Nu_mtop166p5_TuneCP5_13TeV-powheg-pythia8	60.0	88.29
TTTo2L2Nu_mtop169p5_TuneCP5_13TeV-powheg-pythia8	59.8	88.29
TTToHadronic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	138.7	377.96
TTToSemiLeptonic_mtop169p5_TuneCP5_13TeV-powheg-pythia8	194.7	365.34
TTTo2L2Nu_mtop171p5_TuneCP5_13TeV-powheg-pythia8	59.8	88.29
TTToHadronic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	135.3	377.96
TTToSemiLeptonic_mtop171p5_TuneCP5_13TeV-powheg-pythia8	195.2	365.34
TTTo2L2Nu_mtop173p5_TuneCP5_13TeV-powheg-pythia8	59.9	88.29
TTToHadronic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	139.6	377.96
TTToSemiLeptonic_mtop173p5_TuneCP5_13TeV-powheg-pythia8	200.0	365.34
TTTo2L2Nu_mtop175p5_TuneCP5_13TeV-powheg-pythia8	60.0	88.29
TTToHadronic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	138.4	377.96
TTToSemiLeptonic_mtop175p5_TuneCP5_13TeV-powheg-pythia8	199.8	365.34
TTTo2L2Nu_mtop178p5_TuneCP5_13TeV-powheg-pythia8	59.9	88.29
TTToHadronic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	139.7	377.96
TTToSemiLeptonic_mtop178p5_TuneCP5_13TeV-powheg-pythia8	199.9	365.34
TTTo2L2Nu_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	60.0	88.29
TTToHadronic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	139.7	377.96
TTToSemiLeptonic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	193.2	365.34
TTTo2L2Nu_hdampUP_TuneCP5_13TeV-powheg-pythia8	59.8	88.29
TTToHadronic_hdampUP_TuneCP5_13TeV-powheg-pythia8	138.3	377.96
TTToSemiLeptonic_hdampUP_TuneCP5_13TeV-powheg-pythia8	199.4	365.34
TTTo2L2Nu_TuneCP5down_13TeV-powheg-pythia8	60.0	88.29
TTToHadronic_TuneCP5down_13TeV-powheg-pythia8	139.8	377.96
TTToSemiLeptonic_TuneCP5down_13TeV-powheg-pythia8	190.3	365.34
TTTo2L2Nu_TuneCP5up_13TeV-powheg-pythia8	57.9	88.29
TTToHadronic_TuneCP5up_13TeV-powheg-pythia8	139.9	377.96
TTToSemiLeptonic_TuneCP5up_13TeV-powheg-pythia8	199.7	365.34
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	340.5	832
TT_TuneCH3_13TeV-powheg-herwig7	241	832
QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	55.1	315400
QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	58.5	32260
QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	47.7	6830
QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	15.7	1207
QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	10.6	119.1
QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	4.5	25.16
ST_t-channel_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	3.7	35.6
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	23.4	35.6
ST_t-channel_top_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	5.7	119.7
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	47.7	82.52
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	10.1	38.09
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	9.1	38.09
WJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	14.39	2549.0
WJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	9.26	276.5
WJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	13.54	59.25
WJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8	13.6	28.75

7.3 Variables of interest

As the measurement performed in this analysis is differential, it is done with respect to a certain variable. Before mentioning the variables, some introduction needs to be made. In the CMS experiment the reconstructed objects are ordered based on their transverse momentum (p_T). When referring to an object as "leading" we refer to the object with the highest transverse momentum (p_T). This is true for all variables not only for p_T itself. For example, the mass of the leading jet, means the mass of the jet with the highest transverse momentum out of the reconstructed jets.

The first set of variables used are the p_T of the leading and sub-leading jets. The transverse momentum is used as a variable because it is Lorentz invariant. It is easier to study the kinematics of a collision in the center of mass (CM) frame (coordinate system). The CM is the frame at which the total momentum of the system equals to zero. Having that constraint, calculations become easier. On the other hand, the measurements we make with our experimental apparatus are in what is called the laboratory frame. In particle physics this is the frame of reference whose center is at the center of the lab. In order to transfer from one frame to the other we have to transform our measurements with what is called a Lorentz boost. When we say that a variable is Lorentz invariant it means that it stays the same under a Lorentz boost. As it can be understood using variables that have this characteristic make things easier.

Another set of variables used are the rapidity and pseudorapidity η of the leading and sub-leading jets. Consider the 4-momentum of a particle, $(E/c, p_x, p_y, p_z)$. We define the quantity:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z c}{E - p_z c}\right) \tag{7.1}$$

The coordinate system used in collider experiments has the z-axis in the beam direction. Now consider a particle that is the product of a collision, moving in the XY plane, i.e. perpendicular to the beam direction. In this case the p_z will be small and the rapidity will be almost 0. When the particle moves along the z-axis in the positive direction, $E \approx p_z c$ and then $y \to \infty$. In the negative it will be $E \approx -p_z c$ and $y \to -\infty$. Rapidity is also very useful because rapidity differences are Lorenz invariant under boosts in the z-axis.

Although rapidity is a useful property it has the disadvantage that it is hard to measure for high relativistic particles. To calculate rapidity we need to know both the energy and the momentum of a particle. This is not an easy thing to do especially for high values of p_z. However we can define a similar property that is

independent of this things, pseudorapidity. Starting from the definition of rapidity Equation 7.1 using the energy definition we can do

$$\begin{aligned} y &= \frac{1}{2}\ln\left(\frac{E+p_zc}{E-p_zc}\right) \\ &= \frac{1}{2}\ln\frac{\sqrt{p^2c^2+m^2c^4}+p_zc}{\sqrt{p^2c^2+m^2c^4}-p_zc} \end{aligned}$$

Taking into account that pc is way bigger that mc^2 for highly relativistic particles we can use a binomial expansion. Also using that $p_z/p = cos\theta$ where θ is the angle between the particles trajectory and the beam pipe we get

$$\begin{aligned} y &\simeq \frac{1}{2}\frac{\cos^2\frac{\theta}{2}}{\sin^2\frac{\theta}{2}} \\ y &\simeq -\ln\tan\frac{\theta}{2} \end{aligned}$$

which is the definition of pseudorapidity η.

$$\eta = -\ln\tan\frac{\theta}{2} \tag{7.2}$$

So for highly relativistic particles, rapidity and pseudorapidity are approximately equal.

While the quantities already discussed are properties of each individual object, the rest of the variables of interest relate to the total system of the interaction products. The other 3 variables are the m_{JJ} which is the sum of the masses of the two leading jets, the p_{T_JJ} which the sum of the transverse momentum of the two leading jets and finally y_{JJ} which is the sum of the rapidities of the two leading jets.

7.4 Trigger

As mentioned before the detector records only events that are selected by the trigger system of the experiment. The trigger is a modular system that contains many sets of selection criteria. The same architecture are used both in L1 trigger as well as in the HLT. In the context of the L1 trigger these sets are called seeds, while the HLT ones are called paths. Since each event must pass both trigger stages, trigger paths are a combination of an L1 path that is used as a seed (L1

seed) for the HLT trigger. One of the first requirements of an analysis is to select one or more triggers that select events that are of interest for this particular study. Since the detector configuration was different throughout the Run-II data period, different triggers were used.

Table 7.8: Summary of triggers used in the analysis for the 2016 data taking period.

Trigger	Purpose
L1_SingleJet180 OR L1_SingleJet200	L1 seed
HLT_AK8DiPFJet280_200_TrimMass30_BTagCSV_p20	signal HLT path
HLT_AK8PFJet140	control HLT path

The triggers used for 2016 data can be found in Table 7.8. The trigger path at the L1 level requires that the event has at least a jet with $p_T > 180$ GeV. This path is used as a seed for the HLT path. At the high level trigger, the jets are reconstructed from particle flow (PF) candidates using the anti-kt algorithm. They are characterized from the radius of the cone that is used for the reconstruction. For this particular trigger path the radius used is 0.8. From now on these type of jets, i.e. reconstructed from PF (candidates) using the ak algorithm with a radius of 0.8 will be referred to as AK8. They are required to have mass greater than 30 GeV after the trimming of soft particles. The event is must have at least two jets where one of them is b-tagged. The CSV algorithm is used for b-tagging. Finally the jets must have a p_T greater than 280 GeV for the leading and 200 GeV for the sub-leading jet. This trigger run for the whole 2016 data period collecting an integrated luminosity of 35.9 fb^{-1}. The trigger was used unprescaled.

Due to the method used for the calculation of the background, there is a need for a trigger that has no b-tagging requirement. Another trigger was used for that reason. This trigger has the same L1 seed and at the HLT leve requires the presence of a jet with a prt more than 140 GeV. This trigger run again for the full 2016 data taking period but prescaled. It collected an integrated luminosity of 1.67 fb^{-1}.

The detector configuration was different in 2016 compared to the following years of data taking. For that reason a different trigger was used for 2017 and 2018. This trigger path can be seen in Table 7.9. A more complex L1 seed is used. It requires that the total hadronic transverse energy (HTT) is at least 120 GeV. Alternatively the presence of 4 jets in the event with an HTT of at least 280 GeV is required. Out of these jest the leading jet must have an energy of 70 GeV, the subleading 55 GeV and 40 and 34 GeV for the next two jets. At the HLT level again

Table 7.9: Summary of triggers used in the analysis for the 2017 and 2018 data taking period.

Trigger	Purpose
L1_HTT120er OR L1_HTT160er OR L1_HTT200er OR L1_HTT220er OR L1_HTT240er OR L1_HTT255er OR L1_HTT270er OR L1_HTT280er OR L1_HTT280er_QuadJet_70_55_40_35_er2p5 OR L1_HTT300er OR L1_HTT300er_QuadJet_70_55_40_35_er2p5 OR L1_HTT320er OR L1_HTT340er OR L1_HTT380er OR L1_HTT400er OR L1_HTT450er OR L1_HTT500er	L1 seed
HLT_AK8PFHT800_TrimMass50	signal & control HLT path

AK8 jets are used. It is required that the total transverse energy of the event (H_T) is 800 GeV and that the jets have a mass of 50 GeV after trimming. This trigger path run for both 2017 and 2018 unprescaled, gathering an integrated luminosity of 41.5 fb^{-1} and 59.7 fb^{-1} respectively.

A measure of how good a trigger performs is a quantity called trigger efficiency. The trigger efficiency is basically the probability that a trigger selects an event of a specific type. In our case we define how "efficient" the trigger is selecting event that has the characteristics that we want for our analysis. That means how many of the events that should be selected are indeed selected. In order to determine all the events we use what is called a reference trigger is used. This trigger is one that is "orthogonal", i.e. not correlated to the trigger that is of interest. The trigger efficiency is defined as:

$$\epsilon = \frac{N_{ref} + N_{sig} + selection\ criteria}{N_{ref} + selection\ criteria} \tag{7.3}$$

where N_{ref} are the events passing the reference trigger, N_{sig} are the events passing the signal trigger and $selection\ criteria$ refers to any additional criteria that are used in the selection of the events. The reference trigger used in our case is the HLT_IsoMu27 which requires the presence of an isolated muon with p_T greater

than 27 GeV.

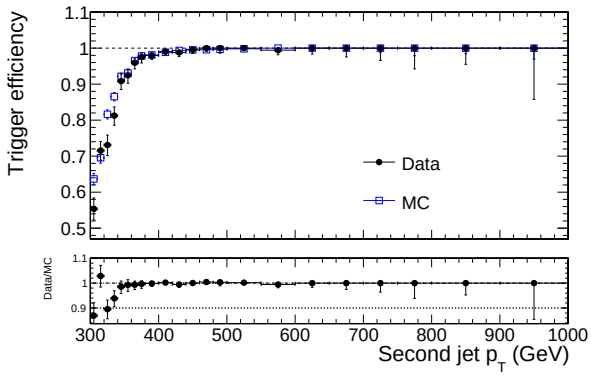

Figure 7.3: Trigger efficiency for the signal path of the analysis vs second leading p_T for 2016

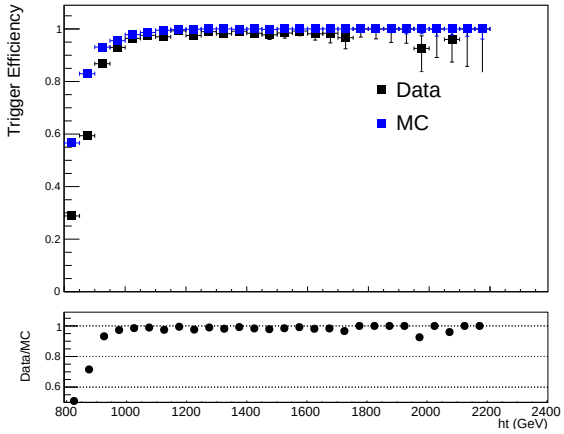

Figure 7.4: Trigger efficiency for the signal path of the analysis vs the total transverse momentum of the event for 2017

The trigger efficiency can be seen in Figure 7.3, Figure 7.4, Figure 7.5 for each year. It is displayed in comparison with the expected efficiency calculated from simulations. For the 2016 trigger the p_T of the sub-leading jet is used while for the 2017 and 2018 trigger the variable H_T which is the p_T sum of all the jets in the event is used.

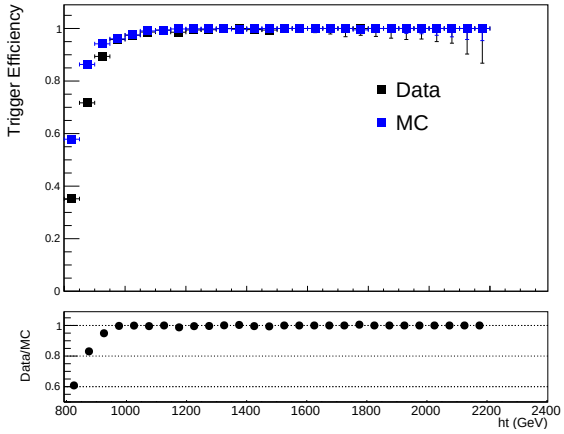

Figure 7.5: Trigger efficiency for the signal path of the analysis vs the total transverse momentum of the event for 2018

7.5 Selection criteria

On top of the selection criteria that are imposed by the trigger selection we need to apply another set of criteria in order to select events that are of interest for this analysis. The selection criteria aim to select events that contain $t\bar{t}$ pairs of high transverse momentum that decay hadronically. Of course this can not be done totally with just the selection criteria and background has to be further removed with the method discussed later, see section 7.6.

The leptons used in this analysis are muons and electrons. The are used to ensure better jet reconstruction. First we require that there are no leptons in the event. To achieve so, leptons need to be reconstructed with a good precision allowing us to discard events that contain them. Secondly leptons can be misinterpreted as jets cause jets also carry an electromagnetic component. This way, indirectly, good lepton identification ensures that also jets are correctly identified. In this analysis muons are reconstructed from PF candidates and are required to have $p_T > 20$ GeV. An additional set of criteria are used based on the isolation and identification algorithms [40]. For muons the medium working point for the identification (ID) algorithm is used while the isolation is required to be less than 0.1. For electrons, the tight ID working point is used and again 0.1 as the required value for isolation.

For jets, PF candidates are used [13]. The Pile Up Per Particle Identification algorithm (PUPPI) is used to remove pile up contributions from the reconstructed

jet. Finally they are reconstructed as AK8 jets, see section 7.4. These jets will be referred to as AK8PFPUPPI jets. They are required to pass the tight cut of the jet identification algorithm used in CMS [41]. In addition to the AK8 jets another set of jets is used. The large-radius AK8 jets contain all the decay products of the t ($\bar{t}$) quark. But we are also interested in the substructure of this jet. For example the b quark inside it will be another smaller jet, or the decay products of the W boson will also be other smaller radius jets. The same algorithms are used to construct a collection of smaller jets, AK4 jets. For the mass of the jets the soft drop mass algorithm was used chapter 4.

The jets are required to contain at least one b-quark. This allows to identify whether a jet is the result of a top quark decay or not. The DeepCSV algorithm [15] is used in order to identify the jets as b-tagged or not. It calculates the probability that a subject of the AK8 jets (AK4) is a b quark. The medium working point is used for each year, according to the CMS recommendations [42], [43], [44], [45]. Finally since leptons can be also reconstructed as jets they are removed by doing a geometrical matching. If the distance between a jet and a lepton $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is less than 0.4 the jet is rejected from the selection.

7.5.1 Baseline Selection

After doing some basic pre-selection on the reconstructed objects from the detector, further criteria are applied. In Table 7.10 the baseline selection of the analysis is shown. This is the basis and is common to all the areas defined later. It requires at least two jets in the events and a veto on leptons, i.e. no leptons in the event. Kinematically, the jets are required to have a transverse momentum of 450 and 400 GeV for the leading and sub-leading jet respectively. They are both required to have an absolute pseudorapidity of less than 2.4. Their soft drop mass is required to be in the range of 50 to 300 GeV. Finally the mass m_{jj} of the system is required to be more than 1000 GeV.

7.5.2 Boosted Decision Tree

In addition to the selection criteria used in the analysis to select events, a Boosted Decision Tree (BDT) is developed and used. A bdt is a machine learning (ML) algorithm used in classification problems. A decision tree Figure 7.6 is a tree structure, starting from a root note. Additional child nodes are defined by applying binary criteria to the starting node until a some conditions are met. Boosting is a technique where multiple weaker classifiers are combined together to form a

Table 7.10: Baseline Selection Criteria

Observable	Requirement
N_{jets}	> 1
N_{leptons}	$= 0$
p_T^{jet1}	$> 450\text{GeV}$
p_T^{jet2}	$> 400\text{GeV}$
$\lvert\eta^{\text{jet1,2}}\rvert$	< 2.4
$m_{SD}^{\text{jet1,2}}$	$(50, 300)\text{GeV}$
$m_{t\bar{t}}$	$> 1000\text{GeV}$

stronger one. In every step weights are applied to each tree in order to account for its accuracy. BDTs are widely used in high energy physics in classification of structures as signal or background. The same is done in our analysis where we use a BDT to classify events.

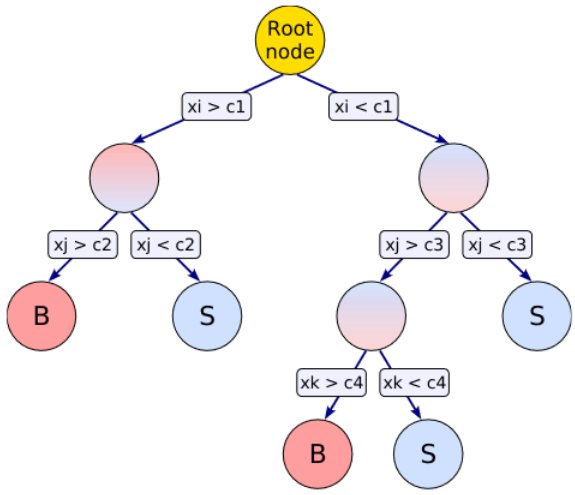

Figure 7.6: Schematic representation of a decision tree [46]. A decision tree is a tree structure, starting from a root note. Additional child nodes are defined by applying binary criteria to the starting node until a some conditions are met.

The BDT developed in this analysis is used to classify jets as top candidates or not. The samples are splitted in 4 categories, based on the p_T of the jet. Splitting the phase space in categories, enhances the performance of the discriminator. The

categories are (400, 600) GeV, (600, 800) GeV, (800, 1200) GeV, (1200, Inf) GeV. In total 500 trees are trained with the Gradient Boost method. The BDT relies on various variables to discriminate the various objects. First variables regarding the structure of the jet are used. The n-subjetiness [47] variables take advantage of the fact that partons that result from the decay of a particle carry a fraction of the particle's momentum. Therefore the momentum inside a jet is not evenly distributed but rather is located in certain parts of the jet ("prongs"), see Figure 7.7. They are a set of variables, calculated by

$$\tau_N = \frac{1}{\sum_k p_{T,k} R} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots \Delta R_{N,k}\}, \tag{7.4}$$

where N is the number of candidate reconstructed subjets, e.g. τ_3 calculates the value for having 3 subjets. k is an index that runs over the constituent particles in the jet. A value close to zero means that most of the constituents are aligned along the subjet axis. If $\tau_N \gg 0$ means that the jet has probably more than N subjets. The BDT developed for this analysis uses $\tau_1,\ \tau_2,\ \tau_3$.

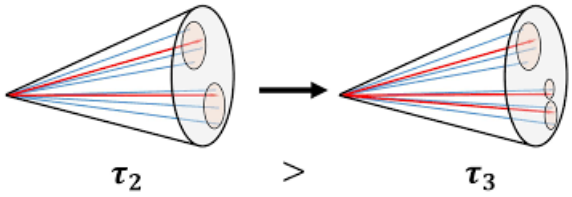

Figure 7.7: Partons that result from the decay of a particle carry a fraction of the particle's momentum. Therefore the momentum inside a jet like that is not evenly distributed but rather is located in certain parts of the jet ("prongs"). The figure displays a jet with two and a jet with three "prongs".

Another set of variables used are the Energy Correlation Functions (ECF) [48]. They are based on the energies and pair-wise angles of particles within a jet .They are also sensitive to N-prong substructure. For hadron colliders they are defined as:

$$ECF(N,\beta) = \sum_{i_1<i_2\cdots<i_N\in J} (\prod_{i=1}^{N} p_{T,i})(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} R_{i_b i_c})^{\beta} \tag{7.5}$$

where R_{ij} is the Euclidean distance between i and j in the rapidity-azimuth angle plane and $R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, with $y_i = \frac{1}{2}\ln\frac{E_i+p_{z_i}}{E_i-p_{z_i}}$. The BDT makes use of the following ECFs ECFB1N2, ECFB1N3, ECFB2N2, ECFB2N3.

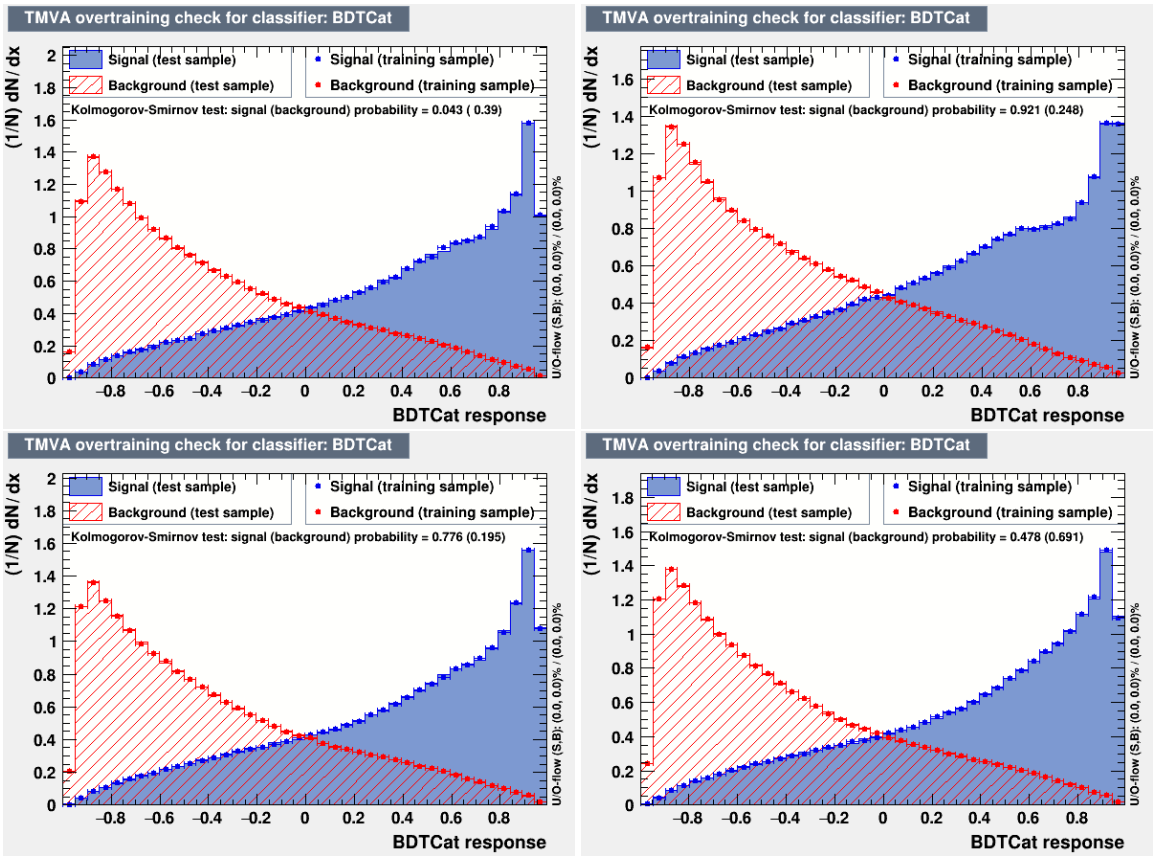

Figure 7.8: Output of the Boosted Decision Tree (BDT). A comparison in the performance of the BDT in classifying actual top quark jets and QCD jets as top quarks. The output is displayed for each year. Top left is 2016_preVFP, top right 2016_postVFP, bottom left 2017 and bottom right 2018

Together with the variables mentioned above the soft drop mass mass of the jet and the ratio of the jet p_t over the sum of the p_T's of all the jets in the event are used. In total it requires 10 variables to classify a jet as a top quark or not. Since each data taking year is different a different BDT was developed for each year. Accordingly a separate working point is used for each year. Figure 7.8 shows the response of the BDT.

7.5.3 Analysis regions

Putting it all together, the following regions are defined in this analysis. The signal region (SR) is where the actual measurement is performed. It contains events that pass the baseline selection (Table 7.10). It requires that both jets have a b-tagged subjet and pass the BDT top tagging selection. Finally they need to pass the mass cut criterion where their soft drop mass needs to be in the region (120, 220) GeV. The next defined region is the extended signal region (SR_A). It

Table 7.11: Selection requirements per analysis region for 2016 pre and post VFP

Region	Trigger	Offline	Purpose
SR	Signal	Baseline + $BDT > 0.2$ + 2 b-tag subjets + $m_{SD}^{jet1,2} \in (120,\ 220) GeV$	Signal Region
SR_A	Signal	Baseline + $BDT > 0.2$ + 2 b-tag subjets + $m_{SD}^{jet1,2} \in (50,\ 300) GeV$	QCD fit region
CR	Control	Baseline + $BDT > 0.2$ + 0 b-tag subjets + $m_{SD}^{jet1,2} \in (120, 220) GeV$	QCD control region
CR_A	Control	Baseline + $BDT > 0.2$ + 0 b-tag subjets + $m_{SD}^{jet1,2} \in (50,\ 300) GeV$	QCD extended CR

is similar to the SR but has a loser selection threshold for the soft drop mass cut, i.e. (50, 300) GeV. This region is used for the fit procedure which is used to define the QCD background, see section 7.6. The next region is the Control Region (CR). It contains events that pass the baseline selection and the BDT top tagging selection. They are also required to have a soft drop mass between 120 and 200 GeV. The jets in this region are required to not b-tagged. This region is used to define the shape of the QCD contribution and then subtract it. The final region is the extended control region (CR_A). It is similar to the CR but has the looser cut for the soft drop mass of the jets, (50, 300) GeV. The different setup of the detector in the various data taking years, yield different threshold values (working points) for the b-tagging algorithm and the BDT top tagging algorithm. For the b-tag the values are 0.6001 , 0.5847, 0.4168, 0.4506 for 2016_preVFP, 2016_postVFP, 2017 and 2018 respectively. For top tagging, the working points are 0.2 for both 2016 periods and 0.0, 0.1 for 2017 and 2018 respectively.

Table 7.12: Selection requirements per analysis region for 2017 (Control and Signal triggers are the same)

Region	Trigger	Offline	Purpose
SR	Signal	Baseline + $BDT > 0.0$ + 2 b-tag subjets + $m_{SD}^{jet1,2} \in (120, 220)$ GeV	Signal Region
SR_A	Signal	Baseline + $BDT > 0.0$ + 2 b-tag subjets + $m_{SD}^{jet1,2} \in (50,\ 300) GeV$	QCD fit region
CR	Signal	Baseline + $BDT > 0.0$ + 0 b-tag subjets + $m_{SD}^{jet1,2} \in (120, 220)$ GeV	QCD control region
CR_A	Control	Baseline+$BDT > 0.0$ + 0 b-tag subjets + $m_{SD}^{jet1,2} \in (50,\ 300) GeV$	QCD extended CR

Table 7.13: Selection requirements per analysis region for 2018 (Control and Signal triggers are the same)

Region	Trigger	Offline	Purpose
SR	Signal	Baseline + $BDT > 0.1$ + 2 b-tag subjets + $m_{SD}^{jet1,2} \in (120, 220) GeV$	Signal Region
SR_A	Signal	Baseline + $BDT > 0.1$ + 2 b-tag subjets + $m_{SD}^{jet1,2} \in (50,\ 300) GeV$	QCD fit region
CR	Signal	Baseline + $BDT > 0.1$ + 0 b-tag subjets + $m_{SD}^{jet1,2} \in (120, 220) GeV$	QCD control region
CR_A	Control	Baseline + $BDT > 0.1$ + 0 b-tag subjets + $m_{SD}^{jet1,2} \in (50,\ 300) GeV$	QCD extended CR

7.5.4 Measurement Areas

As said before, the actual particles of interest can not be detected directly since they decay very quickly. What is actually detected is their decay products. But even then sometimes the decay products also decay themselves before reaching

the detector volume. During the procedure of the experiment until the particle detection we can define various regions that represent the steps that take place, see Figure 7.9.

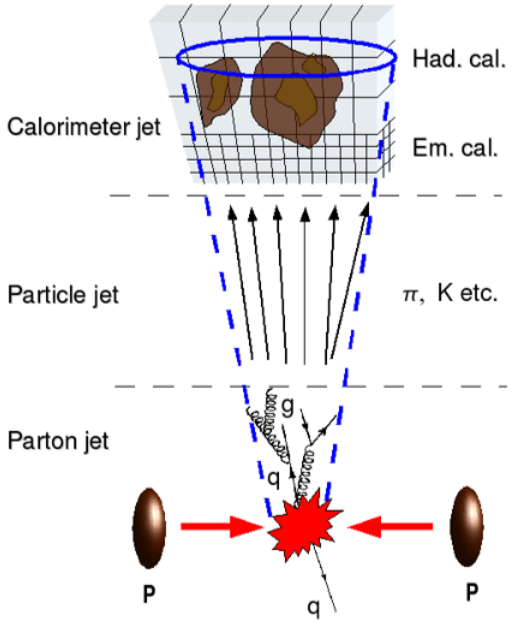

Figure 7.9: Schematic representation of the various phases of the particle lifetime as they are defined from an analysis point of view. Parton level contains the particles that we want to measure. Particle level contains their decay products, what we actually detect. And detector level is what is actually measured by the detector.

The first area we define is the parton level. This contains the particles we actually want to measure, in our case the $t\bar{t}$. The next area is the particle level. After decay, hadronization and other procedures the products that will be detected from the detector are produced. The level that contains the collection of the particles that will be detected is called the particle level. Finally we have the detector level which contains what is being measured by the detector. This is different from what is actually produced contained in the parton level as well as in the particle level. The differences are due to the detector resolution as well as other factors having to do with the modeling of the various processes happening in the decay etc.

In this analysis results are produced both in the parton and particle levels. Although we want to study the $t\bar{t}$ pair which is the parton level the study of the particle level gives us a measurement that is closer to the detector level and thus

can be more accurate as it suffers from less uncertainties. This is limited more by the detector resolution and not so much by the theoretical predictions related to the actual physics process. The parton level measurement involves much more the theoretical predictions as in order to go from the particle level to the parton level we need to have a good way of modeling the transition between these two levels and invert it.

Table 7.14: Definition of parton-level phase space.

Observable	Requirement
Leading Parton $p_T^{t,\bar{t}}$	$> 450\ GeV$
Second leading Parton $p_T^{t,\bar{t}}$	$>\ 400\ GeV$
$\lvert\eta^{t,\bar{t}}\rvert$	$<\ 2.4$
$m_{t\bar{t}}$	$>\ 1000\ GeV$

Table 7.15: Definition of particle-level phase space.

Observable	Requirement
N_{jets}	$>\ 1$
Leading particle jet p_T^{jet1}	$>\ 450\ GeV$
Second Leading particle jetp_T^{jet2}	$>\ 400\ GeV$
$\lvert\eta^{\text{jet1,2}}\rvert$	$<\ 2.4$
$m_{SD}^{\text{jet1,2}}$	$(120,\ 220)\ GeV$
m_{jj}	$>\ 1000\ GeV$

Similar to the definition of the measurement phase space with the cuts mentioned in the previous section we also need to define parton and particle phase spaces with a set of selection criteria. For the parton level, see Table 7.14 we required a p_T cut at 450 GeV and 400 Gev for the leading end sub-leading jets respectively and then a cut for the $|\eta|$ of both jets are 2.4. Finally there is a cut for the invariant mass of the $t\bar{t}$ system at 1000 GeV. For the particle level we require two or more AK8 jets. Only jets originating from the primary vertex are selected. Jets that are geometrically matched within ΔR < 0.4 with a lepton are rejected. The selected jets are then required to have p_T cuts at 450 and 400 GeV same as the detector level. Again $|\eta|$ is required to be less that 0.4 for both, their soft drop mass to be in the range (120, 220) GeV and the sum of their invariant masses to be more that 1000 GeV.

7.6 Signal Extraction

As discussed above, the dominant background contribution on this analysis comes from QCD multijet production. It is possible that jets that originate from the radiation of a parton mimic the jet produced from a top quark. Using the selection criteria described above and in particular b-tagging and the top tagging BDT greatly reduces the contribution of QCD in our signal selection. In order to remove the rest QCD contribution a data driven technique is used. This means that the data sample is used in order to estimate this contribution. The SR selection consists of three parts. The actual $t\bar{t}$ part which we want to keep, the QCD part and the other subdominant contributions. This can be represented by:

$$D(x) = a * S(x) + b * Q(x) + c * B(x) \tag{7.6}$$

where x refers to the measurement variables, $p_T^{t_1,t_2}$, $|y^{t_1,t_2}|$, $m^{t\bar{t}}$, $p_T^{t\bar{t}}$, $y^{t\bar{t}}$. D(x) is the measured distribution in data, S(x) is the signal, the $t\bar{t}$ contribution and Q(x) is the distribution of the QCD contribution. Finally B(x) is the distribution of the subdominant background contributions. $a,\ b,\ c$ are normalization factors for the yield of each contribution respectively.

The method used for the background contribution is based on the assumption that reverting the b-tagging requirement can gives us the same distribution as the QCD contribution. This means that the area we get has the same "shape" and that it is pure meaning that it contains only QCD jets and not other contributions. The following images serve as a closure test Figure 7.10 - Figure 7.15. They show a comparison of the shape of QCD multijets using the SR and CR selection criteria as described in the previous section. The shapes compared are for the SR selection used in the analysis and the CR which is the same but with the b-tagging requirement reverted. The plots show a fairly good agreement which means that the CR can be used to take the shape of the QCD contribution in the SR. In some of the variables like the mass soft drop of the leading jet and the p_T of the jets where there seems to be a bit bigger disagreement in the shapes. To account for this, a correction factor is used. It is extracted by a fit applied in the ratio plot of the two shapes.

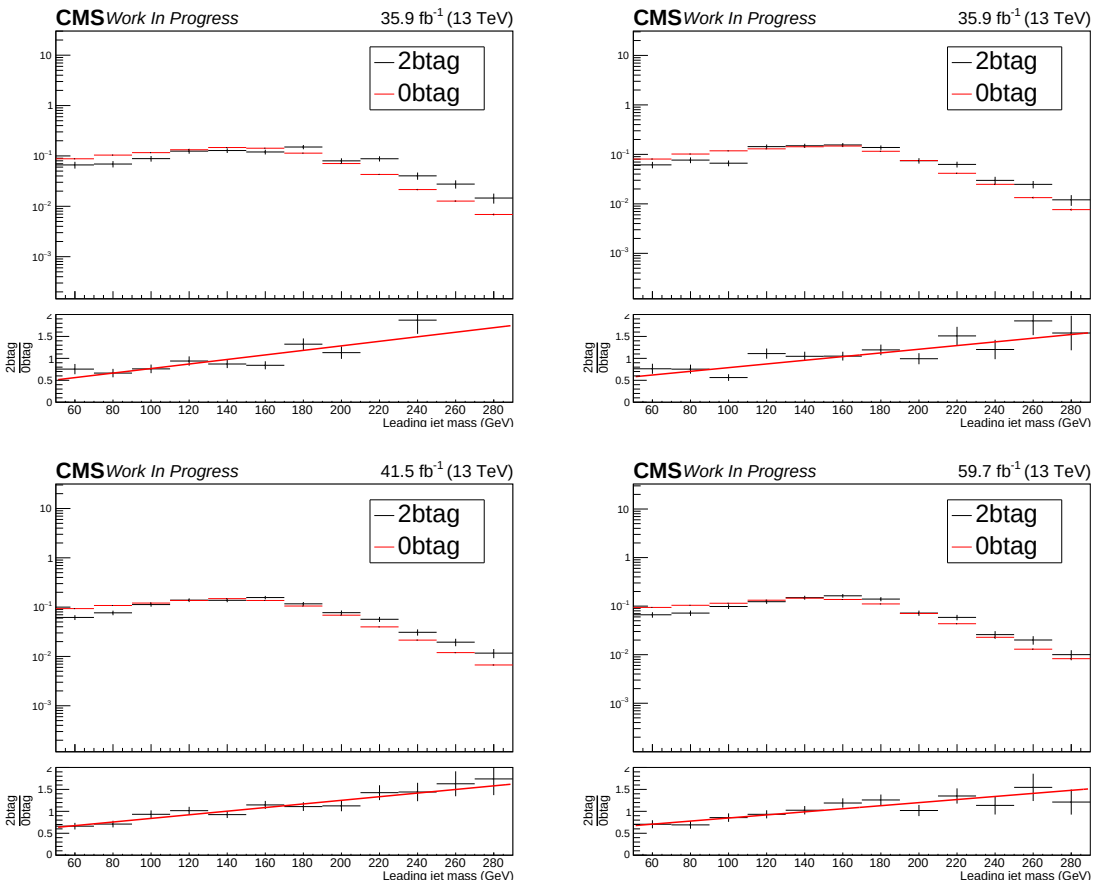

Figure 7.10: Closure test in the QCD simulation for the shape of the soft drop mass of the leading jet in the two possible b-tagging requirements (none, or both jets contain a b-tagged subjet) for the 4 data taking periods. Top left is 2016_preVFP, 2016_postVFP top right, 2017 is bottom left and 2018 is bottom right. A ratio plot is shown in the bottom of each figure. A linear trend can be seen which is used later to correct this effect.

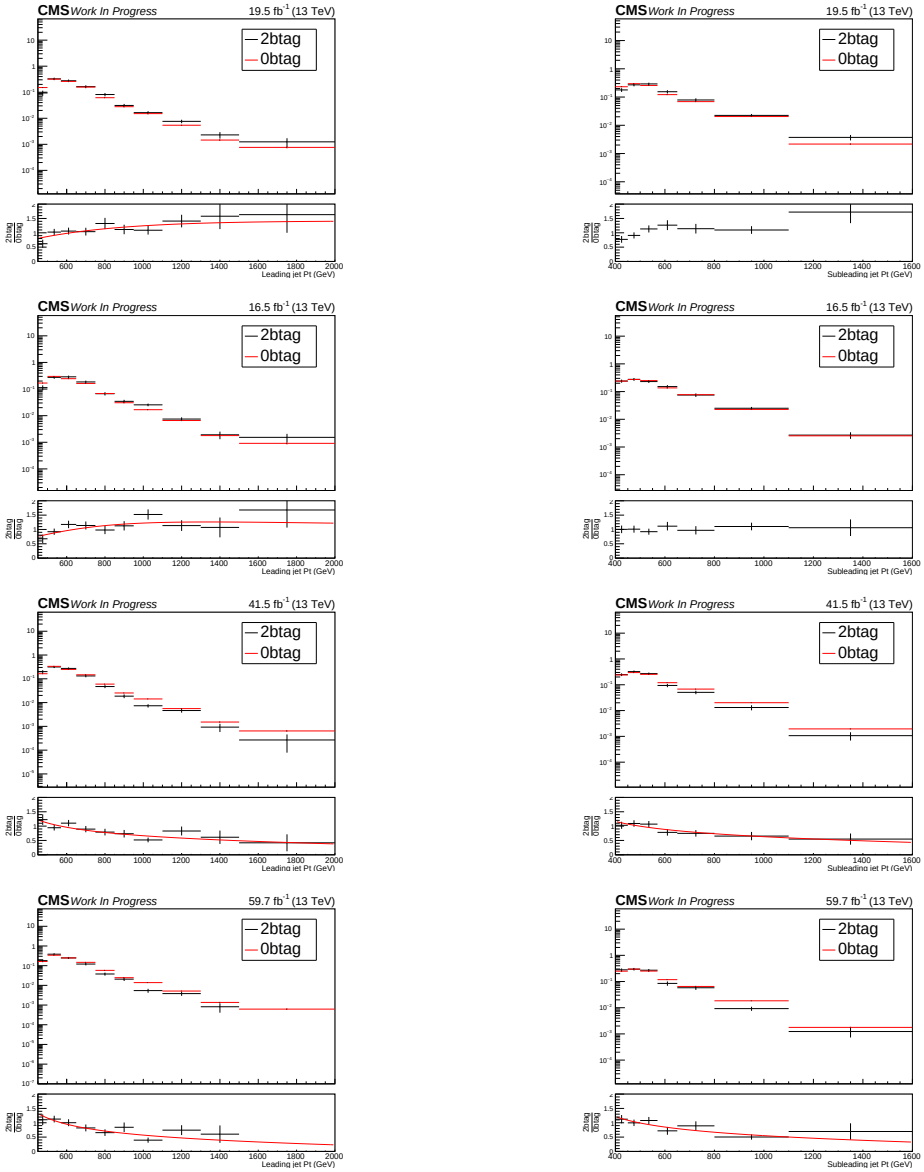

Figure 7.11: Closure test in the QCD simulation for the shape of the leading (left) and subleading (right) p_T of the jet. First row is for 2016_preVFP, second for 2016_postVFP, third 2017 and fourth 2018. A fit is performed for a correction factor in the plots that require it.

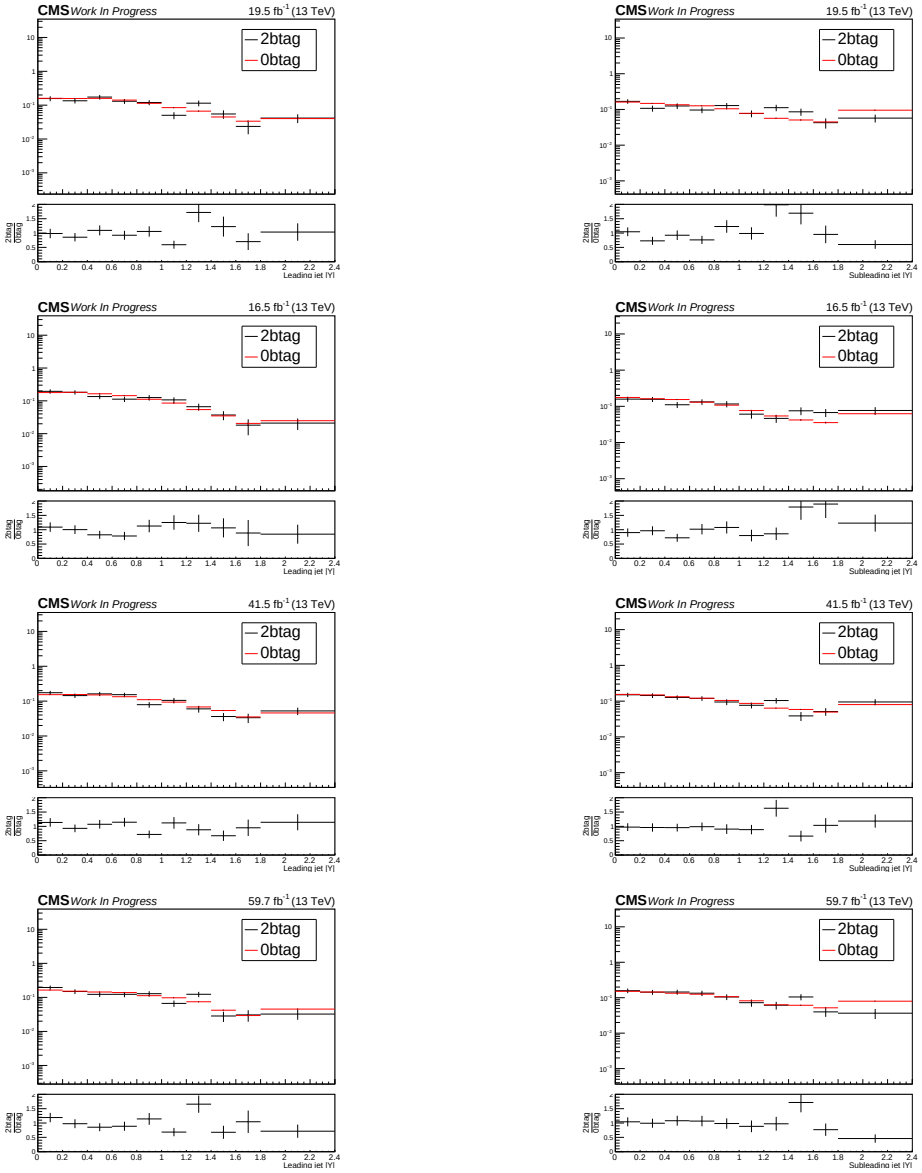

Figure 7.12: Closure test in the QCD simulation for the shape of the leading (left) and subleading (right) $|y|$ of the jet. First row is for 2016_preVFP, second for 2016_postVFP, third 2017 and fourth 2018.

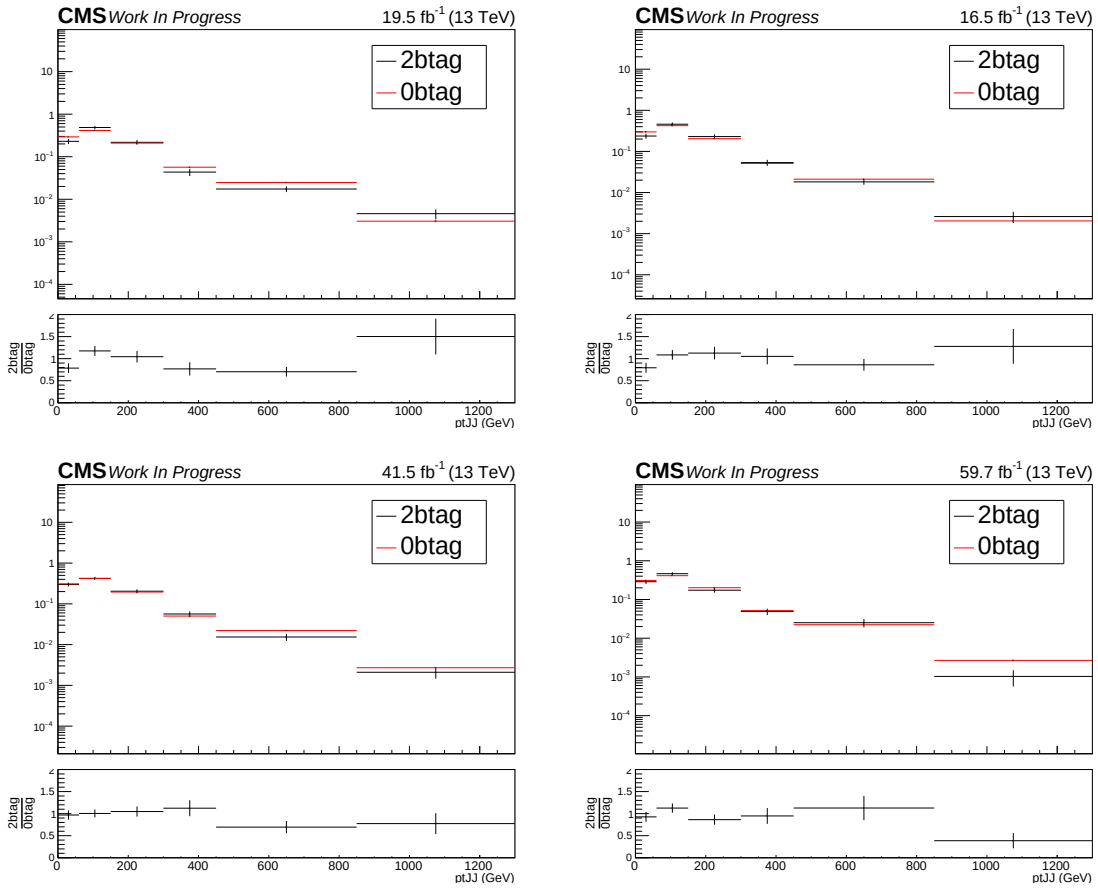

Figure 7.13: Closure test in the QCD simulation for the shape of the pT_{JJ} of the system. First row is for 2016_preVFP, second for 2016_postVFP, third 2017 and fourth 2018.

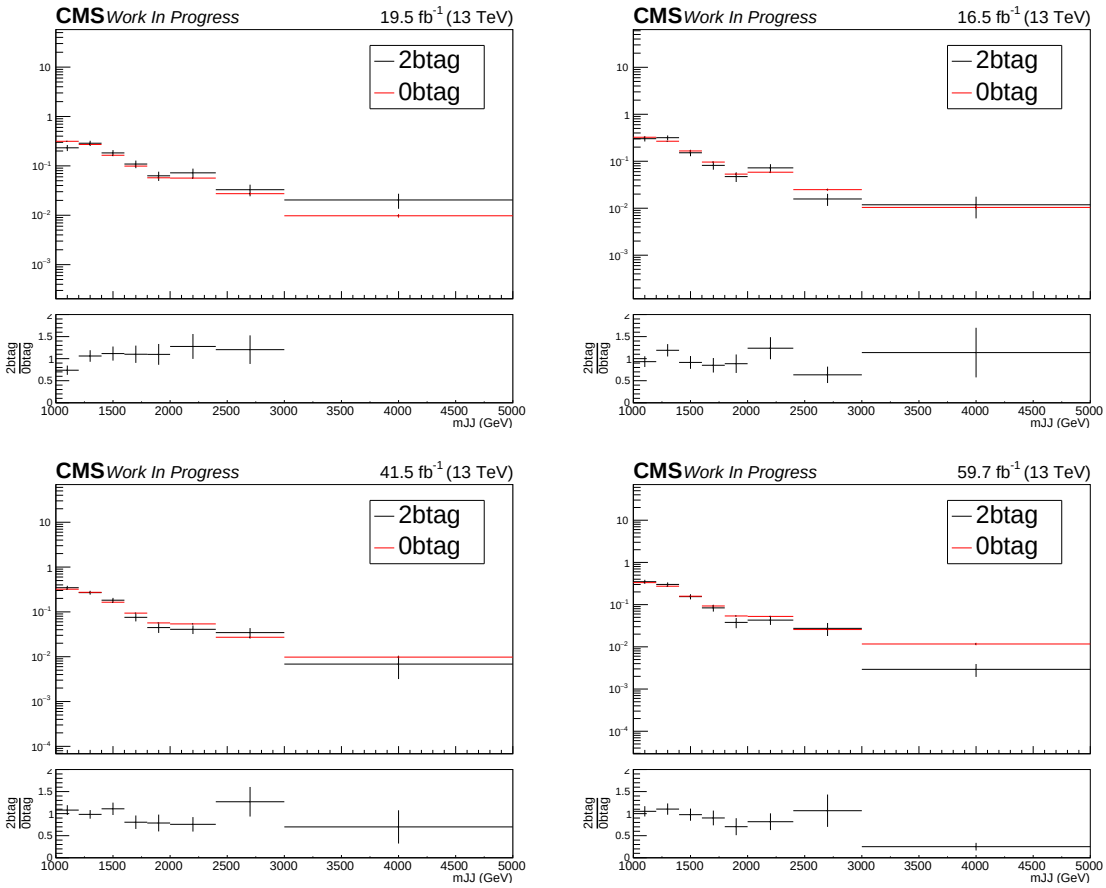

Figure 7.14: Closure test in the QCD simulation for the shape of the m_{JJ} of the system. First row is for 2016_preVFP, second for 2016_postVFP, third 2017 and fourth 2018.

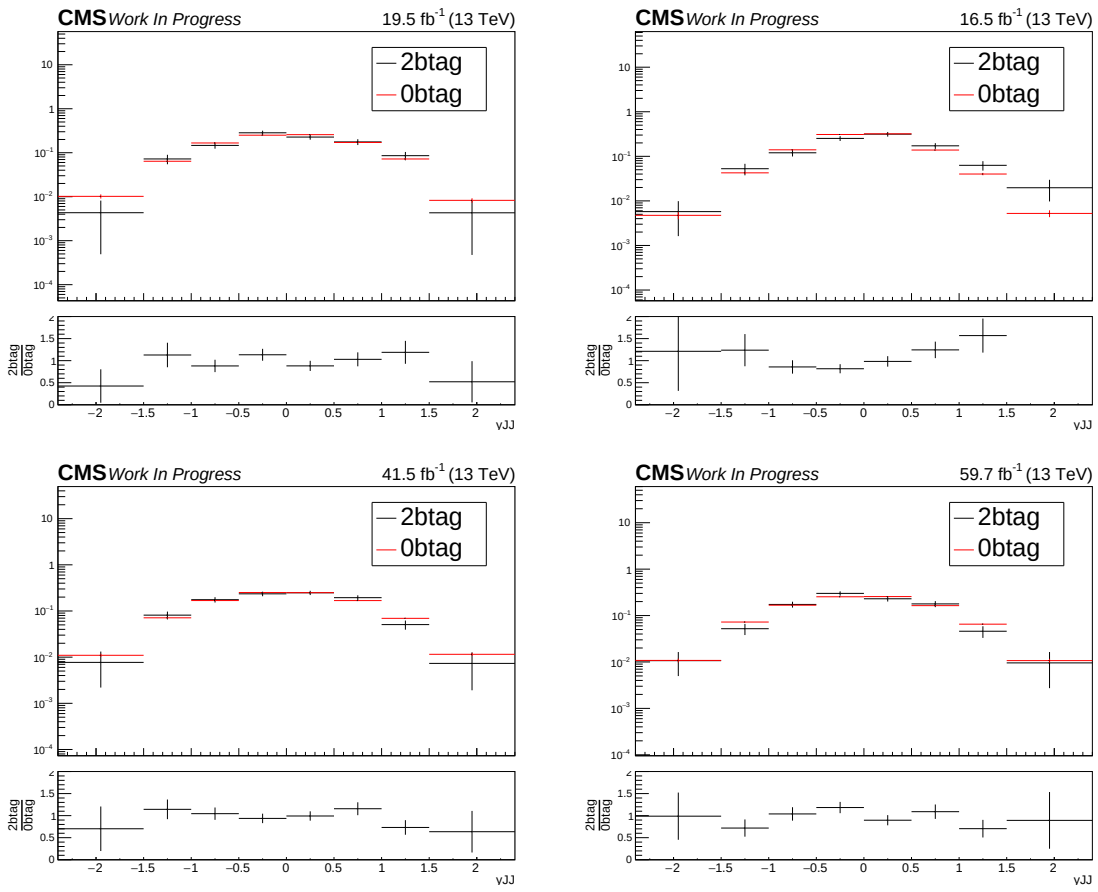

Figure 7.15: Closure test in the QCD simulation for the shape of the y_{JJ} of the system. First row is for 2016_preVFP, second for 2016_postVFP, third 2017 and fourth 2018.

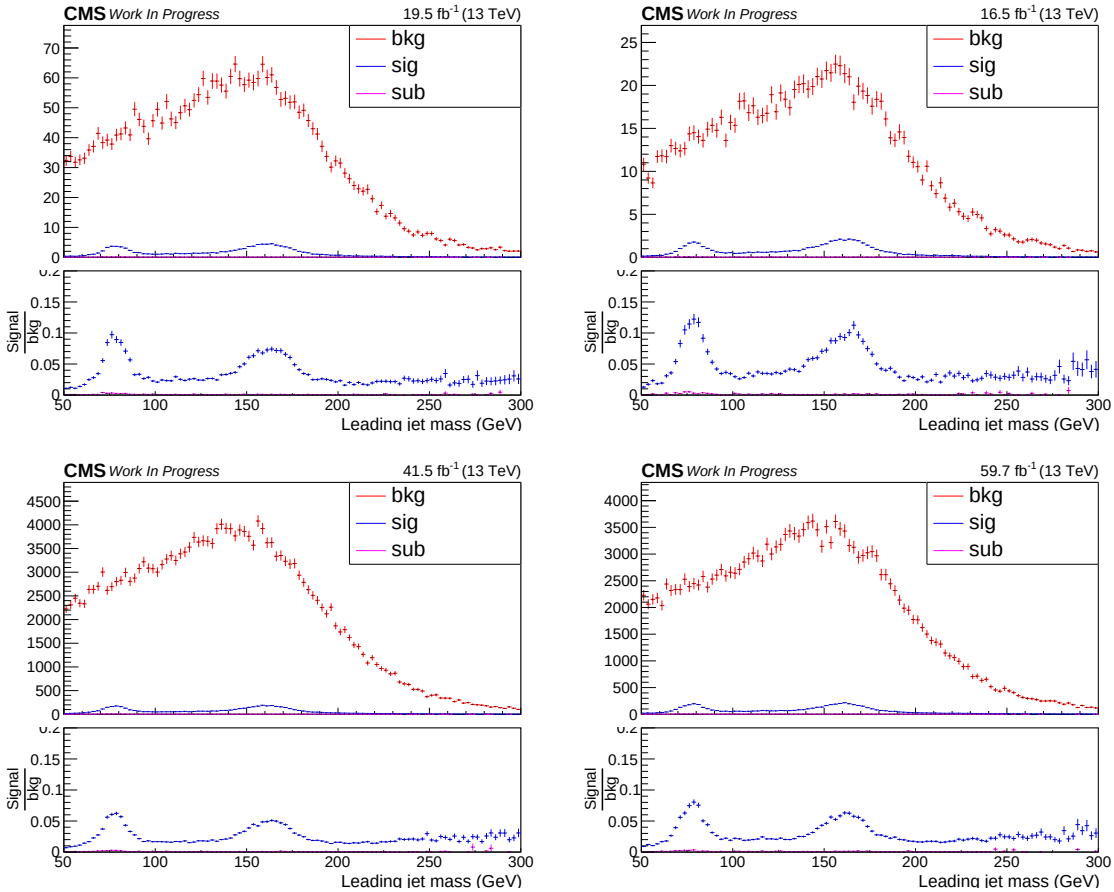

Figure 7.16: Contamination plots for the QCD control region. They display the contamination of the CR from $t\bar{t}$ and subdominant backgrounds as a function of the soft drop mass of the leading p_T jet. On the top left for 2016 preVFP, and top right 2016 postVFP and in the bottom left for 2017 and right for 2018. On top we present the expected yields for each process in the Control Region. The ratio of the plots shows the percentage of contamination in the CR with respect to the present QCD in the CR for each year.

Although reverting the b-tagging requirement give us the shape of the QCD in the SR we need to see if by applying this selection criteria only QCD is selected and not actual signal. If this is not the case some adjustment needs to be made. The presence of signal in our background selection region is called contamination. Figure 7.16 shows the contamination of the QCD control region. The selection criteria are used in monte carlo sample simulating the various processes QCD, $t\bar{t}$

and other subdominant processes. It can be seen that in fact the CR selection has a significant contribution $\sim$ 10% from $t\bar{t}$ events. This contribution is removed before the CR selection is used for background subtraction.

Using the CR a distribution is obtained for the QCD contribution but the yield, the normalization factor needs to be determined. For this reason a fit is performed. The fit is performed in the data and in particular in the SR_A selection. This region is selected as it is wider and includes more background allowing for a better modeling of the various shapes from the fit procedure. The fit is performed on the distribution of the soft drop mass of the leading jet. The equation used for is the following:

$$D_{SR_A}(m^t) = N_{t\bar{t}}T(m^t; k_{\text{scale}}, k_{\text{res}}) + N_{\text{qcd}}(1 + k_{\text{slope}}m^t)Q_{CR_A}(m^t) + N_{\text{bkg}}B(m^t) \quad (7.7)$$

where D_{SR_A} is the distribution of the data where the fit is performed. The shapes $T(m^t)$ and $B(m^t)$ are the shapes of the signal and subdominant background processes and are taken from simulation. The shape of the QCD $Q(m^t)$ is taken from the extended control region (CR_A) from the data.

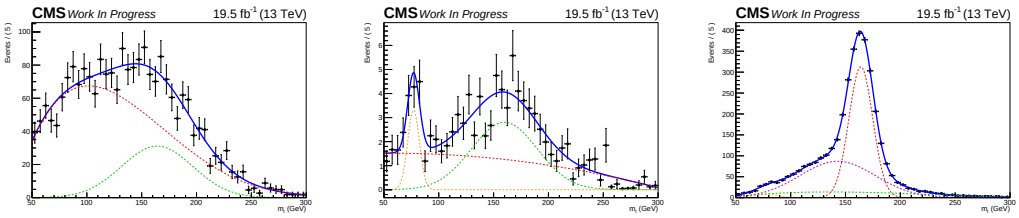

Figure 7.17: Templates of QCD (left), taken from data, of the subdominant backgrounds (center), taken from the simulation and templates of the tt (right) taken from simulation for 2016_preVFP. The different lines show the individual components (frozen in the fit) used to describe the shapes. The QCD shape is composed of a smooth polynomial and a Gaussian, while the shape of the subdominant backgrounds contains a smooth polynomial and two Gaussians (one describes the W resonance from the single top and WJets processes and the other describes the broader peak from the kinematic selections). On the right, template of the tt signal taken from the simulation. The shape consists of a smooth polynomial and two Gaussians (one describes the W resonance from unmerged top decays and the other describes the fully merged top resonance).

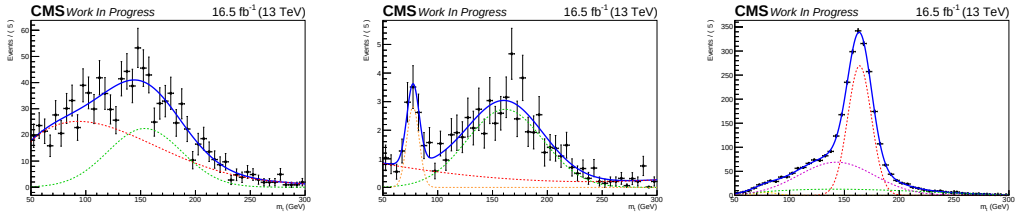

Figure 7.18: Templates of QCD (left), taken from data, of the subdominant backgrounds (center), taken from the simulation and templates of the tt (right) taken from simulation for 2016_postVFP. The different lines show the individual components (frozen in the fit) used to describe the shapes. The QCD shape is composed of a smooth polynomial and a Gaussian, while the shape of the subdominant backgrounds contains a smooth polynomial and two Gaussians (one describes the W resonance from the single top and WJets processes and the other describes the broader peak from the kinematic selections). On the right, template of the tt signal taken from the simulation. The shape consists of a smooth polynomial and two Gaussians (one describes the W resonance from unmerged top decays and the other describes the fully merged top resonance).

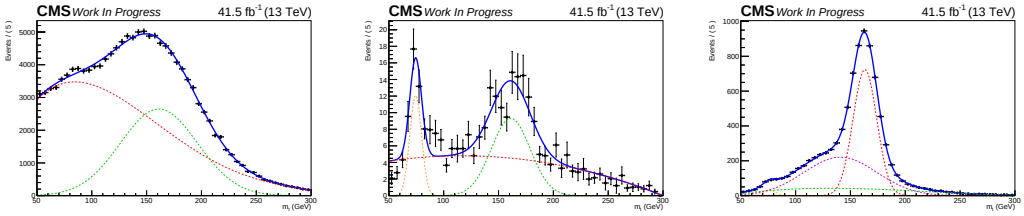

Figure 7.19: Templates of QCD (left), taken from data, of the subdominant backgrounds (center), taken from the simulation and templates of the tt (right) taken from simulation for 2017. The different lines show the individual components (frozen in the fit) used to describe the shapes. The QCD shape is composed of a smooth polynomial and a Gaussian, while the shape of the subdominant backgrounds contains a smooth polynomial and two Gaussians (one describes the W resonance from the single top and WJets processes and the other describes the broader peak from the kinematic selections). On the right, template of the tt signal taken from the simulation. The shape consists of a smooth polynomial and two Gaussians (one describes the W resonance from unmerged top decays and the other describes the fully merged top resonance).

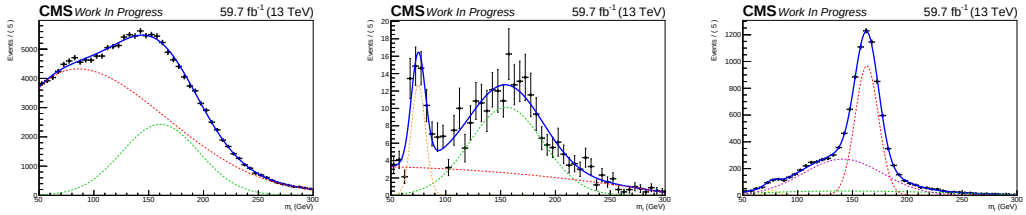

Figure 7.20: Templates of QCD (left), taken from data, of the subdominant backgrounds (center), taken from the simulation and templates of the tt (right) taken from simulation for 2018. The different lines show the individual components (frozen in the fit) used to describe the shapes. The QCD shape is composed of a smooth polynomial and a Gaussian, while the shape of the subdominant backgrounds contains a smooth polynomial and two Gaussians (one describes the W resonance from the single top and WJets processes and the other describes the broader peak from the kinematic selections). On the right, template of the tt signal taken from the simulation. The shape consists of a smooth polynomial and two Gaussians (one describes the W resonance from unmerged top decays and the other describes the fully merged top resonance).

The various shapes used can be seen in the following figures, Figure 7.17 - Figure 7.20. The fit is performed in the soft drop mass variable. The closure test for this variable, Figure 7.10 shows some linear trend in the ratio plot. To account for this discrepancy the correction factor k_{slope} has been introduced and left free in the fit. The correction factor is applied in the form (1 + $k_{slope}m^t$). Two more free parameters are left in the fit k_{scale} and k_{ref}. These are to account for any discrepancies in the modeling of the shape of the top quark $T(m^t)$ and are again left free to be determined by the fit. The are used to change the σ and μ of the Gaussian used to model the top quark peak in the mass. Finally the normalization parameters (yields) $N_{t\bar{t}}$, N_{qcd}, N_{bkg} for all contributions are left free to be determined by the fit.

The fit is performed using the RooFit package [49]. The fit results can be seen in theFigure 7.21 - Figure 7.24 for each year respectively. The fitted parameters, i.e. the free parameters determined by the fit procedure can be seen in Table 7.17 and Table 7.18. For all years the $t\bar{t}$ yield is $\approx$ 35% smaller than what the MC simulations predict (Powheg + Pythia8). To quantify this we introduce the quantity $r_{t\bar{t}}$ called signal strength. It is defined as the number of events measured in data over the number of events predicted by the simulation. The respective values for the signal strength for each year can be seen in Table 7.16.

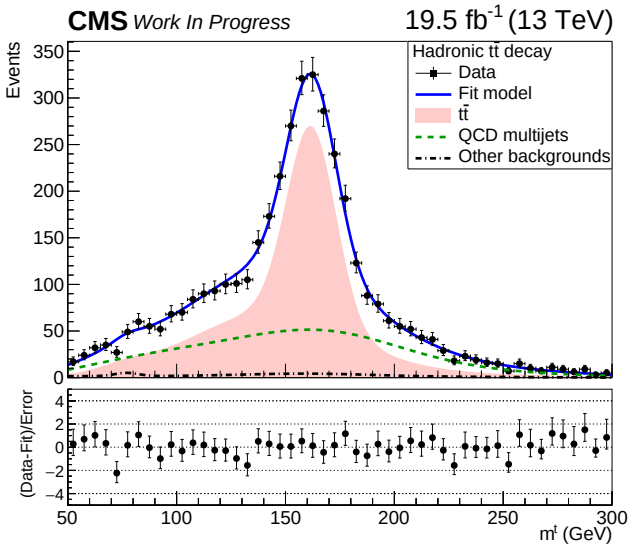

Figure 7.21: Result of the template fit on data for 2016 preVFP. The red line shows the $t\bar{t}$ contribution, the green line shows the QCD, and the brown line shows the subdominant background.

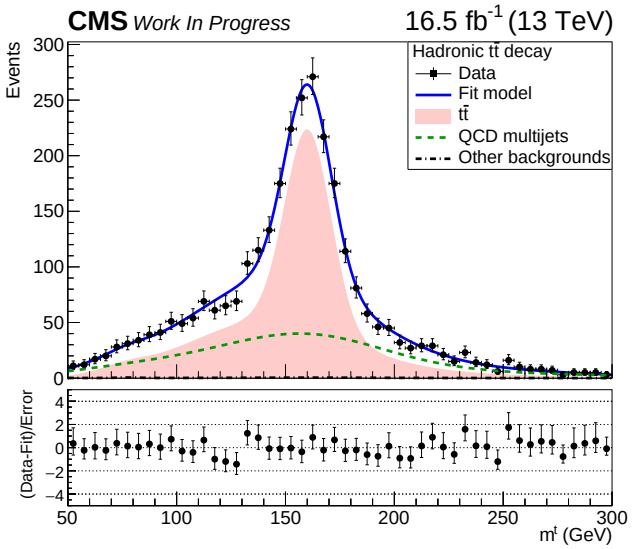

Figure 7.22: Result of the template fit on data for 2016 postVFP. The red line shows the $t\bar{t}$ contribution, the green line shows the QCD, and the brown line shows the subdominant background.

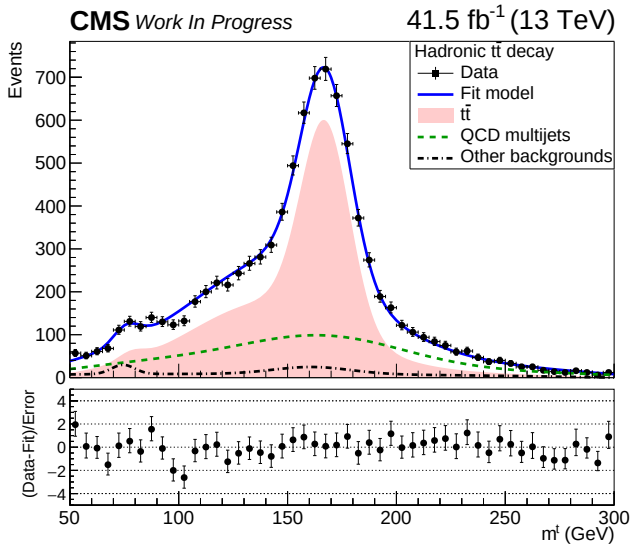

Figure 7.23: Result of the template fit on data for 2017. The red line shows the $t\bar{t}$ contribution, the green line shows the QCD, and the brown line shows the subdominant background.

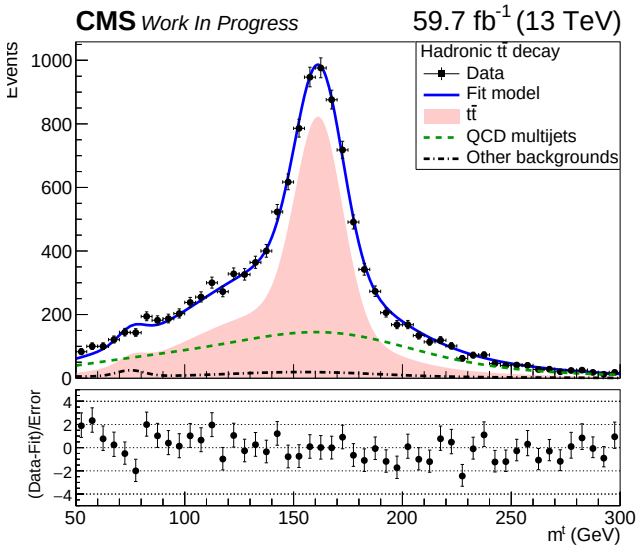

Figure 7.24: Result of the template fit on data for 2018. The red line shows the $t\bar{t}$ contribution, the green line shows the QCD, and the brown line shows the subdominant background.

Table 7.16: Results of the $t\bar{t}$ signal strength for each year.

Year	Value	Error
2016_preVFP	0.691	0.028
2016_postVFP	0.640	0.029
2017	0.665	0.018
2018	0.675	0.016

Table 7.17: Results of the fit in SR_A for 2016 preVFP left, 2016 postVFP right

Parameter	Value	Error
k_{res}	1.032	0.045
k_{scale}	0.987	3.36e-03
k_{slope}	6.91e-02	0.14
N_{bkg}	108	183
N_{qcd}	1321	212
$N_{t\bar{t}}$	2543	104

Parameter	Value	Error
k_{res}	0.976	0.049
k_{scale}	0.979	3.61e-03
k_{slope}	0.185	0.136
N_{bkg}	10	848
N_{qcd}	951	639
$N_{t\bar{t}}$	1977	92

Table 7.18: Results of the fit in SR_A for 2017 left, 2018 right

Parameter	Value	Error
k_{res}	1.03	0.03
k_{scale}	1.025	2.74e-03
k_{slope}	0.2	0.07
N_{bkg}	526	159
N_{qcd}	2514	210
$N_{t\bar{t}}$	6008	160

Parameter	Value	Error
k_{res}	1.01	0.026
k_{scale}	0.992	0.002
k_{slope}	5e-02	3.24e-03
N_{bkg}	479.3	222
N_{qcd}	3900	251
$N_{t\bar{t}}$	7664	177

The signal extraction procedure can be summarised in the following equation:

$$S(x) = D(x) - R_{\text{yield}} N_{\text{qcd}} Q(x) - B(x) \tag{7.8}$$

where x is the variable of interest and S is the desired signal distribution. D is the measured distribution in the SR of the data. B is the contribution of the subdominant background processes taken from the MC. Q is the shape of the QCD multijets distribution taken from the CR of the data sample. That means we invert the b-tagging requirement on the SR selection criteria. N is the number of QCD elements taken from the fit procedure. Since this value is measured in the extended signal region SR_A a transfer factor needs to be added in order to transfer

this value to the SR. This is defined as:

$$R_{\text{yield}} = \frac{N_{SR}}{N_{SR_A}} \tag{7.9}$$

and is taken from simulation.

7.7 Data vs Monte Carlo

A sanity check for the values calculated in the previous section is to do a comparison between the number of events predicted by MC using the scale factors and data. In the following figures, the number of events selected from the data sampled are compared with the number of events predicted by the MC simulation. The simulation samples used are the $t\bar{t}$, QCD and subdominant backgrounds, (W+jets, Z+jets, Single Top). The $t\bar{t}$ number of events is scaled down using the signal strength factor calculated in the previous section. It can be seen tat there is a fairly good agreement. Additionally, figures 7.25 - 7.30 contain comparison plots of the data selection vs the MC predictions. Again the MC simulations are scaled according to the results of the fit from the previous section.

Table 7.19: Expected and observed event yields in the signal region for all analyzed years.

Process	Yield 2016 preVFP	Yield 2016 postVFP	Yield 2017	2018
$t\bar{t}$	2252	1952	5358	6840
QCD	434	329	1182	1280
Subdominant (W+jets, Z+jets, Single Top)	10	8	38	47
Data	2187	1654	4818	6205

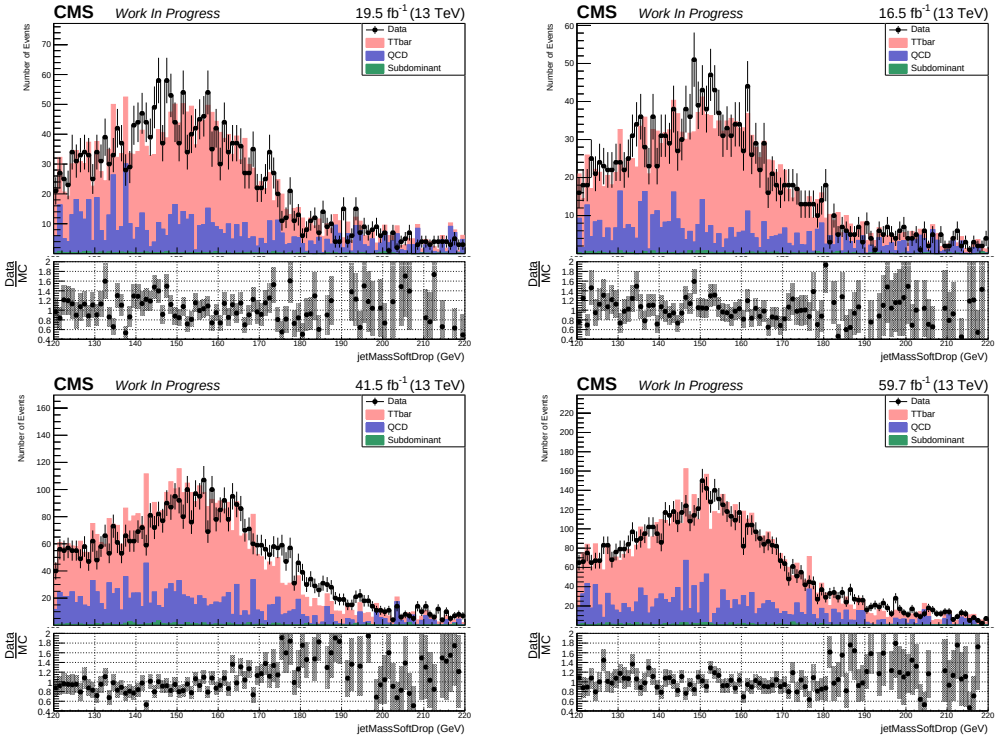

Figure 7.25: SoftDrop mass of the leading AK8 jet after the baseline selection with both AK8 jets containing a b-tagged subjet, with the BDT > 0.0 cut (2016 preVFP top left, 2016 postVFP right), > 0.1 for 2017 bottom left and > 0.2 for 2018 bottom right

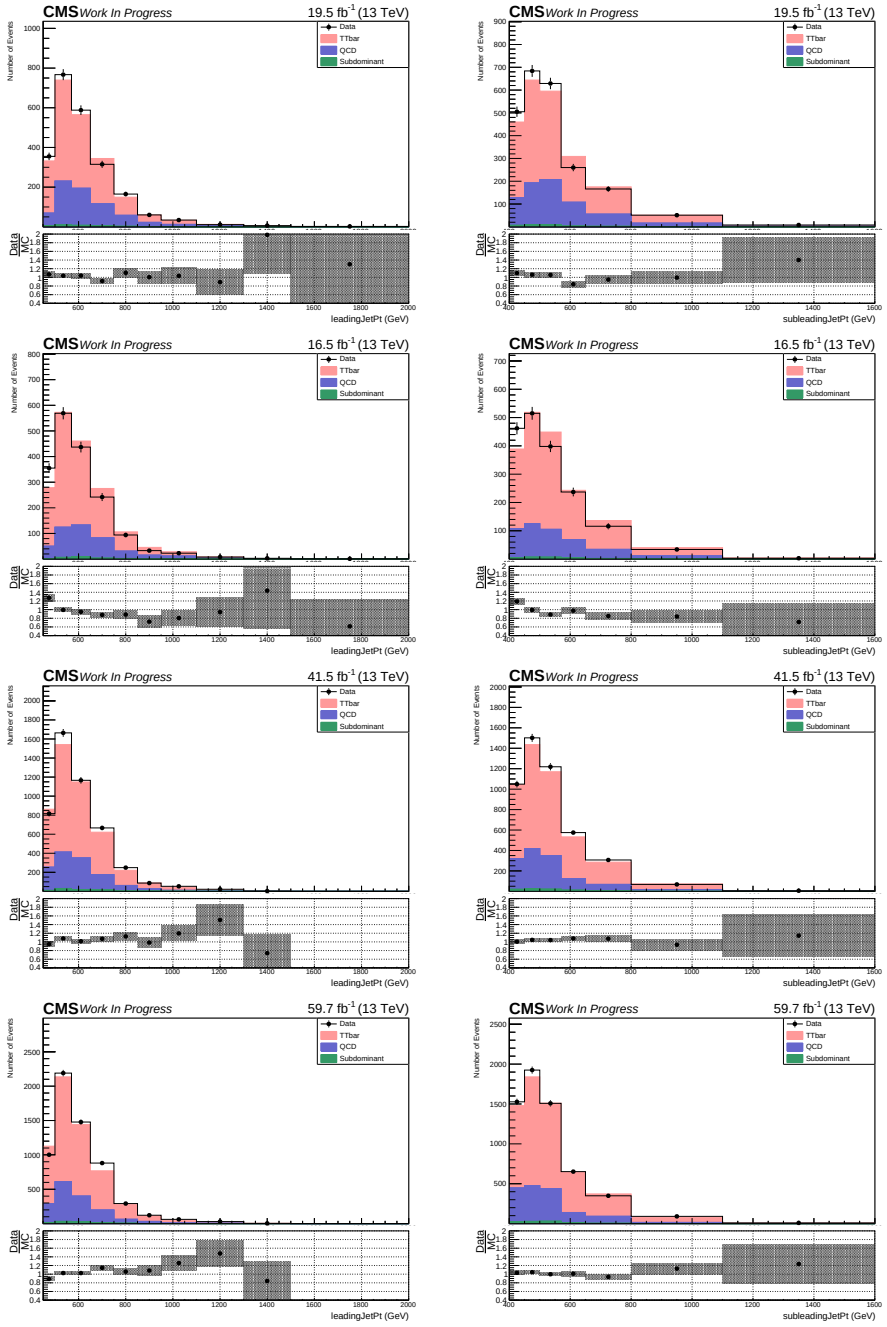

Figure 7.26: Distributions for the transverse momentum of the leading and sub-leading jet in the signal region. On left, the distribution for the leading jet and on the right for the sub-leading jet. Starting from top: 2016 preVFP, 2016 postVFP, 2017, and 2018

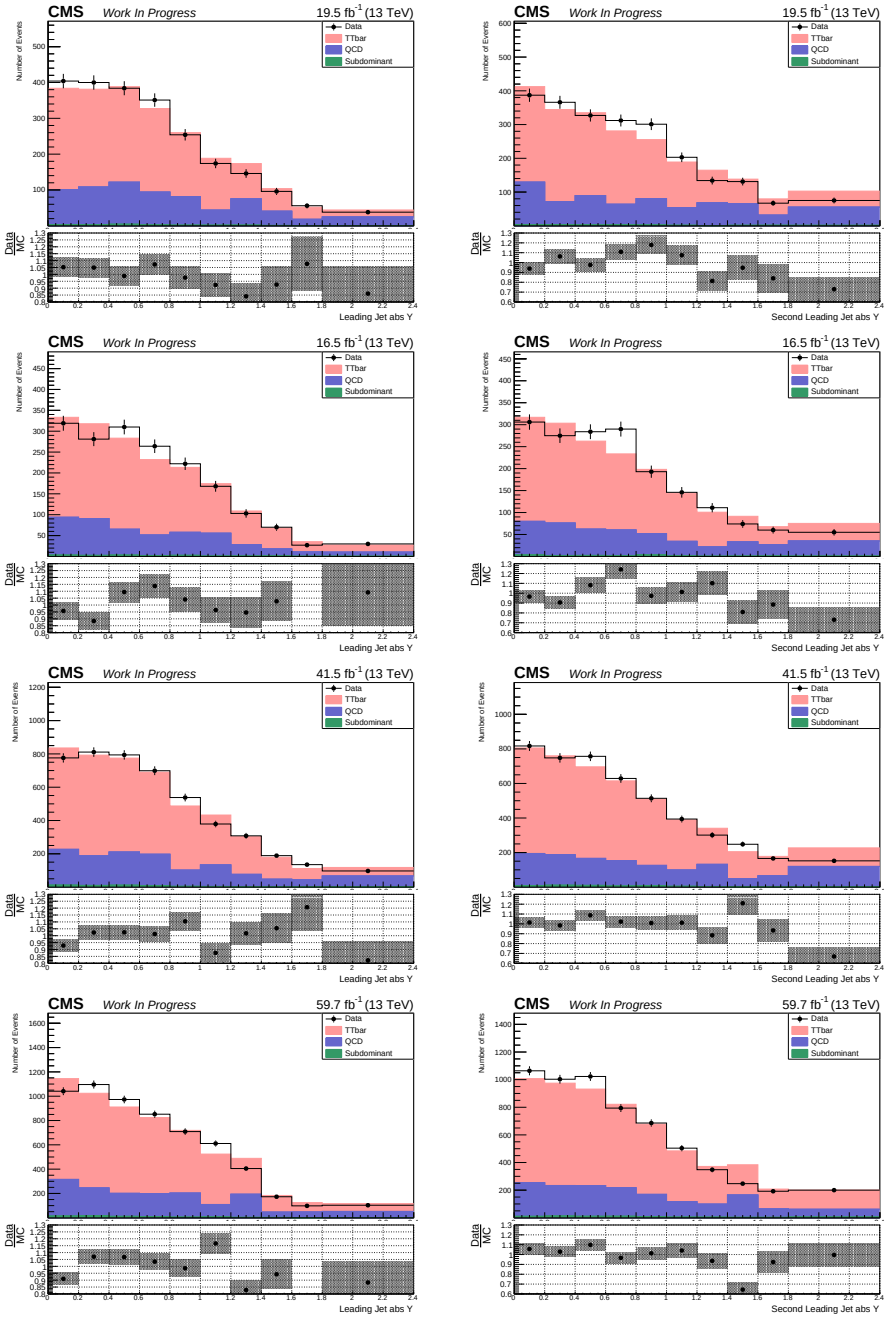

Figure 7.27: Distributions of the rapidity for the leading and sub-leading jet in the signal region. On left, the distribution for the leading jet and on the right for the sub-leading jet. Starting from top: 2016 preVFP, 2016 postVFP, 2017, and 2018

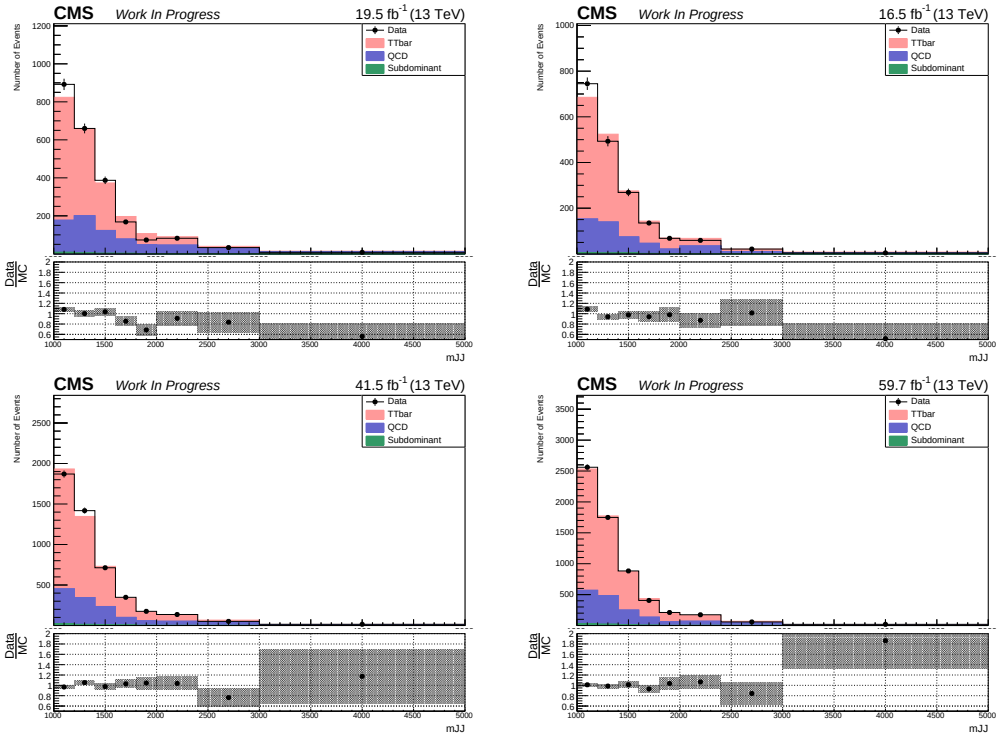

Figure 7.28: Distributions of the rapidity for the dijet mass of the system in the signal region. On top left for 2016 preVFP, top right for 2016 postVFP, bottom left for 2017 and bottom right for 2018.

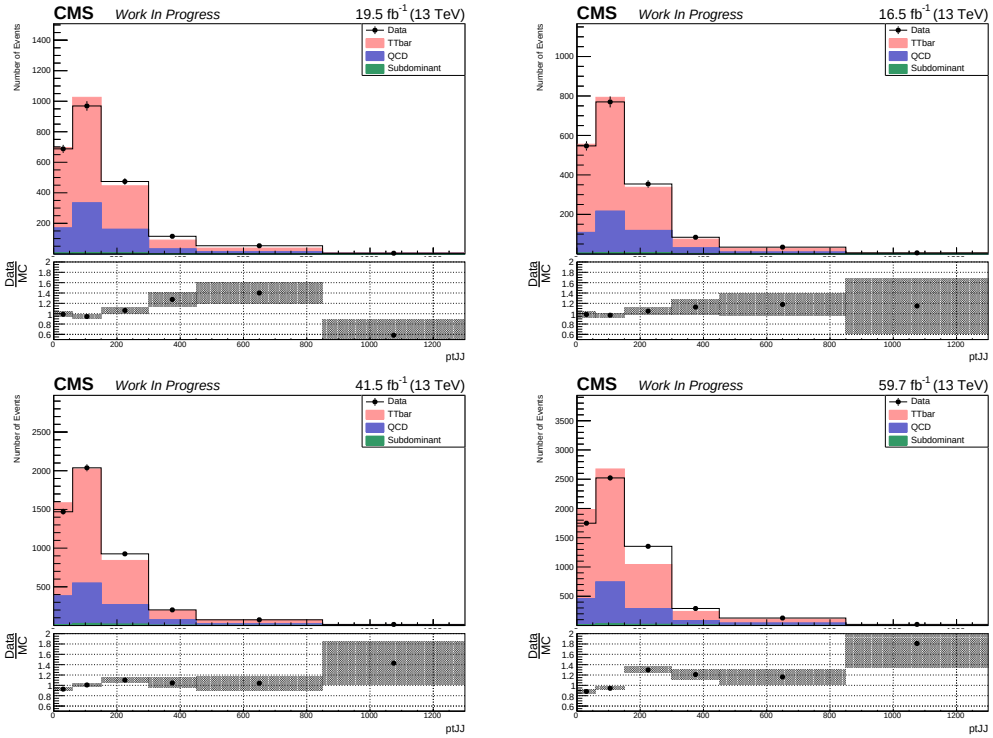

Figure 7.29: Distributions of the rapidity for the for the transverse momentum of the system in the signal region. On top left for 2016 preVFP, top right for 2016 postVFP, bottom left for 2017 and bottom right for 2018.

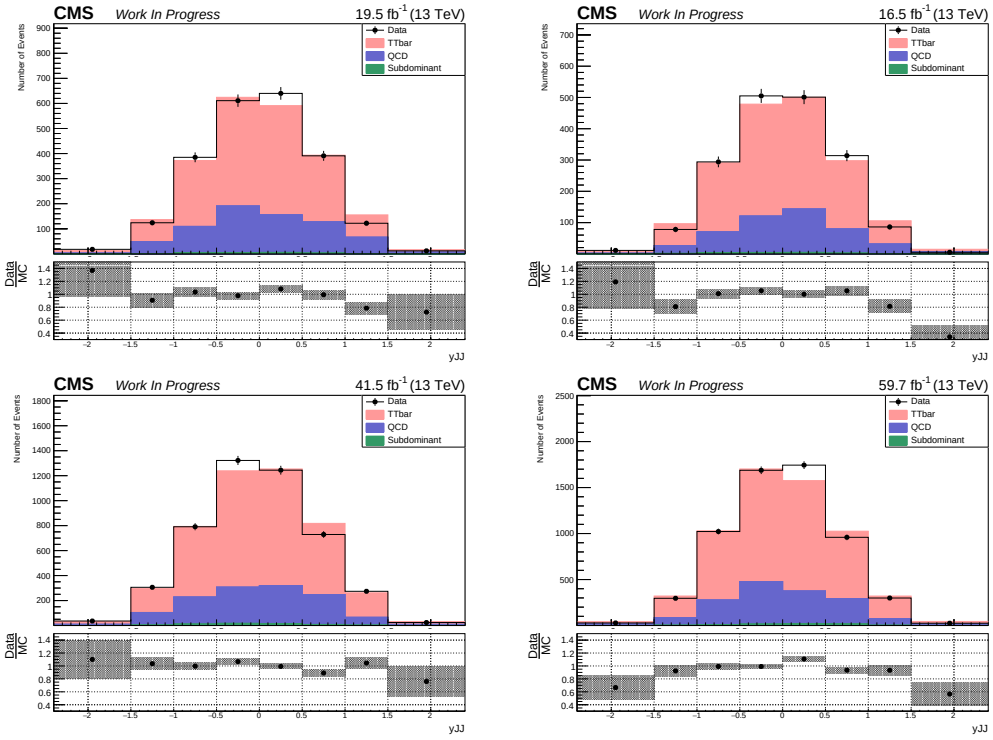

Figure 7.30: Distributions of the rapidity for the for the rapidity of the system in the signal region. On top left for 2016 preVFP, top right for 2016 postVFP, bottom left for 2017 and bottom right for 2018.

7.8 Systematic Uncertainties

Performing a measurement always contains some short of error. Due to the statistical nature of particle physics there is some inherent uncertainty in the measurement which is expressed as a statistical error. This is called random error as performing the same experiment over and over again it will result in some times values larger than the true value and some other times in lower values. Overall it will give us a good estimate of the actual value. Due to the incapability of predicting the actual result of the experiment this error is called random. There is another type of error which is called systematic error. This error can be of many forms but it will always give us a wrong result in the same direction compared to the correct one. For example measuring a temperature with a thermometer that is wrongly calibrated with and offset of 1 degree all our results will be systematically wrong by 1 degree. Figure 7.31 shows a schematic conceptual representation of the two types of errors. This section will discuss and review the systematic uncertainties used in this analysis.

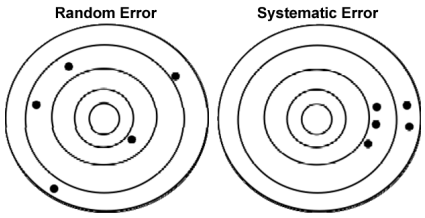

Figure 7.31: Conceptual schematic representation of random and systematic error. Figures are taken from [50]

In this particular analysis systematic uncertainties are divided in two categories experimental and theoretical. Experimental uncertainties refer to uncertainties that are inserted in the object reconstruction where data and simulation predictions do not match. These affect the formation of the objects that are used as a basis in the analysis. The theoretical uncertainties affect mostly the modeling of the various processes and affect the method used in order to extract the final results from the detector level to the parton and particle level. For each source of systematic uncertainty a set of variations are used. In addition, for each of these variations the result, cross section, is re-calculated. Then the difference of the new measured value with respect to the nominal result is taken as the effect of this type of uncertainty. Since the final results are compared with a set of theoretical models, the same procedure for calculating uncertainties is also used for the

theoretical models. For simulation this is done only for the theoretical uncertainties and not the experimental. In the following part of the section we will list the various uncertainties used.

1. Experimental Uncertainties refer to uncertainties that have to do with the measurement conducted by the detector and the various reconstruction techniques. The ones considered in this analysis are the following:

   - *Luminosity*: this refers to the efficient determination of the luminosity recorded by the detector during the data taking period. During the data taking period the rate of collisions is measured online. The integrated luminosity is measured offline where the data collected are re-evaluated in terms of quality. The luminosity uncertainty for each year 1.2%, 2.3% and 2.5% for 2016, 2017 and 2018 respectively. The uncertainty in the measurement of the total integrated luminosity is 1.6% [51], [52].
   - *Pileup*: refers to the efficient determination of the number of pileup during the interactions. The values affecting this parameter is the total inelastic cross section. The number used is 69.2 mb. To estimate its impact the values is varied by $\pm 4.6\%$. The total effect in the result is negligible, less than 1%.
   - *b-tagging*: this uncertainty is related to the efficiency of identifying the AK4 subjets inside the AK8 jets as b quarks, [53]. Scale factors are applied in order for the simulation predictions to match the data one. Instructions on how the scale factors are calculated and applied can be found here [54]. Variations of theses scale factors are applied to identify the uncertainty on the cross section. This uncertainty is one of the leading experimental uncertainties which is expected since the method used relies heavily on b-tagging as it is used to both identify top quarks and to also identify and subtract QCD background. The total uncertainty is around 7%. The value of the uncertainty for this source seems to be similar for all variables across their distributions. In the measurement of the normalized distributions the effects of this uncertainty seem to cancel out.
   - *Jet energy scale*: refers to the unceratinty of the measured energy of each reconstructed jet. Since jets are the main object used in the analysis this uncertainty is by far the leading experimental uncertainty. To evaluate it 30 different independent sources of uncertainty are taken

into account according to prescription used throughout the CMS experiment, [55]. For each of these variations a new jet collection is created meaning that each jet is reconstructed again and the whole event selection procedure is repeated from scratch. This does not only change the energy distribution of the jets but can alter the top candidates themselves meaning that a jet that was passing the criteria before the application of a variation can now fail to pass them or the opposite. The effect for this uncertainty is measured around 10% but can be much bigger in objects with really high p_T.

- *Jet energy resolution*: refers to the uncertainty in the jet resolution. It is determined by smearing the jets according to the JER uncertainty [55]. The effect it at the level of 2%.
- *QCD background prediction*: this source refers to the background normalization predicted by the fit. Here the uncertainty calculated by the fit procedure is used.
- *Trigger*: This uncertainty accounts for differences between the trigger efficiency measured in data and simulation. The impact is below 1%.

2. Theoretical Uncertainties consist of uncertainties used mostly in simulation of the process we are trying to measure as well as the various background processes. As MC samples are used in various steps of this analysis, considering systematics on theory is very important. The theoretical systematic uncertainties used in this analysis are

   - *Parton Distribution Functions*: this has to do with the PDFs used in the MC samples. They are used in the calculation of the cross section but also for modelling the actual collisions that take place in the experiment. As mentioned also above, the pdf used is the NNPDF 3.1 [56], [57]. The uncertainty is calculated using 100 replicas of the nominal PDF. The variations are applied in the form of weights.
   - *Renormalization (μ_R) and factorization (μ_F) scales* are two parameters used in the simulation for the calculation of the matrix elements (ME). The values used for these parameters can affect the final result. These values are varied by a factor of 1/2 or 2 leading to a total of 6 combinations. The variations are applied in the form of weights.
   - *Strong coupling constant*: Again is applied in the form of weights that correspond to the calculation of a different matrix element.

- *Matrix element −parton shower matching*: This refers to the variation of the h_{damp} parameter used in the POWHEG generator. This is a parameter that regulates the radiation of the event. Alternative samples where used that have the values of $h_{damp} = m_t$ and $h_{damp} = 1.379^{+0.926}_{-0.502} m_t$.
- *Underlying event tune*: refers to the additional interactions that happen besides the main beam interaction. For the modeling alternative dedicated samples are used where the CP5 parameters are varied by $\pm\sigma$.
- Initial and Final state Radiation (ISR, FSR): refer to radiation that happens before or after the interaction respectively from the particles. It is affected by the chosen strong coupling constant α_S. Variations are calculated by applying weights.
- Color reconnection

The effect of the uncertainties discussed in this section for the total data sample used can be seen in Figures 7.44 - 7.50.

7.9 Fiducial measurement

Before performing the measurement of the differential cross section in the parton and particle level the measurement if performed in what is called the fiducial level. The fiducial level is the detector level with all the selection cuts, described above applied. The fiducial differential cross section is defined as follows:

$$\frac{d\sigma_i^{\text{fid}}}{dx} = \frac{S_i}{\mathcal{L} \cdot \Delta x_i} \tag{7.10}$$

where subscript i refers to each bin of the distribution and x is the variable of interest. S refers to the signal yield after the selection criteria are applied, $\mathcal{L}$ is the integrated luminosity and Δx_i is the width of the ith bin of the distribution of the variable x. Apart from the differential cross section the normalized differential cross section is calculated. That is the shape of the differential cross section distribution. For the fiducial phase space it is defined as:

$$\frac{1}{\sigma^{\text{fid}}} \frac{d\sigma_i^{\text{fid}}}{dx} = \frac{1}{\sum_k S_k} \cdot \frac{S_i}{\mathcal{L} \cdot \Delta x_i}, \tag{7.11}$$

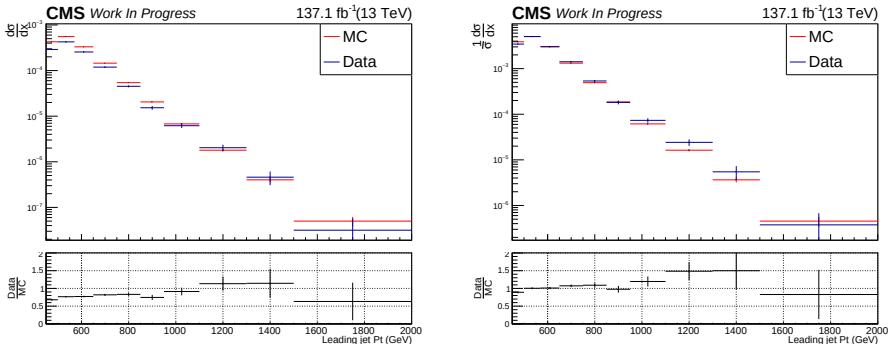

Figure 7.32: Fiducial differential cross section, absolute (left) and normalized (right), as a function of the leading top p_T. The bottom panel shows the ratio theory/data.

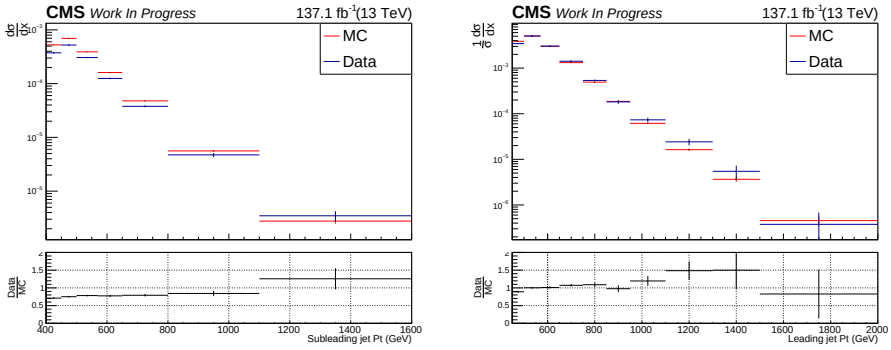

Figure 7.33: Fiducial differential cross section, absolute (left) and normalized (right), as a function of the second top p_T. The bottom panel shows the ratio theory/data.

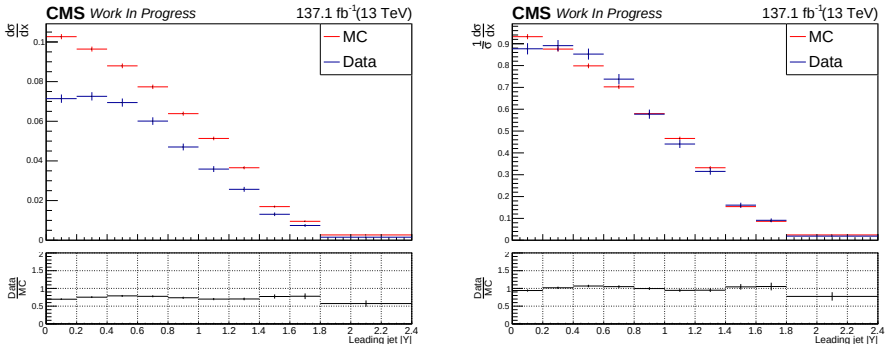

Figure 7.34: Fiducial differential cross section, absolute (left) and normalized (right), as a function of the leading top $|y|$. The bottom panel shows the ratio theory/data.

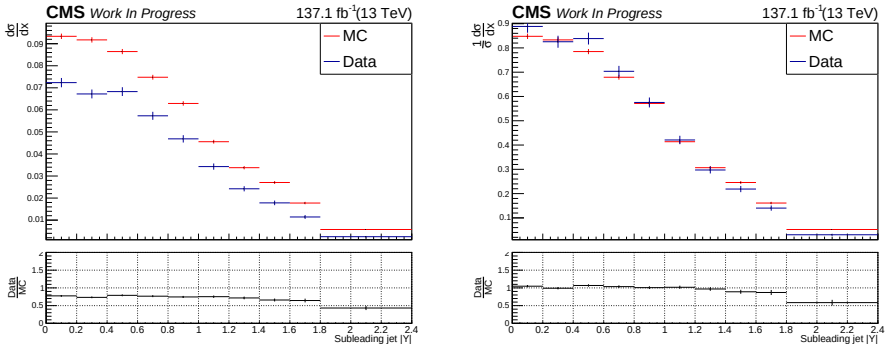

Figure 7.35: Fiducial differential cross section, absolute (left) and normalized (right), as a function of second top $|y|$. The bottom panel shows the ratio theory/data.

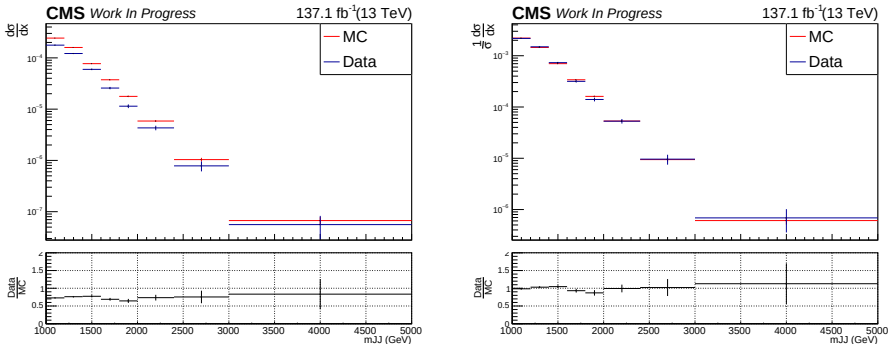

Figure 7.36: Fiducial differential cross section, absolute (left) and normalized (right), as a function of $m^{t\bar{t}}$. The bottom panel shows the ratio theory/data.

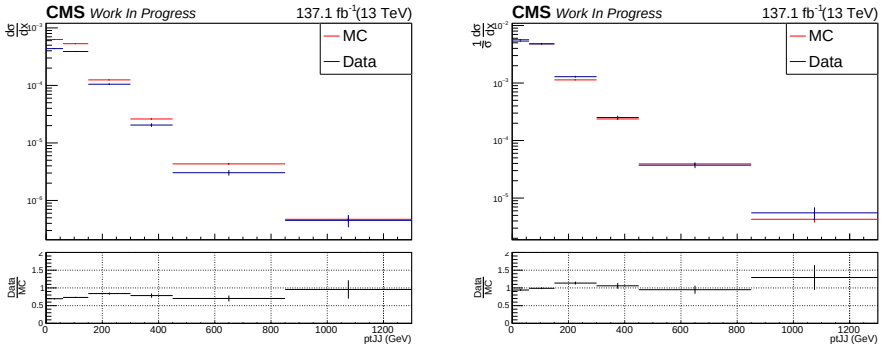

Figure 7.37: Fiducial differential cross section, absolute (left) and normalized (right), as a function of $p_T^{t\bar{t}}$. The bottom panel shows the ratio theory/data.

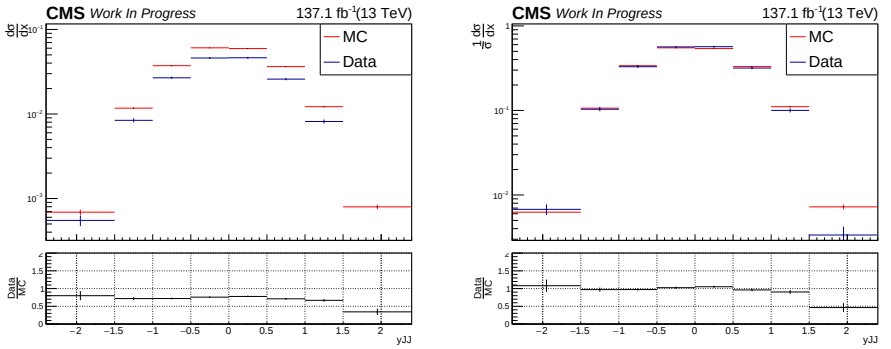

Figure 7.38: Fiducial differential cross section, absolute (left) and normalized (right), as a function of top $y^{t\bar{t}}$. The bottom panel shows the ratio theory/data.

7.10 Unfolded Measurement

The differential cross section measurement is done using the unfolding method for both the parton and particle phase spaces. Unfolding is a mathematical procedure used to solve what are called inverse problems. These are problem where we have the results and we want to get back to the cause of these results. It can be written as:

$$y = Ax \Rightarrow x = A^{-1}x \tag{7.12}$$

where y is the observation or the measurement, and x is the actual real value. Usually we can never measure the exact initial value due to various reason such as the resolution of the instrument used for the measurement. This can be modeled by A which is called the migration or response matrix. When solving an inverse problem we are trying to reverse the effect of A in order to get back x. In our use case, x is the real differential cross section and y is what is measured by the detector. As A we can indicate everything that is affecting the original generated particles in order for us to get the measured result y.

In our case the problem is in reality much more complex as we have to change from different phase spaces where particle decays happen. Since the resolution of our detector is finite and the migration matrix is not exactly diagonal we have a mix between the various bins. In this case a simple inversion of the matrix is not possible and we need to use more sophisticated techniques like unfolding which takes care of uncertainties, correlations etc. The equation used to derive the final result is:

$$\frac{d\sigma_i^{\text{unf}}}{dx} = \frac{1}{\mathcal{L} \cdot \Delta x_i} \cdot \frac{1}{f_{2,i}} \cdot \sum_j \left(R_{ij}^{-1} \cdot f_{1,j} \cdot S_j \right) \tag{7.13}$$

where $\mathcal{L}$ is the total integrated luminosity and Δx_i is the width of the i-th bin of the distribution of the respective observable x. S_j is the number of events in the j-th bin of the corresponding distribution in the reconstructed level. The factor $f_{1,j}$ is the fraction of reconstructed events in the j-th bin that have a corresponding event in then unfolded level (parton or particle). It is basically the probability that an event that is reconstructed and passes our selection criteria originated from an event that passed our selection criteria in the parton and the reconstructed level. It allows us to transfer from a space of events that pass our reco only criteria to events that pass the criteria in both phase spaces. R_{ij} is the the migration matrix between the i-th and j-th bins where i is in the unfolded level and j is in the reconstructed level. R_{ij}^{-1} is the inverse of the response matrix. The second fraction $f_{2,i}$ is the fraction of events at the unfolded level that have an equivalent event at the reconstructed level. It is the probability of an event that passed the parton

selection criteria to have passed both the parton and the reconstructed selection criteria. It allows us to move from the common phase space of events that pass both sets of criteria to the space of all the events that pass the parton criteria which is the area of interest. A visual representation of the unfolding procedure can be seen in Figure 7.39. Figures for the various quantities mentioned here can be seen in the results section 7.12.

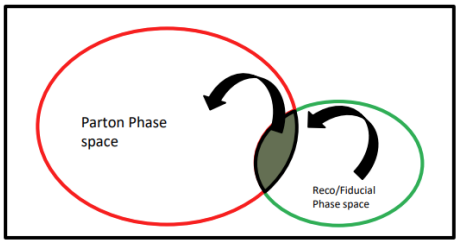

Figure 7.39: Visual representation of the unfolding procedure.

The binning of the various distributions is chosen based in two quantities, purity and stability. Purity is the fraction of reconstructed events that have an equivalent event in the same bin in the unfolded level. Stability is the fraction of events at the unfolded level that have a corresponding reconstructed event in the same bin. The requirement is to have values of more than 50% for each bin. The calculated values can be seen in Figures 7.40 and 7.41. The choice of binning also results in fairly diagonal response matrices, see Figures 7.51 and 7.52. The unfolding procedure can be performed in many ways. In this particular analysis the simple matrix inversion without regularization method was used. This method avoids biases but has a slight increase in uncertainty.

The measurements produced in this thesis are also compared with three theoretical models. These are POWHEG combined with PYTHIA8, aMC@NLO[FxFx] combined PYTHIA8 and POWHEG combined with HERWIG. The comparison is presented as values of χ^2. The values are calculated using the covariance matrix. Using the χ^2 values also p-values are calculated. The theoretical values used for the comparison have also applied systematic uncertainties mentioned in section 7.8. The uncertainties are also used in the calculation of the χ^2 values as they are used in the calculation of the covariance matrix. For the calculation of χ^2 the following equation is used:

$$x^2 = V_{N_b}^T \cdot C_{N_b}^{-1} \cdot V_{N_b} \tag{7.14}$$

where V_{N_b}, $V_{N_b}^T$ are the vector of differences between the measured and the corresponding theoretical predictions and its transpose while $C_{N_b}^{-1}$ corresponds to the

inverse of the covariance matrix. The index N_b corresponds to the number of bins for each measurement. For the calculation of the covariance matrix the following formula is used:

$$C = C_{stat} + C_{syst} \tag{7.15}$$

where C_{syst} corresponds to the covariance matrix for the systematical uncertainties and C_{stat} is the one for the statistical uncertainties. For the systematic uncertainties the covariance matrix is calculated by

$$C_{ij}^{syst} = \sum_{k,l} \frac{1}{N_k} C_{j,k,l} C_{i,k,l},\ 1 \leq i \leq N,\ 1 \leq j \leq N, \tag{7.16}$$

$C_{i,k,l}$ is the systematic uncertainty in the i-th bin of the l-th variations of the k-th source and N_k is the number of variations for the source k. The sum runs over all the sources of systematic unceratinties and their variations.

7.11 Combination of the different years

The measurement performed in this analysis uses the data collected by the CMS detector in the three running years of the Run-II of the LHC. These data have to be combined as the conditions of the experiment and the configuration and settings of the detector were different. This section discusses the method used in combining the data.

For each year the signal extraction method is performed separately. This is also done for every uncertainty source for every year. At this point for every year there is a nominal, central, value and one value for each uncertainty source for every year. So one quartet of values for every uncertainty variation plus one for the central value. The signal which is the number of events is then combined by adding the four individual years. So from quartets we end up with one value for each uncertainty source and the nominal value. Since the various values are correlated between the data taking periods we make use of the correlation found in Table 7.20 values while adding them.

To compute the differential cross section Equation 7.13, the combined response matrix is needed as well as the fractions f1 and f2. The response matrix of each year is normalized to the luminosity of the corresponding year and the 4 response matrices are added. During the unfolding procedure the combined response matrix is normalized to unity so the luminosity of each year acts as a weight with the years with higher luminosity contributing more. This is also done for each uncertainty source. There is one response matrix for each uncertainty variation

Table 7.20: Correlations for systematic uncertainties between the various years

Source	2016_pre - 2016_post	2016_pre - 2017	2016_pre - 2018	2016_post - 2017	2016_post - 2018	2017 - 2018
pile - up	100%	100%	100%	100%	100%	100%
b-Tagging Unc.	See [58]					
Parton Shower Unc.	100%	100%	100%	100%	100%	100%
CP5 Tune Unc.	100%	100%	100%	100%	100%	100%
hDamp Samples Unc.	100%	100%	100%	100%	100%	100%
JES	See [59]					
Scale Unc.	100%	100%	100%	100%	100%	100%
PDF Unc.	100%	100%	100%	100%	100%	100%
Luminosity	30%	30%	30%	30%	30%	30%

plus one for the central value. For the two fractions, their numerators and denominators are added and from their combination, the total fractions are calculated. The procedure is again repeated for every uncertainty source. The result of the differential cross section is then computed for the central value and for each uncertainty source. The results from the uncertainty sources are then used to compute the total uncertainty of the measurement.

7.12 Results

In this section we present the result for the measurement of the differential cross section measurement. First a table with the χ^2 and p values is presentad and then the plots of the measurement.

Table 7.21: χ^2 Values of the Absolute differential cross sections at Parton level.

Variable	NDOF	χ^2 Powheg Pythia8	χ^2 AMC@NLO Pythia8	χ^2 Powheg Herwig	P-value Powheg Pythia8	P-value AMC@NLO Pythia8	P-value Powheg Herwig
Leading Parton p_T	10	9.5	13	7.2	0.49	0.22	0.71
Sub-leading Parton p_T	7	5.3	8.4	3.2	0.63	0.3	0.87
Parton $y_{t\bar{t}}$	8	9.2	7	9.2	0.33	0.54	0.33
Parton p_T^{tt}	6	4.2	8.3	4.5	0.65	0.21	0.61
Parton $m_{t\bar{t}}$	8	4.6	7	3.6	0.8	0.53	0.89
Leading Parton $\|Y\|$	10	4	4.9	3.2	0.95	0.9	0.98
Sub-leading Parton $\|Y\|$	10	16	14	14	0.1	0.17	0.17

Table 7.22: χ^2 Values of Normalized differential cross sections at Parton level.

Variable	NDOF	χ^2 Powheg Pythia8	χ^2 AMC@NLO Pythia8	χ^2 Powheg Herwig	P-value Powheg Pythia8	P-value AMC@NLO Pythia8	P-value Powheg Herwig
Leading Parton p_T	9	6.7	7.5	6.7	0.67	0.59	0.66
Sub-leading Parton p_T	6	1.2	1.6	1.2	0.98	0.95	0.98
Parton $y_{t\bar{t}}$	7	14	15	15	0.057	0.039	0.038
Parton p_T^{tt}	5	9.4	4.7	12	0.094	0.45	0.033
Parton $m_{t\bar{t}}$	7	2.1	2.5	1.8	0.95	0.93	0.97
Leading Parton $\|Y\|$	9	2	2.3	2.1	0.99	0.99	0.99
Sub-leading Parton $\|Y\|$	9	15	15	14	0.1	0.092	0.11

Table 7.23: χ^2 Values of the Absolute differential cross sections at Particle level.

Variable	NDOF	χ^2 Powheg Pythia8	χ^2 AMC@NLO Pythia8	χ^2 Powheg Herwig	P-value Powheg Pythia8	P-value AMC@NLO Pythia8	P-value Powheg Herwig
Leading Particle p_T	10	16	22	21	0.11	0.014	0.021
Sub-leading Particle p_T	7	6.2	7.6	6.8	0.52	0.37	0.45
Particle $y_{t\bar{t}}$	8	17	7.2	21	0.031	0.52	0.0077
Particle p_T^{tt}	6	5	15	6.6	0.55	0.019	0.36
Particle $m_{t\bar{t}}$	8	5.4	5.9	5.3	0.71	0.66	0.73
Leading Particle $\|Y\|$	10	6.6	6	7.1	0.77	0.81	0.72
Sub-leading Particle $\|Y\|$	10	20	14	21	0.031	0.19	0.021

Table 7.24: χ^2 Values of Normalized differential cross sections at Particle level.

Variable	NDOF	χ^2 Powheg Pythia8	χ^2 AMC@NLO Pythia8	χ^2 Powheg Herwig	P-value Powheg Pythia8	P-value AMC@NLO Pythia8	P-value Powheg Herwig
Leading Particle p_T	9	4.5	4.6	5	0.88	0.87	0.83
Sub-leading Particle p_T	6	1.5	2.3	1.2	0.96	0.89	0.98
Particle $y_{t\bar{t}}$	7	14	13	15	0.054	0.074	0.035
Particle $p_T^{t\bar{t}}$	5	4.3	5.5	6.5	0.5	0.36	0.26
Particle $m_{t\bar{t}}$	7	2.5	3.5	2.8	0.93	0.83	0.91
Leading Particle $\|Y\|$	9	2.1	3.3	3.1	0.99	0.95	0.96
Sub-leading Particle $\|Y\|$	9	15	14	15	0.096	0.11	0.087

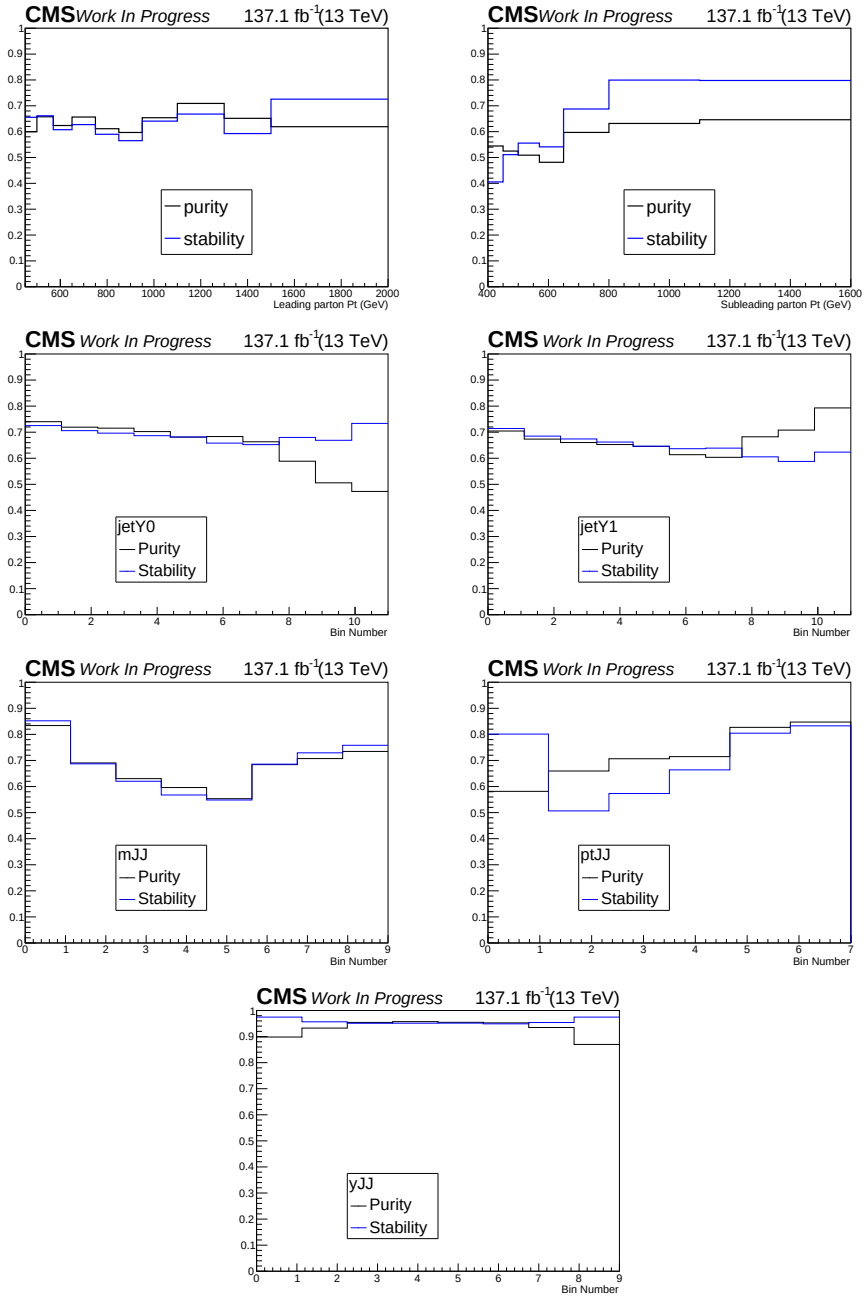

Figure 7.40: Purity and Stability for the parton-level selection as a function of the various observables.

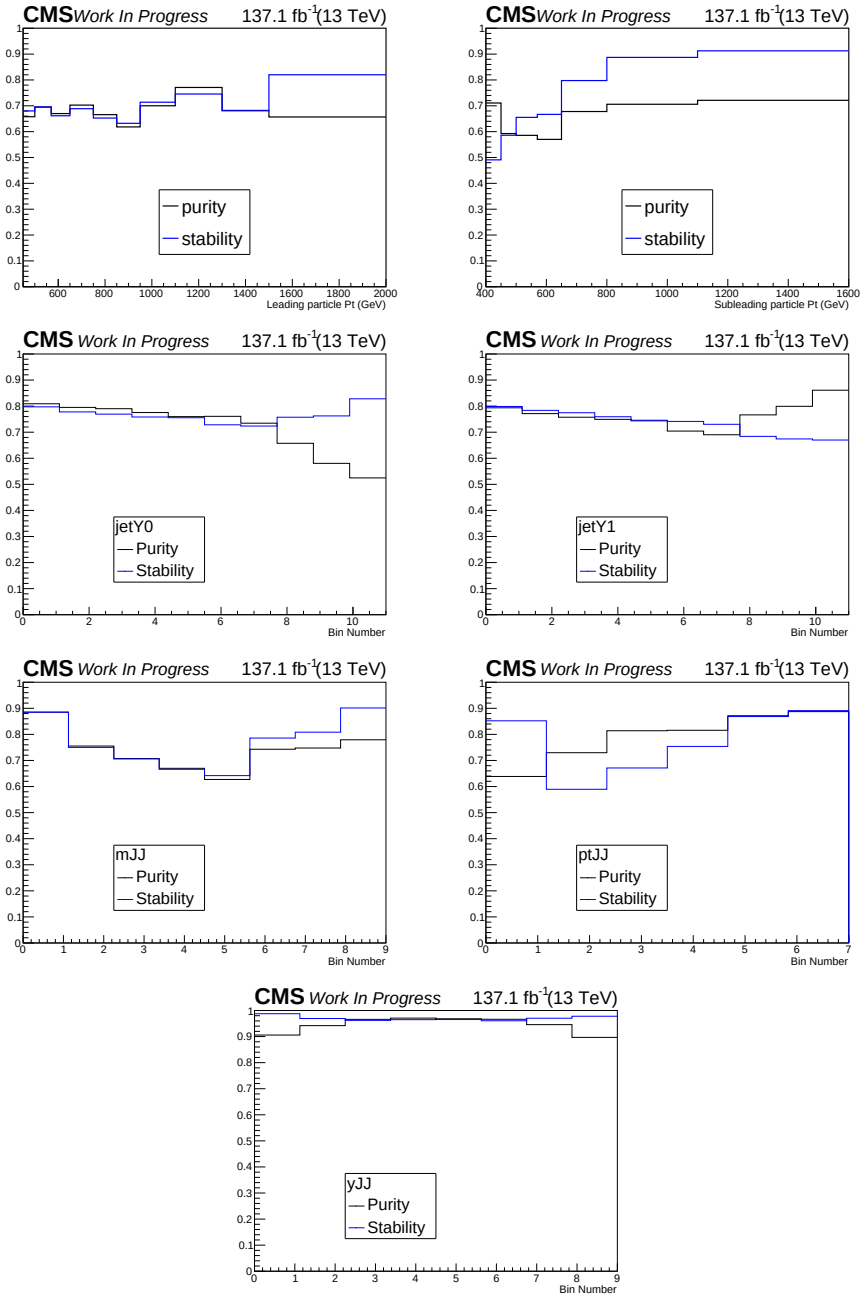

Figure 7.41: Purity and Stability for the particle-level selection as a function of the various observables.

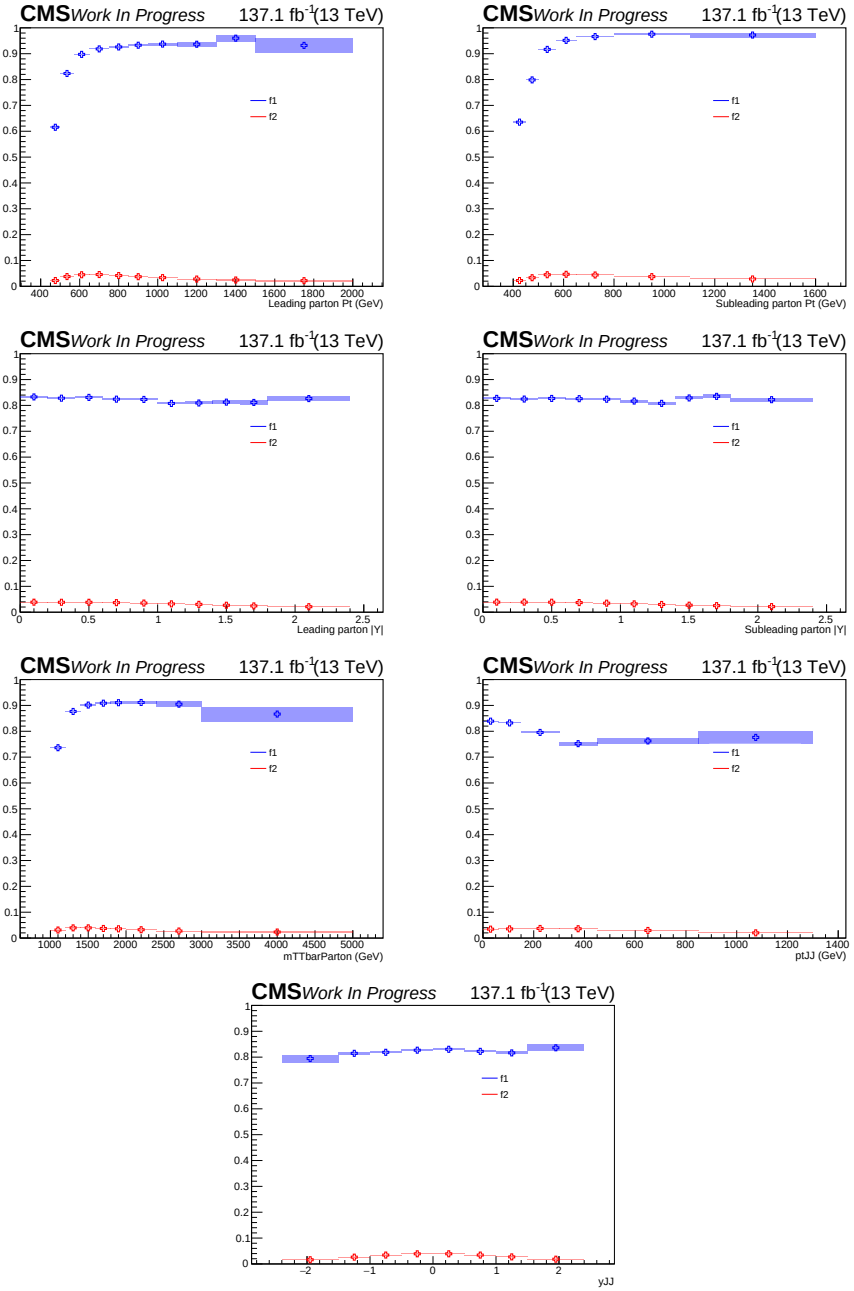

Figure 7.42: Acceptance and efficiency for the parton-level selection as a function of the various observables.

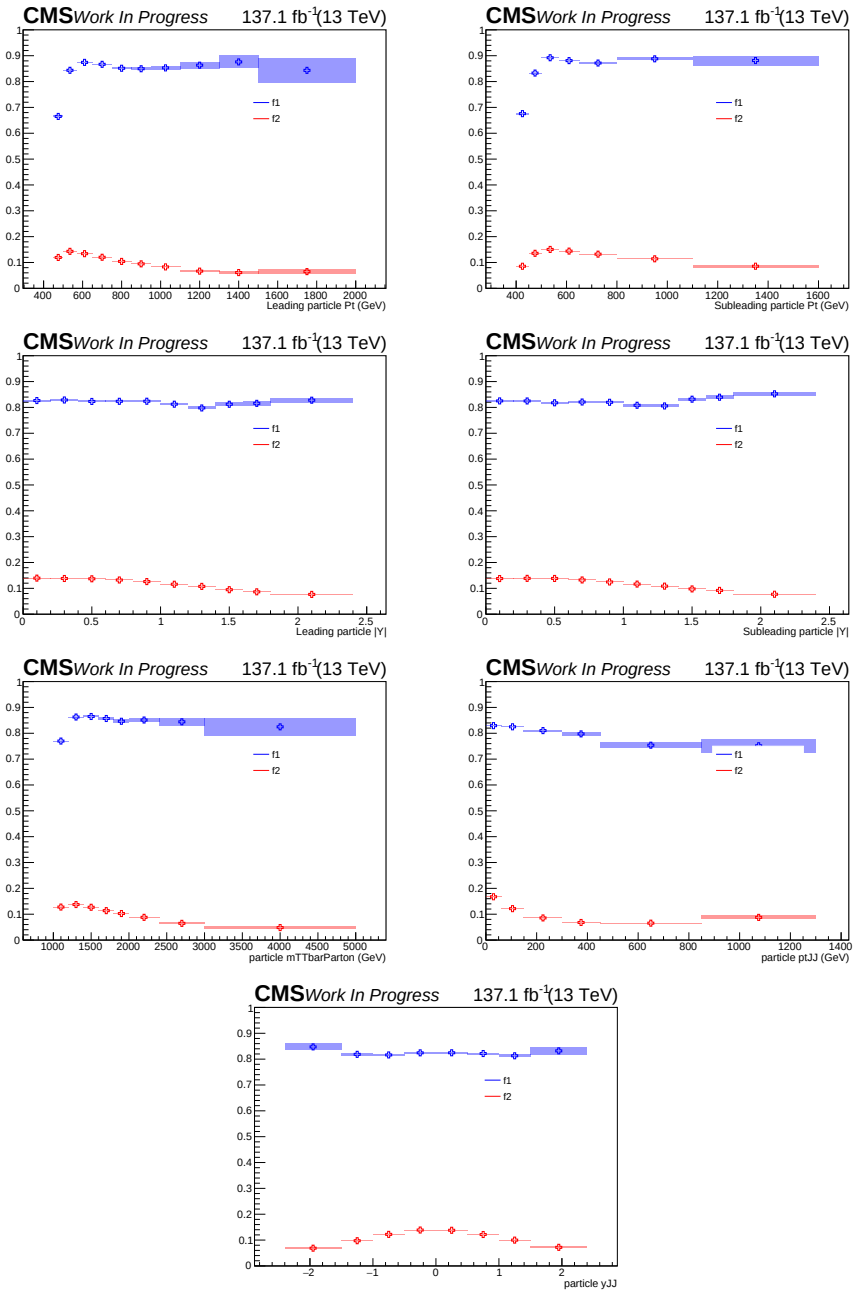

Figure 7.43: Acceptance and efficiency for the particle-level selection as a function of the various observables.

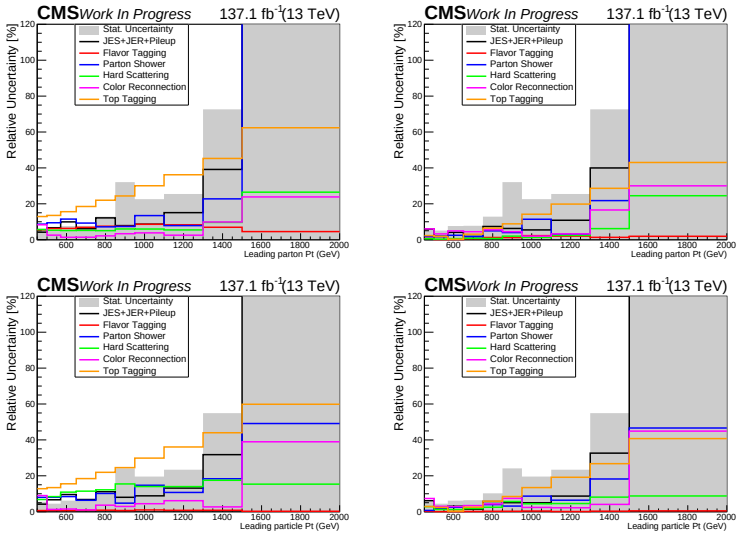

Figure 7.44: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the leading top p_T.

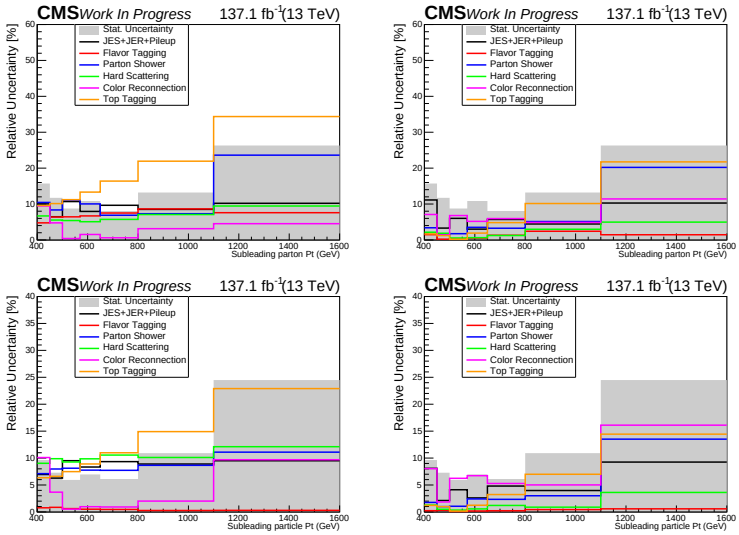

Figure 7.45: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the second leading top p_T.

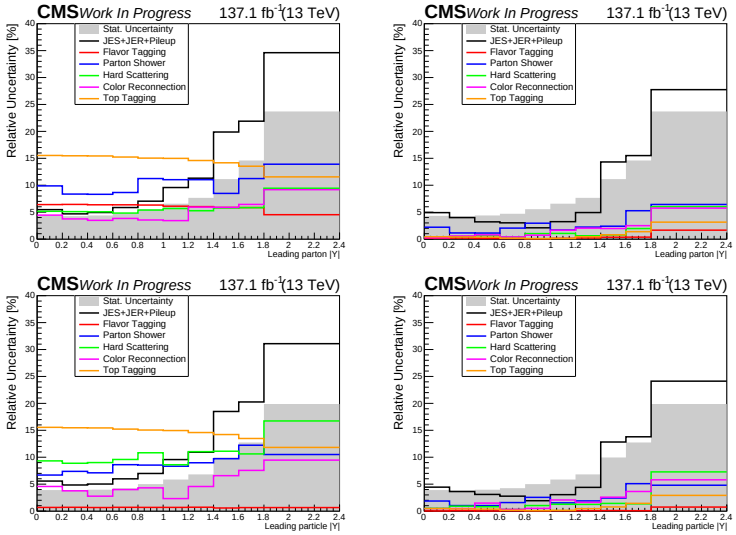

Figure 7.46: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the leading top absolute rapidity y.

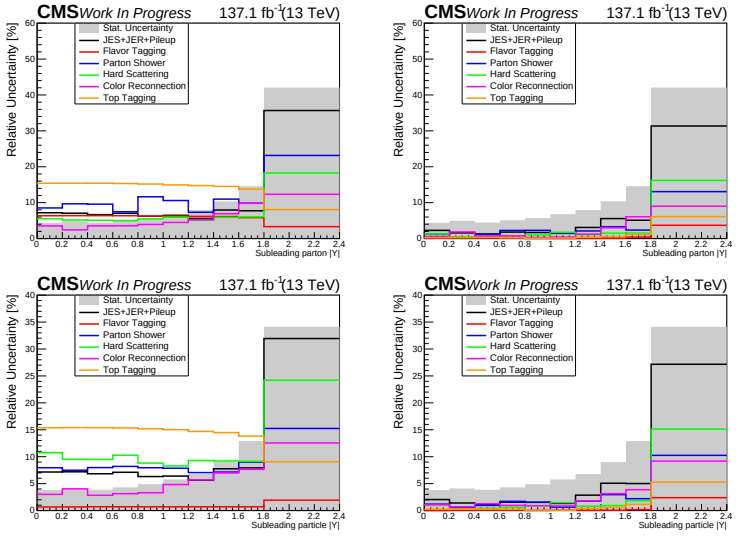

Figure 7.47: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the second leading top absolute rapidity y .

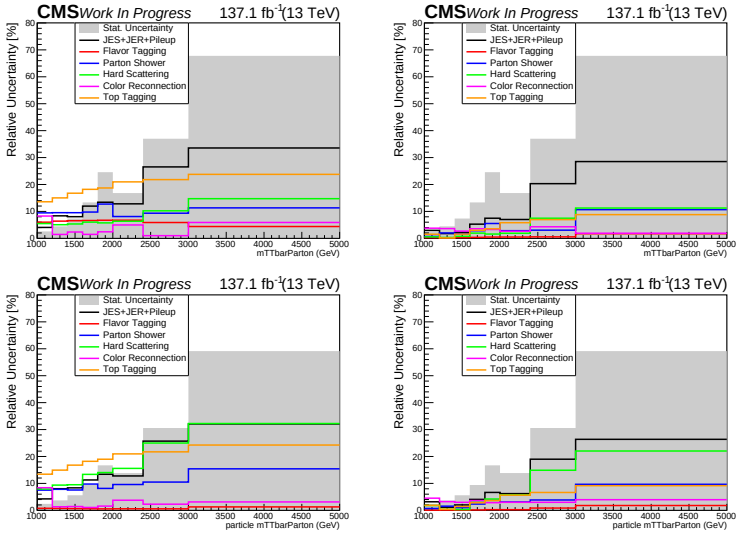

Figure 7.48: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the $m_t\bar{t}$ dijet mass.

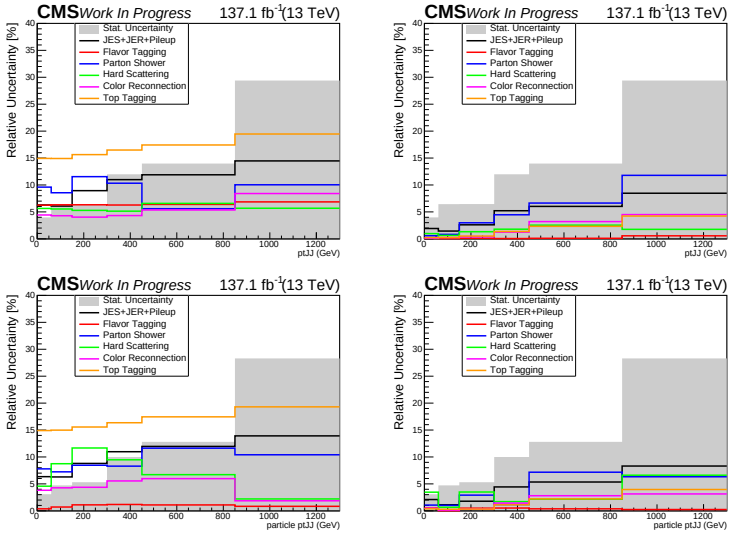

Figure 7.49: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the $p_T^t\bar{t}$ dijet p_T.

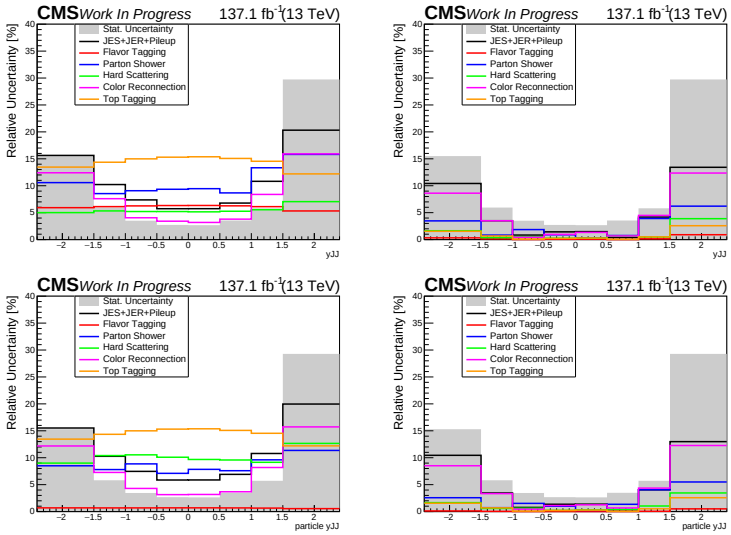

Figure 7.50: Decomposition of uncertainties for the parton- and particle-level measurement (left: absolute, right: normalized) as a function of the $y_t\bar{t}$ dijet rapidity.

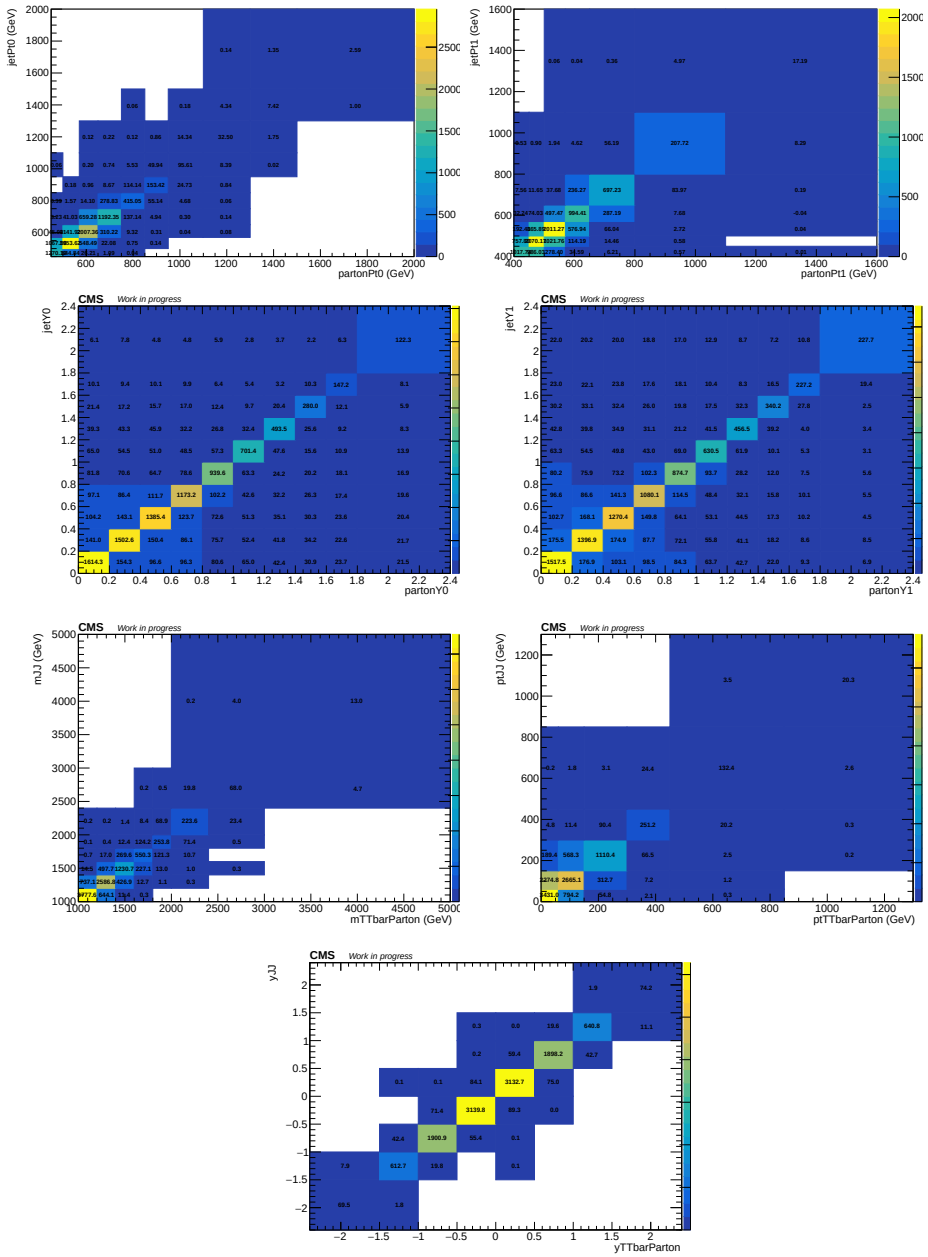

Figure 7.51: Simulated migration matrices at parton level.

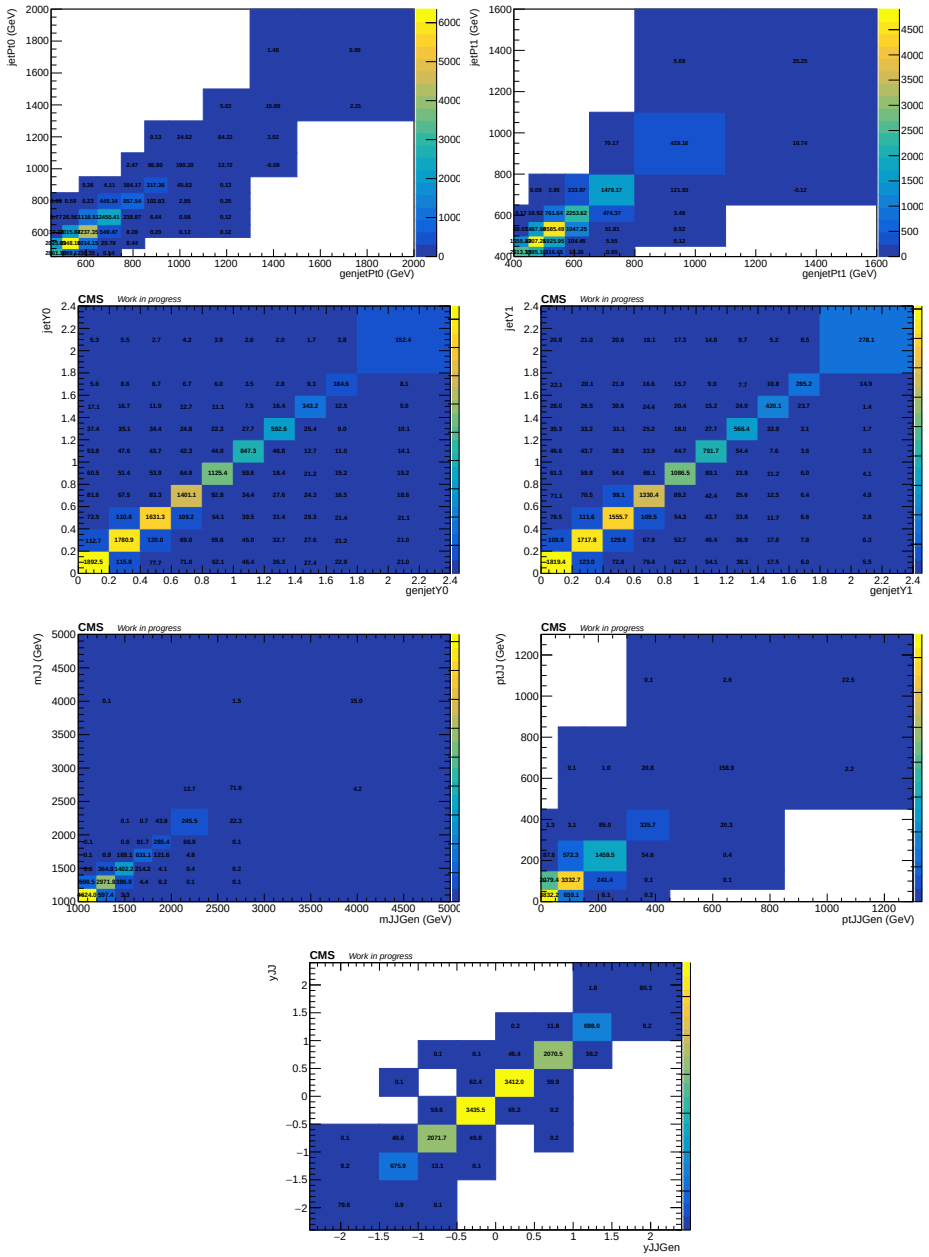

Figure 7.52: Simulated migration matrices at particle level.

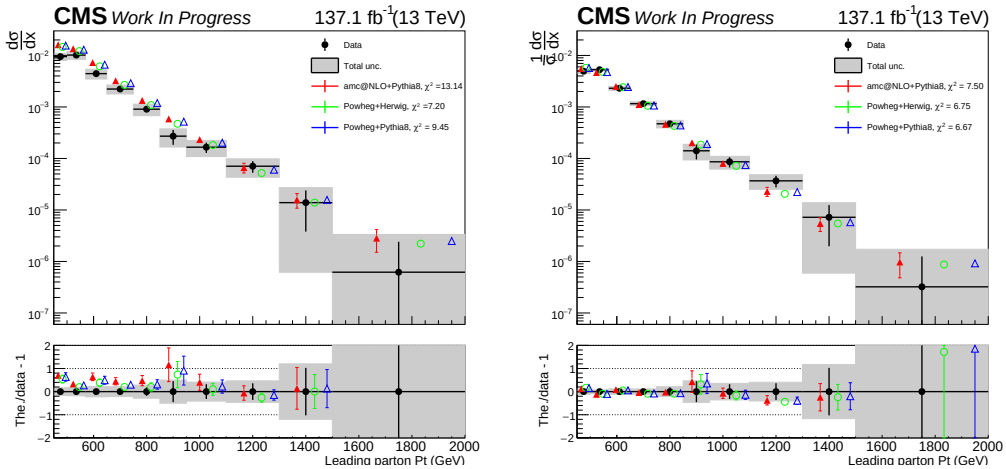

Figure 7.53: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of leading jet p_T. The bottom panel shows the ratio (theory - data)/data.

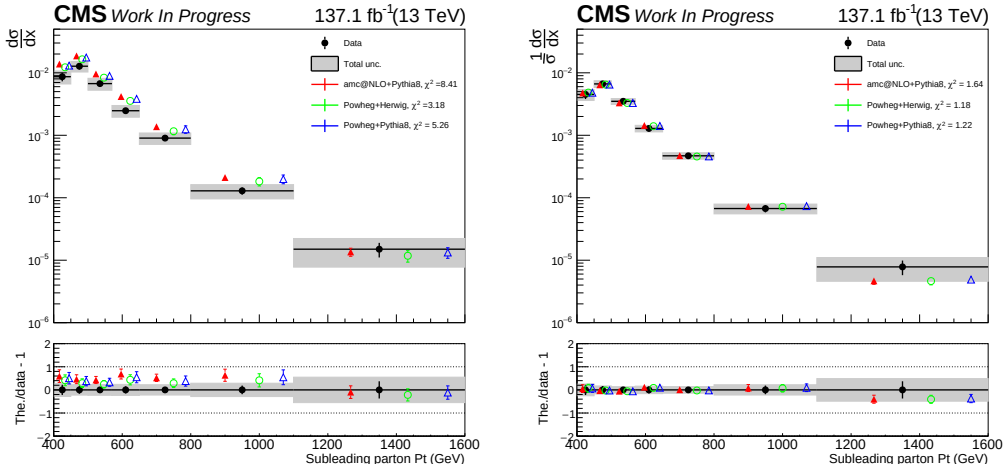

Figure 7.54: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of second leading jet p_T. The bottom panel shows the ratio (theory - data)/data.

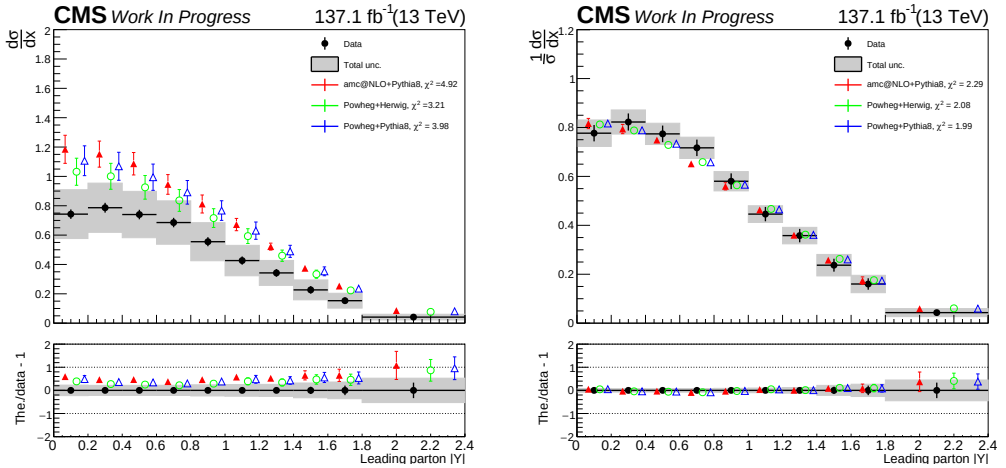

Figure 7.55: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of absolute leading jet rapidity. The bottom panel shows the ratio (theory - data)/data.

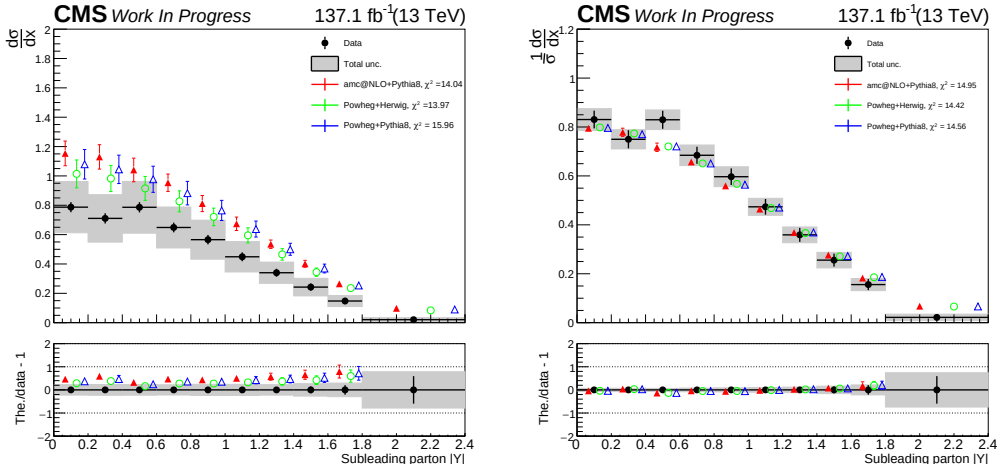

Figure 7.56: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of absolute second leading jet rapidity. The bottom panel shows the ratio (theory - data)/data.

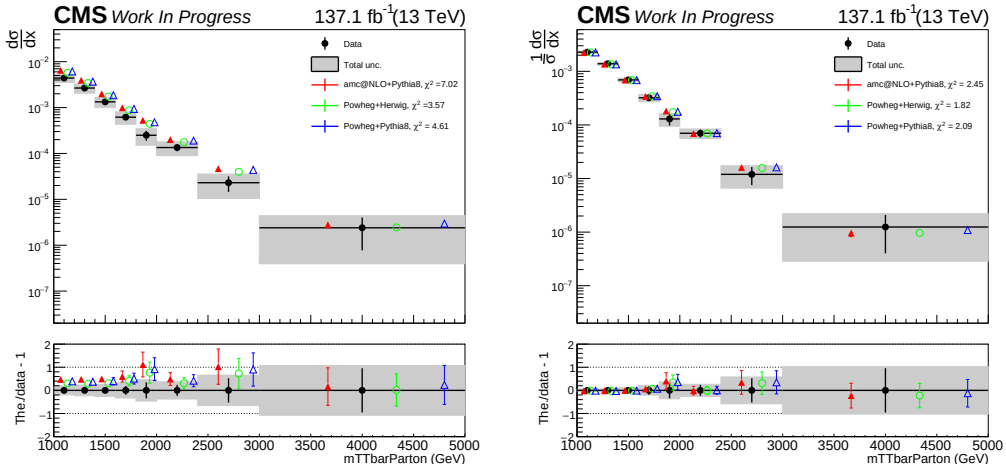

Figure 7.57: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of the dijet $t\bar{t}$ mass. The bottom panel shows the ratio (theory - data)/data.

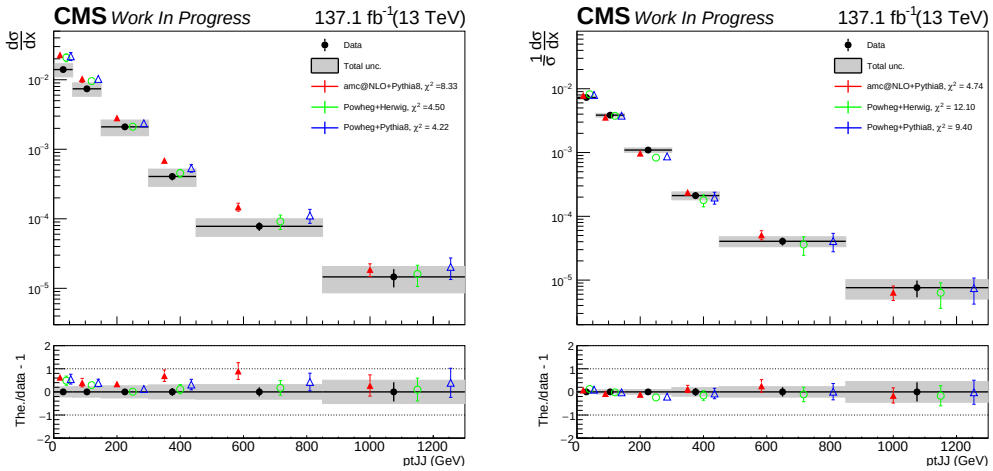

Figure 7.58: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of $p_T^{t\bar{t}}$. The bottom panel shows the ratio (theory - data)/data.

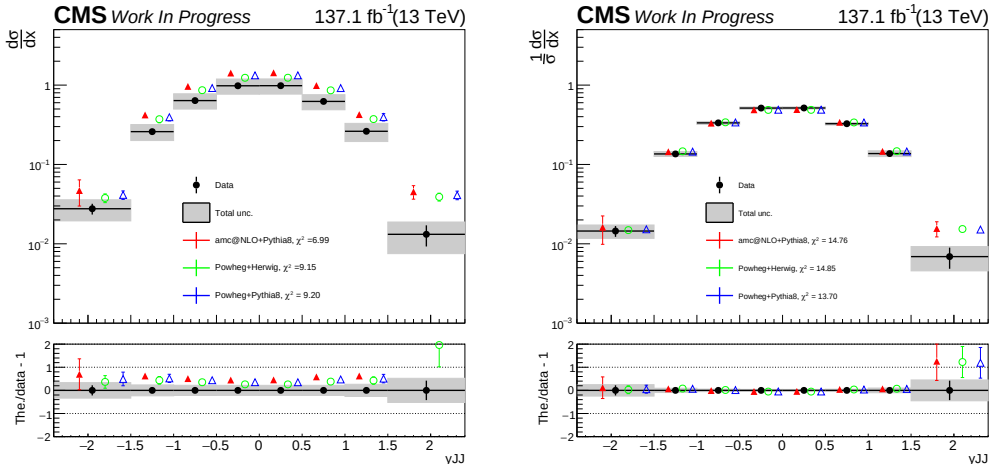

Figure 7.59: Differential cross section unfolded to parton level, absolute (left) and normalized (right), as a function of the dijet rapidity $y^{t\bar{t}}$. The bottom panel shows the ratio (theory - data)/data.

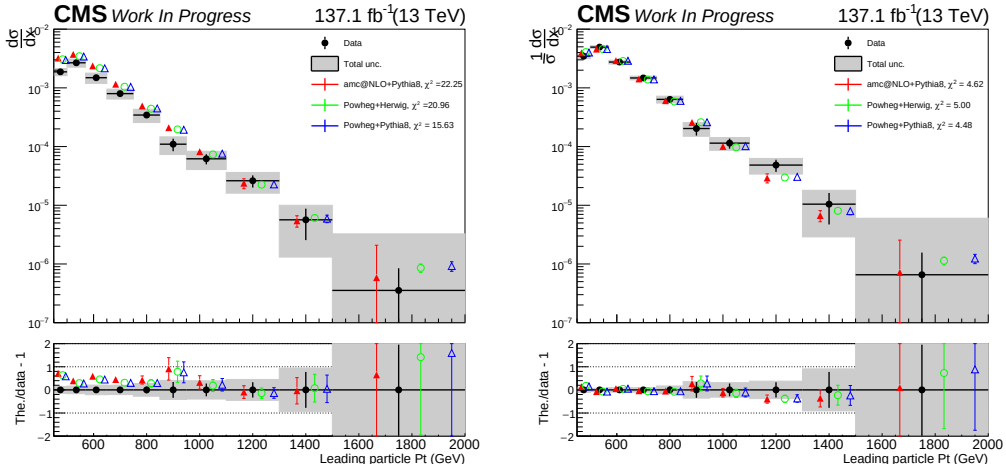

Figure 7.60: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of leading jet p_T. The bottom panel shows the ratio (theory - data)/data.

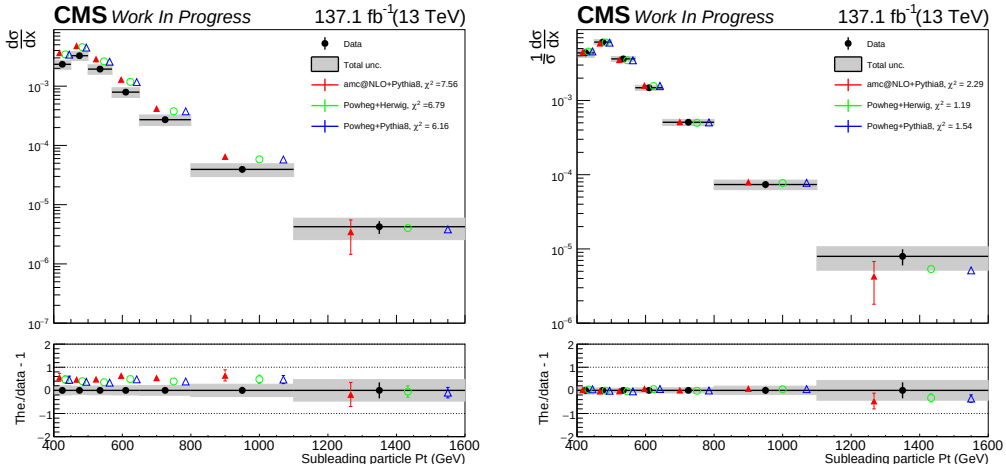

Figure 7.61: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of second leading jet p_T. The bottom panel shows the ratio (theory - data)/data.

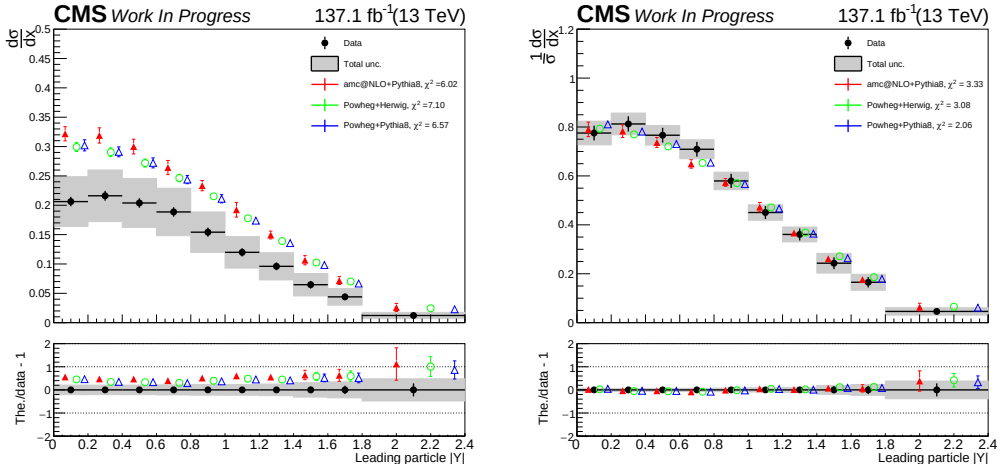

Figure 7.62: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of absolute leading jet rapidity. The bottom panel shows the ratio (theory - data)/data.

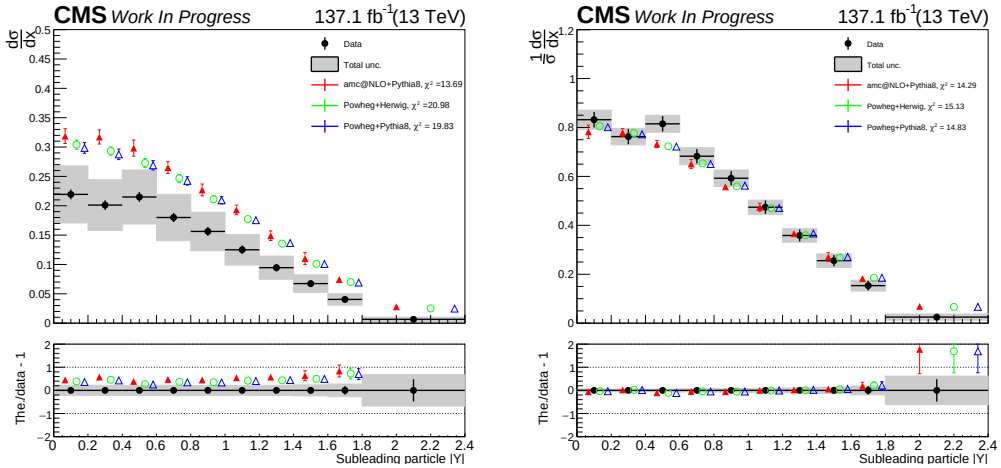

Figure 7.63: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of absolute second leading jet rapidity. The bottom panel shows the ratio (theory - data)/data.

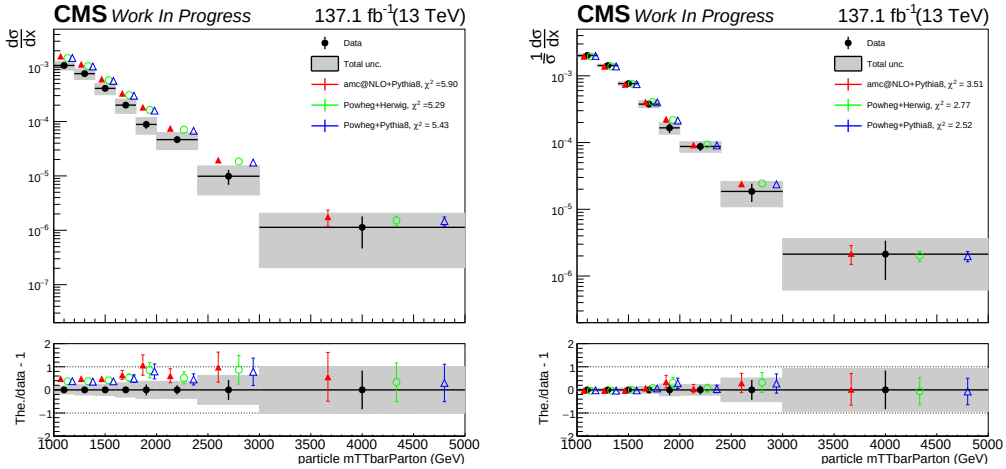

Figure 7.64: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of the dijet $t\bar{t}$ mass. The bottom panel shows the ratio (theory - data)/data.

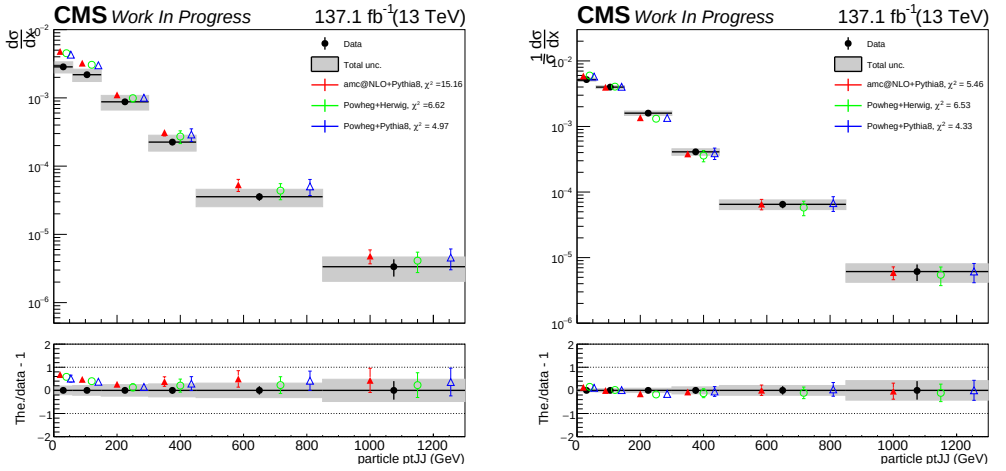

Figure 7.65: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of $p_T^{t\bar{t}}$. The bottom panel shows the ratio (theory - data)/data.

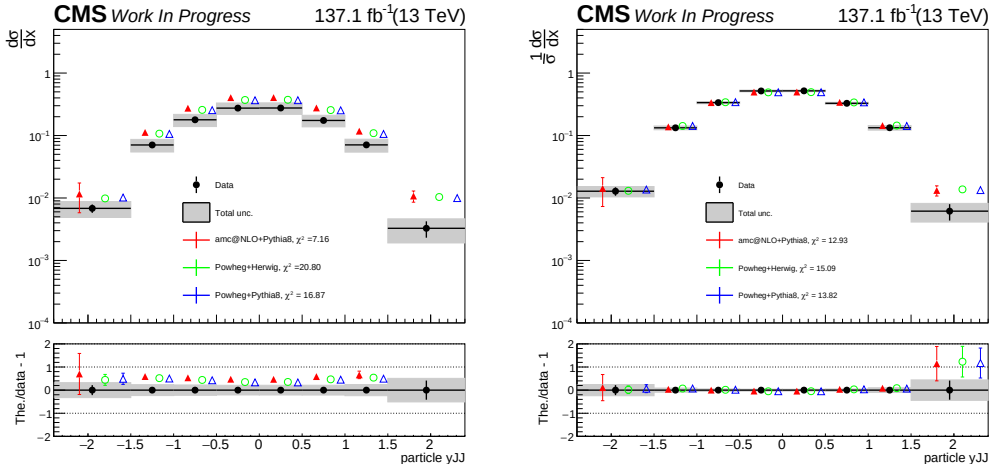

Figure 7.66: Differential cross section unfolded to particle level, absolute (left) and normalized (right), as a function of the dijet rapidity $y^{t\bar{t}}$. The bottom panel shows the ratio (theory - data)/data.

Chapter 8

Conclusions

A measurement of the differential cross section of the production of highly boosted pairs of top anti-top quarks has been presented in this thesis. To conduct this measurement the data collected by the CMS detector during the 2016-2018 period (Run-II) were used. The top quark and the measurement of its properties is of fundamental importance for the field of particle physics. As mentioned in detail in chapter 2 it gives us the ability to study a bare quark. Also it plays an important role in the discovery of new particles as it is a background in many of these processes. Thus the very accurate measurement is of big importance. Precision measurements like this allows us to very accurately test the underlying theory and in particular perturbative qcd which is still not very well understood. Although measurements like this have been conducted before the use of more data improves the accuracy of the measurement and allows for better tuning of the methods used. Also as time passes the instruments used for measuring it, the particle detectors, are also advancing and are better tuned.

The measurements conducted during this thesis have been presented and compared with three theoretical models. It is clear that there is a discrepancy between the measurement and all the theoretical models used. The discrepancy is in the order of 30-40% percent in the measurement of the cross section between the measurement and all models. It looks like the amount of top anti-top pairs predicted by all models is greater that what is actually measured. On the other hand in the normalized cross section there is a very good agreement. This means that the way that these top pairs are produced and distributed across the phase space is in good agreement with the theory in all observable variables. The results are in agreement with previous similar studies both from CMS and ATLAS. This measurement needs to be repeated with even more data in order to improve the significance of

this measurement and better understand the source of the discrepancy between the data and the theoretical predictions.

References

[1] John Dalton. *A new system of chemical Philosophy*. Vol. 2. S. Russell, 1810.

[2] Günther Dissertori. \The Determination of the Strong Coupling Constant". In: *Advanced Series on Directions in High Energy Physics*. WORLD SCIENTIFIC, Aug. 2016, pp. 113{128. DOI: 10.1142/9789814733519_0006. URL: https://doi.org/10.1142/9789814733519_0006.

[3] Giovanni Gallavotti. *Perturbation Theory*. 2007. DOI: 10.48550/ARXIV.0711.2544. URL: https://arxiv.org/abs/0711.2544.

[4] Roger A Hegstrom and Dilip K Kondepudi. \The handedness of the universe". In: *Scientific American* 262.1 (1990), pp. 108{115.

[5] Nicola Cabibbo. \Unitary symmetry and leptonic decays". In: *Physical Review Letters* 10.12 (1963), p. 531.

[6] Makoto Kobayashi and Toshihide Maskawa. \CP-violation in the renormalizable theory of weak interaction". In: *Progress of theoretical physics* 49.2 (1973), pp. 652{657.

[7] R. L. Workman et al. \Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[8] D0 Collaboration. \Observation of the Top Quark". In: *Physical Review Letters* 74.14 (May 1995), pp. 2632{2637. DOI: 10.1103/physrevlett.74.2632. URL: https://doi.org/10.1103/PhysrevLett.74.2632.

[9] D0 Collaboration. \Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab". In: *Physical Review Letters* 74.14 (May 1995), pp. 2626{2631. DOI: 10.1103/physrevlett.74.2626. URL: https://doi.org/10.1103/PhysrevLett.74.2626.

[10] Knut Zoch. *Cross-section measurements of top-quark pair production in association with a hard photon at 13 TeV with the ATLAS detector*. 2020. DOI: 10.48550/ARXIV.2007.14701. URL: https://arxiv.org/abs/2007.14701.

[11] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. \The anti-k_t jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063{063. DOI: 10.1088/1126-6708/2008/04/063. URL: https://doi.org/10.1088/1126-6708/2008/04/063.

[12] *Pileup Removal Algorithms*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1751454.

[13] Daniele Bertolini et al. \Pileup per particle identification". In: *Journal of High Energy Physics* 2014.10 (Oct. 2014). DOI: 10.1007/jhep10(2014)059. URL: https://link.springer.com/article/10.1007/JHEP10(2014)059.

[14] The CMS collaboration. \Determination of jet energy calibration and transverse momentum resolution in CMS". In: *Journal of Instrumentation* 6.11 (Nov. 2011), P11002{P11002. DOI: 10.1088/1748-0221/6/11/p11002. URL: https://doi.org/10.1088/1748-0221/6/11/p11002.

[15] CMS Collaboration. \Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011{P05011. DOI: 10.1088/1748-0221/13/05/p05011. URL: https://doi.org/10.1088/1748-0221/13/05/P05011.

[16] The CMS collaboration. \Identification of b-quark jets with the CMS experiment". In: *Journal of Instrumentation* 8.04 (Apr. 2013), P04013{P04013. DOI: 10.1088/1748-0221/8/04/p04013. URL: https://doi.org/10.10881748-0221/8/04/p04013.

[17] CMS Collaboration. \Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". In: *Journal of Instrumentation* 14.07 (July 2019), P07004{P07004. DOI: 10.1088/1748-0221/14/07/p07004. URL: https://doi.org/10.1088/1748-0221/14/07/p07004.

[18] P Golonka and M Gonzales-Berges. \Integrated access control for PVSS-based SCADA systems at CERN". In: *Proc. ICALEPCS*. Vol. 2011. 2009.

[19] *Lightweight Directory Access Protocol (LDAP)*. https://ldap.com/.

[20] Lorenzo Masetti et al. \Increasing Availability by Implementing Software Redundancy in the CMS Detector Control System". In: (2015), WEPGF013. DOI: 10.18429/JACoW-ICALEPCS2015-WEPGF013. URL: https://cds.cern.ch/record/2213496.

[21] *The automation matrix of the CMS detector control system*. https://cmsonline.cern.ch/webcenter/portal/cmsonline/pages_common/dcs/automation.

[22] M. Aaboud et al. \Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector". In: *JHEP* 11 (2017), p. 191. DOI: 10.1007/JHEP11(2017)191. arXiv: 1708.00727 [hep-ex].

[23] Morad Aaboud et al. \Measurements of top-quark pair differential cross-sections in the $e\mu$ channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector". In: *Eur. Phys. J.* C77.5 (2017), p. 292. DOI: 10.1140/epjc/s10052-017-4821-x. arXiv: 1612.05220 [hep-ex].

[24] M. Aaboud et al. \Measurement of lepton differential distributions and the top quark mass in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Eur. Phys. J.* C77.11 (2017), p. 804. DOI: 10.1140/epjc/s10052-017-5349-9. arXiv: 1709.09407 [hep-ex].

[25] Georges Aad et al. \Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions using the ATLAS detector". In: *Phys. Rev.* D93.3 (2016), p. 032009. DOI: 10.1103/PhysRevD.93.032009. arXiv: 1510.03818 [hep-ex].

[26] Morad Aaboud et al. \Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s}$ = 7 and 8 TeV with ATLAS". In: *Phys. Rev.* D94.9 (2016), p. 092003. DOI: 10.1103/PhysRevD.94.092003. arXiv: 1607.07281 [hep-ex].

[27] Georges Aad et al. \Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS detector". In: *JHEP* 06 (2015), p. 100. DOI: 10.1007/JHEP06(2015)100. arXiv: 1502.05923 [hep-ex].

[28] Georges Aad et al. \Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV". In: *Eur. Phys. J.* C73.1 (2013), p. 2261. DOI: 10.1140/epjc/s10052-012-2261-1. arXiv: 1207.5644 [hep-ex].

[29] A. M. Sirunyan et al. \Measurement of normalized differential $\mathrm{t\bar{t}}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 04 (2018), p. 060. DOI: 10.1007/JHEP04(2018)060. arXiv: 1708.07638 [hep-ex].

[30] *Measurements of normalised multi-differential cross sections for top quark pair production in pp collisions at* $\sqrt{s} = 13$ TeV *and simultaneous determination of the strong coupling strength, top quark pole mass and parton distribution functions*. Tech. rep. CMS-PAS-TOP-18-004. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2647989.

[31] Albert M Sirunyan et al. \Measurements of $\mathrm{t\bar{t}}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons". In: *Submitted to: JHEP* (2018). arXiv: 1811.06625 [hep-ex].

[32] Vardan Khachatryan et al. \Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV". In: *Phys. Rev.* D95.9 (2017), p. 092001. DOI: 10.1103/PhysRevD.95.092001. arXiv: 1610.04191 [hep-ex].

[33] A. M. Sirunyan et al. \Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 06 (2018), p. 002. DOI: 10.1007/JHEP06(2018)002. arXiv: 1803.03991 [hep-ex].

[34] Albert M Sirunyan et al. \Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions". In: *Eur. Phys. J.* C77.7 (2017), p. 459. DOI: 10.1140/epjc/s10052-017-4984-5. arXiv: 1703.01630 [hep-ex].

[35] Vardan Khachatryan et al. \Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J.* C75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].

[36] Vardan Khachatryan et al. \Measurement of the differential cross sections for top quark pair production as a function of kinematic event variables in pp collisions at $\sqrt{s}$=7 and 8 TeV". In: *Phys. Rev.* D94.5 (2016), p. 052006. DOI: 10.1103/PhysRevD.94.052006. arXiv: 1607.00837 [hep-ex].

[37] Serguei Chatrchyan et al. \Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV". In: *Eur. Phys. J.* C73.3 (2013), p. 2339. DOI: 10.1140/epjc/s10052-013-2339-4. arXiv: 1211.2220 [hep-ex].

[38] A. M. Sirunyan et al. \Measurement of differential $\mathrm{t\bar{t}}$ production cross sections using top quarks at large transverse momenta in pp collisions at $\sqrt{s} = 13$ TeV". In: *American Physical Society (APS)* 5 (2021), p. 103. DOI: 10.1103/physrevd.103.052008. arXiv: 2008.07860 [hep-ex].

[39] G Petrucciani, A Rizzi, and C Vuosalo. \Mini-AOD: A New Analysis Data Format for CMS". In: *Journal of Physics: Conference Series* 664.7 (Dec. 2015), p. 072052. DOI: 10.1088/1742-6596/664/7/072052. URL: https://doi.org/10.1088/1742-6596/664/7/072052.

[40] CMS Collaboration. \Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of Instrumentation* 13.06 (June 2018), P06015{P06015. DOI: 10.1088/1748-0221/13/06/p06015. URL: https://doi.org/10.1088/1748-0221/13/06/P06015.

[41] *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256875.

[42] *Recommendations for b-tagging for 2016_preVFP data*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL16preVFP.

[43] *Recommendations for b-tagging for 2016_postVFP data*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL16postVFP.

[44] *Recommendations for b-tagging for 2017 data*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL17.

[45] *Recommendations for b-tagging for 2018 data*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL18.

[46] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2007. DOI: 10.48550/ARXIV.PHYSICS/0703039. URL: https://arxiv.org/abs/physics/0703039.

[47] Jesse Thaler and Ken Van Tilburg. \Identifying boosted objects with N-subjettiness". In: *Journal of High Energy Physics* 2011.3 (Mar. 2011). DOI: 10.1007/jhep03(2011)015. URL: https://link.springer.com/article/10.1007/JHEP03(2011)015.

[48] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. \Energy Correlation Functions for Jet Substructure". In: (2013). DOI: 10.48550/ARXIV.1305.0007. URL: https://arxiv.org/abs/1305.0007.

[49] Wouter Verkerke and David P. Kirkby. \The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[50] *Methods of information collection*. URL: https://ori.hhs.gov/module-4-methods-information-collection-section-2-3.

[51] Albert M Sirunyan et al. \Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *Eur. Phys. J. C* 81.9 (2021), p. 800. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[52] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13\ TeV$. CMS Physics Analysis Summary CMS-PAS-LUM-18-002. 2019. URL: http://cds.cern.ch/record/2676164.

[53] Albert M Sirunyan et al. \Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[54] *Methods to apply b-tagging efficiency scale factors*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.

[55] Vardan Khachatryan et al. \Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[56] Richard D. Ball et al. \Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[57] Richard D. Ball et al. \Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (Oct. 2017). DOI: 10.1140/epjc/s10052-017-5199-5. URL: https://doi.org/10.1140/epjc/s10052-017-5199-5.

[58] https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation.

[59] https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources

Ελληνική Περίληψη

Κεφάλαιο 9

Εισαγωγή

Στόχος της Φυσικής είναι η μελέτη και η κατανόηση της φύσης. Προσπαθεί να βρει τα δομικά της στοιχεία καθώς και τους κανόνες που διέπουν τη συμπεριφορά των στοιχείων αυτών. Η σωματιδιακή φυσική είναι ο τομέας της φυσικής που μελετάει και προσπαθεί να εξηγήσει τη δομή του σύμπαντος. Για το σκοπό αυτό μελετάει σωματίδια, τα οποία είναι στοιχειώδη και αποτελούν τα δομικά στοιχεία της φύσης. Ο απώτερος σκοπός είναι να βρεθεί μια θεωρία, ένα σύνολο κανόνων, που μπορεί να εξηγήσει όλα όσα μπορούν να παρατηρηθούν στο σύμπαν και όλα όσα θα παρατηρηθούν.

Κάθε τομέας της φυσικής έχει δύο βασικούς κλάδους που λειτουργούν μαζί και ενάντια ο ένας στον άλλον. Τη θεωρία και το πείραμα. Δουλεύουν μαζί με την έννοια του ότι έχουν το ίδιο αντικείμενο μελέτης. Από την άλλη εργάζονται ο ένας εναντίον στον άλλο διότι αμφισβητούν ο ένας τον άλλον με σκοπό την δημιουργία νέας γνώσης και περεταίρω κατανόησης στον συγκεκριμένο τομέα. Η πρόοδος μπορεί να συμβεί με δύο τρόπους. Όταν μια νέα θεωρία αναπτύσεται με σκοπό την εξήγηση ενός φυσικού φαινομένου, είναι δουλειά του πειράματος να αμφισβητήσει τη θεωρία αυτή και να προσπαθήσει να τη καταρρίψει. Μία άλλη εναλλακτική είναι κατά τη διάρκεια ενός πειράματος να παρατηρηθέι ένα νέο φαινόμενο ή αποτέλεσμα το οποίο δεν προβλέπετα ή εξηγείται από κάποια θεωρία. Τότε είναι η ώρα της θεωρίας να προσπαθήσει να το εξηγήσει το φαινόμενο αυτό.

Στην διαδικασία της μελέτης του σύμπαντος οι σωματιδιακοί φυσικοί έχουν αναπτύξει το Καθιερωμένο Πρότυπο (Standrad Model, SM) της σωματιδιακής φυσικής. Είναι η πιο πλήρης και ακριβής θεωρία που έχει ποτέ αναπτυχθεί στον τομέα της σωματιδιακής φυσικής. Προβλέπει την ύπαρξη ενός συνόλου σωματιδίων τα οποία ονομάζονται στοιχειώδη. Ο όρος στοιχειώδη προέρχεται από το γεγονός ότι δεν έχουν δομή. Πέρα από την ύπαρξη των σωματιδίων αυτών, το καθιερωμένο πρότυπο προβλέπει τους τρόπους με τους οποίους τα σωματίδια αυτά αλληλεπιδρούν μεταξύ τους.

Χρησιμοποιώντας τα ως βάση η θεωρία προσπαθεί να εξηγήσει τη «φύση» της ύλης καθώς και τη συμπεριφορά της.

Πολλά από τα σωματίδια που προβλέπει η θεωρία αυτή δεν υπάρχουν ελεύθερα στη φύση. Η ύπαρξη τους όμως είναι θεμελιώδης για την εξήγηση του σύμπαντος και την κατάσταση στην οποία βρίσκεται σήμερα. Το γεγονός αυτό, οτι τα σωματίδια αυτά δεν υπάρχουν πλέον ελεύθερα στη φύση είναι κάτι το οποίο προβλέπεται από τη συγκεκριμένη θεωρία. Το καθιερωμένο πρότυπο εκτός από την ύπαρξη των σωματιδίων αυτών, προβλέπει και πώς μπορούν να παραχθούν. Για το λόγο αυτό έχουν αναπτυχθεί επιταχυντές, όπως ο μεγάλος επιταχυντής αδρονίων (LHC). Μέσα από εκτεταμένες δοκιμές και πειράματα, συμπεριλαμβανομένου του LHC το καθιερωμένο πρότυπο έχει ελεγχθεί εκτενώς και έχει καταφέρει να προβλέψει όλα τα αποτελέσματα των διαφόρων πειραμάτων με μεγάλη ακρίβεια. Όλα τα σωματίδια τα οποία προβλέπει, οι ιδιότητές τους καθώς και οι αλληλεπιδράσεις τους έχουν ανιχνευτεί. Το τελευταίο ήταν το μποζόνιο Χιγκς, του οποίου η ανίχνευση ανακοινώθηκε από δύο από τα πειράματα του LHC συμπληρώνοντας έτσι τα σωματίδια τα οποία προβλέπει το καθιερωμένο πρότυπο.

Παρά την επιτυχία του, γνωρίζουμε ότι το καθιερωμέμο πρότυπο δεν είναι τέλειο. Υπάρχουν φαινόμενα στη φύση που δεν περιέχονται στη θεωρία καθώς και διεργασίες τις οποίες δεν προβλέπει. Ένα παράδειγμα είναι η ύπαρξη της βαρύτητας. Γνωρίζουμε ότι η βαρύτητα είναι μια θεμελιώδης δύναμη του σύμπαντος αλλά δεν περιλαμβάνεται στο καθιερωμένο πρότυπο. Το θέμα ασυμμετρία ύλης-αντιύλης. Σύμφωνα με το καθιερωμένο πρότυπο, η ύλη και η αντιύλη παράγονται σε ίσες ποσότητες. Το σύμπαν όμως στη σημερινή του μορφή αποτελείται κυρίως από ύλη. Το γεγονός αυτό, σημαίνει ότι σε κάποια φάση του σύμπαντος η ύλη επικράτησε της αντιύλης. Το ερώτημα είναι πώς και πότε συνέβη αυτό· Γιατί τα νετρίνα έχουν μάζα. Τα νετρίνα είναι σωματίδια που προβλέπονται από το καθιερωμένο πρότυπο και η ύπαρξή τους έχει αποδειχθεί. Σύμφωνα με τη θεωρία δεν έχουν μάζα κάτι το οποίο γνωρίζουμε ότι δεν είναι σωστό καθώς έχει μετρηθλει πειραματικά οτι έχουν μάζα. Γιατί έχουν μάζα και ποια είναι η τιμή της; Τέλος, η σκοτεινή ύλη και η σκοτεινή ενέργεια είναι δύο οντότητες που γνωρίζουμε ότι υπάρχουν στο σύμπαν αλλά δεν προβλέπονται από το καθιερωμένο πρότυπο.

Έχοντας υπόψη τα παραπάνω είναι ξεκάθαρο ότι το καθιερωμλένο πρότυπο σαν θεωρία ισχύει αλλά δεν μπορεί να είναι η τελική θεωρία που εξηγεί τα πάντα. Η μεγάλη επιτυχία του μπορεί να σημαίνει οτι είναι υποσύνολο μιας άλλης μεγαλύτερης θεωρίας με πιο ευρεία ισχύ. Πολλοί επιστήμονες προσπαθούν να δοκιμάσουν νέες θεωρίες που είναι προεκτάσεις του καθιερωμένου προτύπου. Άλλοι προσπαθούν να μετρήσουν πράγματα που προβλέπει με μεγαλύτερη ακρίβεια ώστε να βρούνε ασυνέπειες στη θεωρία και να την αποδείξουν, εν μέρει, λάθος. Η ιστορία συνεχίζεται

με μια ατελείωτη εναλλαγή μεταξύ θεωρίας και πειράματος μέχρι να αναπτυχθεί μια γενική ενοποιημένη θεωρία για τα πάντα. Μια απάντηση στα μεγαλύτερα ερώτηματα, ποιοι και τι είμαστε, από τι είμαστε φτιαγμένοι·

Κεφάλαιο 10

Θεωρία

Οι άνθρωποι ήταν πάντα προβληματισμένοι με τη δομή του κόσμου. Ο Δημόκριτος (430 - 370 π.Χ.) στην αρχαία Ελλάδα ήταν ο πρώτος που διατύπωσε την ατομική θεωρία. Η ιδέα του ήταν ότι αν πάρουμε ένα κομμάτι ύλης και το κόψουμε στη μέση, έπειτα πάρουμε ένα από τα δύο κομμάτια και επαναλάβουμε τη διαδικασία κάποια στιγμή θα φτάσουμε σε ένα κομμάτι που δεν μπορεί να κοπεί περαιτέρω. Ονόμασε εκείνο το τμήμα άτομο που σημαίνει κάτι που δεν μπορεί να κοπεί. Η σύγχρονη έννοια των ατόμων εισήχθη από τον χημικό John Dalton [1]. Πρότεινε ότι τα στοιχεία αποτελούνται από πολύ μικρά σωματίδια που ονόμασε άτομα. Πρότεινε επίσης ότι τα άτομα του ίδιου στοιχείου είναι πανομοιότυπα. Γνωρίζουμε πλέον ότι αυτό δεν ισχύει και ότι τα άτομα μπορούν να «κοπούν» σε μικρότερα κομμάτια. Ωστόσο, η ίδια φιλοσοφία οδήγησε την αναζήτηση των δομικών στοιχείων του σύμπαντος. Αυτό είχε ως αποτέλεσμα την ανάπτυξη και τη διαμόρφωση του SM της σωματιδιακής φυσικής. Αυτό το κεφάλαιο θα είναι μια θεωρητική εισαγωγή στο SM και στο top κουάρκ.

10.1 Το καθιερωμένο πρότυπο της σωματιδιακής φυσικής

Το SM διαμορφώθηκε στα τέλη της δεκαετίας του ′60 και στις αρχές της δεκαετίας του ′70. Είναι μια Κβαντική Θεωρία Πεδίου που προβλέπει την ύπαρξη ενός συνόλου στοιχειωδών σωματιδίων, σωματιδίων δηλαδή χωρίς υποδομή, Εικόνα 2.1. Χρησιμοποιώντας αυτά τα σωματίδια περιγράφει το σχηματισμό όλης της ορατής ύλης καθώς και τη συμπεριφορά της.

Σύμφωνα με το SM τα σωματίδια χωρίζονται σε δύο μεγάλες κατηγορίες, τα μποζόνια και τα φερμιόνια. Η διαφοροποίηση γίνεται με βάση το σπιν τους που είναι εσωτερική ιδιότητα όλων των σωματιδίων. Τα φερμιόνια έχουν ημι-ακέραιο σπιν, δηλαδή 1/2 ενώ τα μποζόνια έχουν ακέραια τιμή για το σπιν.

10.1.1 Τα Φερμιόνια

Τα φερμιόνια είναι τα σωματίδια που συνθέτουν την ύλη. Χωρίζονται σε δύο κατηγορίες, τα λεπτόνια και τα κουάρκ. Υπάρχουν συνολικά 6 κουάρκ (6 γεύσεις), τα up (u), down (d), charm (c), strange (s), top (t) και bottom (b). Υπάρχουν επίσης 6 λεπτόνια. Τρία είναι ηλεκτρικά φορτισμένα, το ηλεκτρόνιο (e), το μιόνιο (μ) και το ταυ (τ). Για κάθε φορτισμένο λεπτόνιο υπάρχει και ένα το οποίο είναι ουδέτερα φορτισμένο και ονομάζεται νετρίνο (ν_e, ν_μ, ν_τ). Χρησιμοποιώντας ως μέτρο το ηλεκτρικό φορτίο του ηλεκτρονίου, λέμε ότι το ηλεκτρόνιο έχει φορτίο -1. Τα υπόλοιπα φορτισμένα λεπτόνια έχουν επίσης φορτίο -1, ενώ τα νετρίνα έχουν 0. Τα κουάρκ από την άλλη πλευρά μπορούν να έχουν φορτίο 2/3 για τα u, c , t και -1/3 για τα d, s και b.

Η μαθηματική διατύπωση φερμιονίων στο SM (εξίσωση Dirac) προβλέπει την ύπαρξη σωματιδίων με αρνητική ενέργεια. Αυτό έχει αποδειχθεί και πειραματικά. Έτσι για κάθε φερμιόνιο υπάρχει και ένα αντισωματίδιο, δηλαδή ένα σωματίδιο που έχει ακριβώς τις ίδιες ιδιότητες με το σωματίδιο αλλά αντίθετα φορτία πρόσημου.

Τα φερμιόνια φέρουν επίσης έναν άλλο τύπο φορτίου που ονομάζεται χρώμα. Μπορεί να έχει τρεις πιθανές τιμές κόκκινο πράσινο και μπλε. Τα λεπτόνια είναι άχρωμα ενώ τα κουάρκ μπορεί να έχουν οποιοδήποτε από τα πιθανά χρώματα. Αυτή η ιδιότητα σχετίζεται με την ικανότητα ενός σωματιδίου να είναι ελεύθερο ή σε δεσμευμένη κατάσταση. Στη φύση μόνο άχρωμες καταστάσεις μπορούν να βρεθούν. Αυτό σημαίνει ότι τα λεπτόνια μπορούν να βρεθούν ελεύθερα στη φύση αφού δεν έχουν χρώμα. Από την άλλη, τα κουάρκ βρίσκονται μόνο ομαδοποιημένα σχηματίζοντας πιο πολύπλοκα σωματίδια μέσω μιας διαδικασίας που ονομάζεται αδρονοποίηση, hadronization στα αγγλικά. Αυτό συμβαίνει ώστε να σχηματίζουν άχρωμες καταστάσεις. Οι συνδυασμοί κουάρκ είναι συνήθως κουάρκ αντι-κουάρκ, που ονομάζονται μεσόνια, ή συνδυασμοί περισσότερων κουάρκ που ονομάζονται βαρυόνια. Μέχρι πρόσφατα είχαν βρεθεί μόνο βαρυόνια με 3 κουάρκ. Το πείραμα LHCb στο CERN ανακοίνωσε την ανακάλυψη κάποιων εξωτικών σωματιδίων που έχουν 4 ή και 5 κουάρκ και ονομάζονται τέτρα και πεντα-κουάρκ αντίστοιχα. Τα φερμιόνια χωρίζονται επίσης σε 3 άλλες κατηγορίες που ονομάζονται γενιές. Κάθε γενιά περιέχει 2 κουάρκ, ένα φορτισμένο λεπτόνιο και ένα ουδέτερο λεπτόνιο, ένα νετρίνο. Τα σωματίδια του ίδιου τύπου έχουν αυξημένη μάζα σε κάθε κατηγορία. Με την πρώτη γενιά να περιέχει τα ελαφρύτερα και την τρίτη τα πιο βαριά.

10.1.2 Τα Μποζόνια

Τα μποζόνια είναι οι φορείς δυνάμης σύμφωνα με το καθιερωμένο πρότυπο. Υπάρχουν 4 θεμελιώδεις δυνάμεις στο σύμπαν όπου κάθε μία από αυτές εκφράζεται μέσω ενός σωματιδίου, ενός μποζονίου. Αυτό σημαίνει ότι για να συμβεί κάθε αντί-

δραση ένα από αυτά τα σωματίδια πρέπει να εμπλέκεται. Οι 4 δυνάμεις που αναφέρθηκαν είναι η ισχυρή, η ασθενής, η ηλεκτρομαγνητική και η βαρύτητα. Ο φορέας για την ισχυρή δύναμη είναι το γκλουόνιο (g) το οποίο δεν έχει μάζα. Για την ασθενή δύναμη υπάρχουν τρεις φορείς, οτ W^+, οτ W^- και το Z^0. Τα δύο πρώτα είναι θετικά και αρνητικά/φορτισμένα αντίστοιχα ενώ το τρίτο είναι ουδέτερο. Και οι τρεις φορείς της ασθενής αλληλεπίδρασης, έχουν μάζα. Το φωτόνιο (γ) είναι ο φορέας δύναμης της ηλεκτρομαγνητικής δύναμης και δεν έχει μάζα. Τέλος, η βαρύτητα εκφράζεται μέσω του βαρυτονίου, το οποίο είναι ένα σωματίδιο που έχει προβλεφθεί θεωρητικά αλλά δεν έχει ανιχνευθεί ποτέ. Στο πλαίσιο της σωματιδιακής φυσικής η επίδραση της βαρύτητας είναι τόσο μικρή που μπορεί να αγνοηθεί.

10.2 Το top κουάρκ

Το top κουάρκ είναι το βαρύτερο από τα κουάρκ στο SM Figure 2.3 με μάζα $172,69 \pm 0,30$ [7]. Ανακαλύφθηκε το 1995 από τα πειράματα CDF και D0 στον επιταχυντή Tevatron στο Fermilab [8], [9]. Ήταν το τελευταίο από τα κουάρκ που ανακαλύφθηκαν και από τότε δεν έχουν εντοπιστεί άλλα κουάρκ. Με τόσο υψηλή μάζα έχει πολύ μικρό χρόνο ζωής, περίπου 5×10^{-25} s. Αυτή η τιμή είναι μικρότερη από τον χρόνο αδρονοποίησης 10^{-24} s που σημαίνει ότι διασπάται πριν μπορέσει να σχηματίσει αδρόνια. Το γεγονός ότι διασπάται πριν σχηματίσει ένα αδρόνιο επιτρέπει τη μελέτη των ιδιοτήτων ενός ελεύθερου κουάρκ ακόμη και έμμεσα, μελετώντας τις ιδιότητες της διάσπασης. Επιπλέον λόγω της υψηλής μάζας αλληλεπιδρά με το μποζόνιο Higgs που σημαίνει ότι παίζει σημαντικό ρόλο στον μηχανισμό Higgs. Τέλος, το top κουάρκ αποτελεί υπόβαθρο για πολλές από τις διαδικασίες που προβλέπονται από επεκτάσεις του καθιερεομένου προτύπου. Αυτό σημαίνει ότι παράγεται μαζί με άλλα νέα υποθετικά σωματίδια σε αυτές τις διαδικασίες. Η μέτρηση των ιδιότήτων του με πολύ μεγάλη ακρίβεια θα επιτρέψει στους επιστήμονες να αφαιρέσουν τις συνεισφορές από top κουάρκ στις μετρήσεις τους ώστε αυτό που θα μετρήσουν να είναι μόνο κάποιο νέο σωματίδιο εφόσον αυτό υπάρχει.

10.2.1 Παραγωγή του top κουάρκ

Ένα top κουάρκ μπορεί να παραχθεί μέσω της ηλεκτρασθενούς αλληλεπίδρασης μαζί με ένα κουάρκ b και ένα μποζόνιο W. Τα διαγράμματα Feynman της πρώτης τάξης (leading order, LO) των πιθανών αλληλεπιδράσεων φαίνονται στην Εικόνα 2.4. Μια άλλη δυνατότητα είναι η παραγωγή ενός ζεύγους top (t), anti-top ($\bar{t}$) ($t\bar{t}$) μέσω της ισχυρής αλληλεπίδρασης. Αυτό μπορεί να συμβεί είτε με την αλληλεπίδραση δύο γκλουονίων που ονομάζεται σύντηξη γκλουονίων είτε μέσω της εξαΰλωσης δύο

κουάρκ. Τα διαγράμματα Feynman πρώτης τάξης των δύο διεργασιών φαίνονται στις εικόνες 2.5 και 2.6 αντίστοιχα.

Στους επιταχυντές σωματιδίων, η παραγωγή top κουάρκ γίνεται κυρίως με τη διαδικασία παραγωγής ζεύγους $t\bar{t}$ μέσω της ισχυρής αλληλεπίδρασης. Στον επιταχυντή Tevatron όπου ανακαλύφθηκε το top κουάρκ, τα πειράματα διεξήχθησαν με τη σύγκρουση πρωτονίων και αντιπρωτονίων. Η παραγωγή γινόταν κυρίως με τη διαδικασία εξαΰλωσης ενός κουάρκ και ενός αντι-κουάρκ. Στον LHC λόγω υψηλότερων ενεργειών η παραγωγή γίνεται κυρίως μέσω της διαδικασίας σύντηξης γκλουονίων σε ποσοστό περίπου 90%.

10.2.2 Διάσπαση του top κουάρκ

Το top κουάρκ όπως αναφέρθηκε προηγουμένως διασπάται πολύ γρήγορα λόγω της μεγάλης του μάζας. Η διάσπαση συμβαίνει μέσω της ηλεκτροασθενούς αλληλεπίδρασης καθώς είναι η μόνη που μπορεί να αλλάξει τη γεύση ενός κουάρκ. Με την εκπομπή ενός μποζονίου W, το top κουάρκ μπορεί να διασπαστεί σε κουάρκ d, s ή b. Ωστόσο, η διάσπαση γίνεται σχεδόν αποκλειστικά σε β κουάρκ με πιθανότητα 99,8%.

Το μποζόνιο W που παράγεται, θα διασπαστεί επίσης πριν από την ανίχνευση του. Υπάρχουν δύο κατηγορίες με βάση τη φύση των προϊόντων διάσπασης, λεπτονικά ή αδρονικά. Με τον λεπτονικό τρόπο, το W διασπάται σε ένα λεπτόνιο και το αντίστοιχο νετρίνο του ενώ με τον αδρονικό τρόπο διασπάται σε ένα ζεύγος κουάρκ και αντικουάρκ. Όσον αφορά τα λεπτόνια, καθένα από τα τρία είναι δυνατό να παραχθεί με ίδια πιθανότητα. Για τα αδρόνια, όλα τα κουάρκ μπορούν να παραχθούν εκτός από το top λόγω της διατήρησης της ενέργιας. Στην περιπτωση αυτή τα ποσοστά δεν είναι ίδια με το b κουάρκ να παράγεται κυρίως. Τα ποσοστά για κάθε πιθανή διάσπαση μπορούν να φανούν στην Εικόνα 2.8.

Λαμβάνοντας υπόψη ότι στο ζεύγος $t\bar{t}$ έχουμε δύο top τα οποία διαστπώνται, θα υπάρχουν δύο μποζόνια W. Σε αυτή την περίπτωση μπορούμε να έχουμε οποιοδήποτε από τα παρακάτω σενάρια:

- **Πλήρως αδρονικό κανάλι**[1]. Σε αυτό το κανάλι και τα δύο μποζόνια W διασπώνται σε ένα ζευγάρι κουάρκ. Αυτό το κανάλι είναι το πιο πιθανό από τα τρία με πιθανότητα 45,4%

- **Ημί-λεπτονικό κανάλι**. Σε αυτή την περίπτωση, ένα από τα μποζόνια διασπάται αδρονικά ενώ το άλλο διασπάται λεπτονικά. Αυτή η διαδικασία έχει επίσης μεγάλη πιθανότητα, 44,1%

[1] Στην ορολογία της σωματιδιακής φυσικής ένα κανάλι είναι μια διαδικασία

- **Διλεπτονικό κανάλι**. Η τελευταία περίπτωση είναι όταν και τα δύο μποζόνια W διασπώνται λεπτονικά. Αυτή η διαδικασία συμβαίνει σε ποσοστό περίπου 10,7%

Στην εικόνα 2.9 φαίνονται όλα τα πιθανά σενάρια διάσπασης μαζί με τις αντίστοιχες πιθανότητες.

10.3 Ενεργός διατομή

Η ενέργός διατομή (σ) μιας διεργασίας είναι ένα μέτρο της πιθανότητας να πραγματοποιηθεί η διαδικασία αυτή. Υπολογίζεται θεωρητικά χρησιμοποιώντας τα διαγράμματα Feynman, όπως αυτά στην εικόνα 2.5 για παράδειγμα. Ο υπολογισμός γίνεται αθροίζοντας όλα τα πιθανά διαγράμματα. Τα διαγράμματα στην εικόνα 2.5 ονομάζονται διαγράμματα πρώτης τάξης (LO) αφού περιέχουν μόνο τα σωματίδια αρχικής και τελικής κατάστασης συν τους φορείς αλληλεπίδρασης. Γνωρίζουμε ότι τα σωματίδια μπορούν να ακτινολβολήσουν ένα άλλο σωματίδιο. Έχοντας αυτό κατά νου μπορεί κανείς να δημιουργήσει πιο περίπλοκα διαγράμματα όπως αυτά στην εικόνα 2.10. Αυτά τα διαγράμματα ονομάζονται διαγράμματα δεύτερης τάξης (Next to Leading Order, NLO). Η τάξη του διαγράμματος καθορίζεται ανάλογα με τον αριθμό των επιπλέον σωματιδίων που εκπέμπονται ή αλλιώς του αριθμού των κορυφών. Μια κορυφή είναι πρακτικά ένα σημείο όπου συμβαίνει μια αλληλεπίδραση.

Αυξάνοντας την πολυπλοκότητα των διαγραμμάτων μπορούμε να δούμε ότι πρακτικά μπορεί κανείς να δημιουργήσει και να υπολογίσει έναν άπειρο αριθμό διαγραμμάτων. Αθροίζοντας όλα τα διαγράμματα, η ενέργός διατομή μπορεί να υπολογιστεί σύμφωνα με την εξίσωση 2.2 Ωστόσο, δεδομένου ότι το α_S είναι μικρότερο από 1, μπορούμε να δούμε ότι καθώς τα διαγράμματα γίνονται όλο και πιο πολύπλοκα, συμβάλλουν όλο και λιγότερο στη συνολική ενεργό διατομή. Εκτός από αυτό, τα διαγράμματα γίνονται όλο και πιο πολύπλοκα στον υπολογισμό καθώς αυξάνεται η τάξη τους. Συνήθως γίνεται υπολογισμός μέχρι τον τρίτο όρο, Next to Next to Leading Order (NNLO).

Τα πειράματα επιταχυντών από την άλλη παρέχουν έναν ευκολότερο τρόπο υπολογισμού της ενεργού διατομής. Μπορεί να αποδειχθεί ότι είναι ανάλογη με τον αριθμό των μετρούμενων γεγονότων και υπολογίζεται με την εξίσωση 2.3. Όπου N είναι ο συνολικός αριθμός των μετρούμενων γεγονότων για τη συγκεκριμένη διαδικασία και L_{int} είναι η ολοκληρωμένη φωτεινότητα. Η φωτεινότητα (L) είναι ένα μέτρο του αριθμού των συγκρούσεων που μπορούν να προκληθούν σε έναν ανιχνευτή ανά cm^2 ανά δευτερόλεπτο. Η ολοκληρωμένη φωτεινότητα είναι η φωτεινότητα που «συλλέγεται» σε όλο το πείραμα. Είναι πρακτικά το ολοκλήρωμα της φωτεινότητας

για τη χρονική διάρκεια του πειράματος. Μπορούμε να δούμε με αυτόν τον τρόπο να δούμε ότι τα πειράματα επιταχυντών είναι ένας καλός τρόπος για τον υπολογισμό της ενεργού διατομής και επομένως την επαλήθευση του SM.

Κεφάλαιο 11

Το πείραμα Compact Muon Solenoid

Στο κεφάλαιο αυτό θα γίνει μια περιγραφή του πειράματος Compact Muon Solenoid (CMS). Θα αναλυθούν οι αρχές λειτουργίας του καθώς και τα μέρη από τα οποία αποτελείται.

11.1 Ο Μεγάλος Επιταχυντής Αδρονίων

Ο Μεγάλος Επιταχυντής Αδρονίων (LHC) είναι ο μεγαλύτερος επιταχυντής σωματιδίων που κατασκευάστηκε ποτέ. Βρίσκεται στις εγκαταστάσεις του Ευρωπαϊκού Συμβουλίου για την Πυρηνική Έρευνα (Conseil Européen pour la Recherche Nucléaire CERN), στη Γενεύη της Ελβετίας και άρχισε να λειτουργεί το 2008. Είναι ένας κυκλικός επιταχυντής με περιφέρεια 27 χιλιομέτρων (km) που βρίσκεται 100 μέτρα κάτω από το έδαφος στα σύνορα της Γαλλίας με την Ελβετία μεταξύ της Γενεύης και της οροσειράς Jura. Η Εικόνα 3.1 δείχνει μια αεροφωτογραφία της περιοχής όπου βρίσκεται ο LHC. Λειτουργεί επιταχύνοντας δύο δέσμες πρωτονίων, μία δεξιόστροφα και μία αριστερόστροφα. Οι δέσμες συγκρούονται σε 4 προκαθορισμένα σημεία αλληλεπίδρασης. Ανιχνευτές σωματιδίων έχουν κατασκευαστεί και τοποθετηθεί σε κάθε ένα από τα σημεία αλληλεπίδρασης οι οποίοι καταγράφουν τα προϊόντα των συγκρούσεων.

Οι δύο επιταχυνόμενες δέσμες πρωτονίων συγκρούονται μεταξύ τους προκειμένου να μελετηθούν διάφορες φυσικές διεργασίες. Οι συγκρούσεις λαμβάνουν χώρα σε 4 προκαθορισμένα σημεία της περιμέτρου του επιταχυντή όπου είναι τοποθετημένοι ανιχνευτές, Εικόνα 3.3. Οι 4 ανιχνευτές σωματιδίων LHC είναι:

- **A T**oroidal **L**HC **A**pparatu**S** (**ATLAS**)
- **C**ompact **M**uon **S**olenoid (**CMS**)

- **LHC b**eauty (**LHCb**)
- **A L**arge **I**on **C**ollider **E**xperiment (**ALICE**)

Οι δύο πρώτοι ανιχνευτές ATLAS και CMS, είναι ανιχνευτές γενικής χρήσης, ενώ το LHCb και το ALICE είναι κατασκευασμένοι για τη μελέτη συγκεκριμένων φυσικών διεργασιών. Το LHCb έχει σχεδιαστεί για να μελετά το b (beauty) κουάρκ και τη μέτρηση της παραβίασης της συμμετρίας ομοτιμλίας-φορτίου (Charge Parity (CP)). Με λίγα λόγια αυτό ονομάζεται παραβίαση CP. Το ALICE είναι κατασκευασμένο για να μελετά τις συγκρούσεις δεσμών βαρέων ιόντων και επίσης αυτές μεταξύ βαρέων ιόντων και πρωτονίων.

11.2 Ο ανιχνευτής Compact Muon Solenoid

Ο ανιχνευτής CMS, είναι ένα πείραμα φυσικής υψηλής ενέργειας γενικού ενδιαφέροντος. Έχει κυλινδρικό σχήμα με διάφορα ομοαξονικά στρώματα γύρω από τον άξονα της δέσμης σα βαρέλι. Για να ανιχνευθούν όλα τα παραγόμενα σωματίδια πρέπει να ερμητικά κλειστός. Για να το επιτευχθεί αυτό, χρησιμοποιούνται 2 κατακόρυφοι δίσκοι, κάθετοι στον άξονα της δέσμης που ονομάζονται endcaps. Ο ανιχνευτής έχει μήκος 28,7 μέτρα (m) και ακτίνα 15 μέτρα και βάρος 14.000 τόνους. Είναι σχετικά μικρός για την ποσότητα της ύλης που περιέχει, δικαιολογώντας τη λέξη συμπαγής, Compact.

Ένα από τα πιο ενδιαφέροντα σωματίδια που στοχεύει να ανιχνεύσει το CMS είναι τα μιόνια. Τα μιόνια είναι ενδείξεις ότι μια «ενδιαφέρουσα» φυσική διαδικασία έχει συμβεί κατά τη διάρκεια των συγκρούσεων. Λαμβάνοντας υπόψη τη σημασία των μιονίων, το CMS έχει σχεδιαστεί και κατασκευαστεί με τέτοιο τρόπο ώστε να ανιχνεύει μιόνια με πολύ υψηλή ακρίβεια. Από εκεί προέρχεται η λέξη muon. Τέλος, το CMS χρησιμοποιεί έναν μεγάλο σωληνοειδή μαγνήτη ικανό να παράγει μαγνητικό πεδίο έως και 4 T παράλληλο στον άξονα της δέσμης. Ο μαγνήτης χρησιμοποιείται για να κάμψει τις τροχιές των σωματιδίων κάθετα στον άξονα της δέσμης. Η παρουσία του συγκεκριμένου μαγνήτη δικαιολογεί τη λέξη σωληνοειδές στο όνομα του ανιχνευτή.

Το CMS έχει σχεδιαστεί με τρόπο που αποτελείται από πολλαπλά στρώματα, το καθένα από τα οποία παίζει διαφορετικό ρόλο στην ανίχνευση σωματιδίων. Κάθε ένα από αυτά τα στρώματα είναι ένα ξεχωριστό σύστημα ανιχνευτή που ονομάζεται υποανιχνευτής. Το κάθε στρώμα παρέχει ένα διαφορετικό κομμάτι από τη συνολική πληροφορία που απαιτείται για την ανίχνευση ενός σωματιδίου. Όλα τα κομμάτια συλλέγονται και συνδυάζονται προκειμένου να ταυτοποιηθούν τα σωματίδια που παράγονται κατά τη σύγκρουση πρωτονίων. Το CMS έχει συνολικά 4 συστήματα

υποανιχνευτών, Εικόνα 3.4. Ο ανιχνευτής τροχιών, ο οποίος είναι υπεύθυνος για την ανίχνευση της τροχιάς των σωματιδίων, το Ηλεκτρομαγνητικό θερμιδόμετρο (ECAL), το αδρονικό θερμιδόμετρο (HCAL) και τους ανιχνευτές μιονίων. Εκτός από αυτά τα συστήματα, το CMS περιέχει επίσης και τον σωληνοειδή μαγνήτη που αναφέρθηκε και προηρουμένως. Ο μαγνήτης δεν χρησιμοποιείται για την ανίχνευση των σωματιδίων αλλά είναι ένα ζωτικό μέρος στη λειτουργία του ανιχνευτή. Η διάταξη που χρησιμοποιείται για τα συστήματα υποανίχνευσης είναι η ακόλουθη. Το Tracker, ECAL και HCAL βρίσκονται μέσα στον μαγνήτη ενώ τα συστήματα μιονίων είναι έξω από αυτόν. Κάθε ένας από τους υποανιχνευτές έχει ένα μέρος στον κύριο όγκο του ανιχνευτή το βαρέλι και ένα μέρος στα καπάκια.

11.2.1 Ο ανιχνευτής τροχιών

Ο ανιχνευτής τροχιών είναι το πρώτο σύστημα ανίχνευσης που διαθέτει το CMS και είναι εκείνο που βρίσκεται πιο κοντά στο σημείο αλληλεπίδρασης των δύο δεσμών πρωτονίων. Χρησιμοποιείται για την ανακατασκευή των τροχιών των σωματιδίων που παράγονται από τις συγκρούσεις και επίσης για τη μέτρηση της ορμής τους. Μπορεί να μετρήσει με πραγματικά υψηλή ακρίβεια την ορμή των μιονίων, των ηλεκτρονίων και των φορτισμένων αδρονίων. Μπορεί επίσης να ανιχνεύσει κορυφές (vertex), οι οποίες είναι τα σημεία στα οποία τα σωματίδια που παράγονται από τις συγκρούσεις διασπώνται σε άλλα δευτερεύοντα σωματίδια. Λόγω της υψηλής φωτεινότητας των δεσμών και της θέσης του, ο ανιχνευτής τροχιών πρέπει να έχει πολύ καλή χωρική διακριτική ικανότητα αλλά και να είναι ανθεκτικός στην ακτινοβολία. Πρέπει επίσης να έχει πολύ καλό χρόνο απόκρισης για να καταγράφει όσο το δυνατόν περισσότερα γεγονότα. Το υλικό που χρησιμοποιείται για τον συγκεκριμένο ανιχνευτή είναι πυρίτιο.

11.2.2 Το Ηλεκτρομαγνητικό θερμιδόμετρο (ECAL)

Το δεύτερο σύστημα ανίχνευσης του CMS είναι το ηλεκτρομαγνητικό θερμιδόμετρο (Εικόνα 3.7). Χρησιμοποιείται για τη μέτρηση της ενέργειας των ηλεκτρονίων και των φωτονίων που παράγονται κατά τη διάρκεια των συγκρούσεων. Το Ecal σχεδιάστηκε έτσι ώστε να μπορεί να ανιχνεύει με πραγματικά υψηλή ακρίβεια την ενέργεια των φωτονίων. Αυτό συμβαίνει επειδή ένας από τους πιο συνηθισμένους τρόπους διάσπασης ενός μποζονίου Higgs είναι μέσω των αλληλεπιδράσεων $H \rightarrow \gamma\gamma$ σε δύο πολύ ενεργητικά φωτόνια. Επιπλέον, το θερμιδόμετρο σχεδιάστηκε για να έχει πολύ καλή διακριτική ικανότητα συγκεκριμένα στην ανίχνευση δύο φωτονίων. Αυτό, διότι δύο φωτόνια μπορούν επίσης να παραχθούν από τη διάσπαση ενός ουδέτερου πιονίου,

$\pi^0 \to \gamma\gamma$. Συνεπώς η διάκριση αυτών των δύο διεργασιών είναι ζωτικής σημασίας για την ανίχνευση του μποζονίου Higgs. Το Ecal κατασκευάστηκε με χρήση κρυστάλλων μολύβδου-βολφραμίου ($PbWO_4$). Η μέτρηση της ενέργειας των σωματιδίων βασίζεται σε μια αρχή που ονομάζεται σπινθηρισμός. Όταν ένα σωματίδιο που αλληλεπιδρά ηλεκτρομαγνητικά διασχίζει τον όγκο του ανιχνευτή αλληλεπιδρά με το υλικό του ανιχνευτή και παράγει φως. Το παραγόμενο φως μπορεί να συλλεχθεί και μετρώντας την ποσότητα του παραγόμενου φωτός μπορεί να μετρηθεί η ενέργεια του σωματιδίου.

11.2.3 Το αδρονικό θερμιδόμετρο (HCAL)

Το τρίτο στρώμα του ανιχνευτή CMS είναι το αδρονικό θερμιδόμετρο, Hadronic CALorimenter (HCAL). Αυτό το υποσύστημα είναι υπεύθυνο για τη μέτρηση της ενέργειας των αδρονίων και των προϊόντων διάσπασής τους. Η ανίχνευση σωματιδίων βασίζεται στους αδρονικές πίδακες. Είναι το αποτέλεσμα της αλληλεπίδρασης των αδρονίων, που διασχίζουν το επίπεδο του ανιχνευτή, με τον πυρήνα των ατόμων του υλικού του ανιχνευτή. Το «ενεργό» υλικό που χρησιμοποιείται στο HCAL είναι φθορίζον πλαστικό. Το HCAL είναι ένα θερμιδόμετρο δειγματοληψίας, sampling calorimeter, που σημαίνει ότι αποτελείται από στρώματα ανιχνευτικού υλικού και απορροφητή. Ως απορροφητής χρησιμοποιήθηκε ο ορείχαλκος μαζί με τον χάλυβα. Όταν ένα σωματίδιο περνά μέσα από το ενεργό υλικό του ανιχνευτή παράγεται φως. Αυτό το φως συλλέγεται από ίνες και στη συνέχεια μεταφέρεται σε ειδικούς ανιχνευτές που ονομάζονται Hybrid Photodiodes (HPD) όπου ενισχύεται και μετατρέπεται σε ηλεκτρικό σήμα. Από εκεί μεταφέρεται τελικά στη μονάδα συλλογής δεδομένων, Εικόνα 3.8

Εκτός από την άμεση ανίχνευση σωματιδίων, το ECAL και το HCAL μετρούν επίσης σωματίδια έμμεσα. Αυτό είναι με τη μορφή έλλειψης ορμής. Μερικά από τα σωματίδια που παρουσιάζουν μεγάλο ενδιαφέρον στη σύγχρονη φυσική, όπως τα νετρίνα, αλληλεπιδρούν πολύ σπάνια ή μερικές φορές και ποτέ με την ύλη. Αυτό καθιστά πρακτικά αδύνατη την ανίχνευσή τους. Χρησιμοποιώντας την αρχή διατήρησης της ορμής και υπολογίζοντας το έλλειμα ορμής, μπορεί κανείς να μετρήσει έμμεσα την ύπαρξη ενός σωματιδίου που παρήχθη και δεν ανιχνεύτηκε. Για το λόγο αυτό το ECAL και το HCAL πρέπει να είναι ερμητικά κλειστά γιατί δεν πρέπει να υπάρχει πιθανότητα να διαφύγουν τα συνηθισμένα σωματίδια από τον ανιχνευτή και να ερμηνευθούν εσφαλμένα ως ελλείπουσα ορμή.

11.2.4 Οι ανιχνευτές μιονίων

Το τελευταίο στρώμα του CMS είναι το σύστημα ανίχνευσης μιονίων. Όπως αναφέρθηκε παραπάνω, η ανίχνευση μιονίων είναι πολύ σημαντική για τα πειράματα του LHC. Τα μιόνια είναι πολύ διεισδυτικά και μπορούν να διασχίσουν μεγάλη απόσταση μέσα σε ένα υλικό χωρίς να αλληλεπιδρούν και επομένως χωρίς να ανιχνευτούν. Για την ανίχνευση τους, χρησιμοποιείται ένα ειδικό σύστημα ανιχνευτών. Το σύστημα μιονίων παίζει επίσης βασικό ρόλο στο σύστημα σκανδαλισμού του CMS. Για το σύστημα μιονίων, χρησιμοποιούνται τρεις διαφορετικοί τύποι ανιχνευτών:

- Drift Tubes (DT).
- Cathode Strip Chambers (CSC).
- Resistive Plate Chambers (RPC).

Καθένας από αυτούς τους τύπους χρησιμοποιείται για διαφορετικό σκοπό. Τα δύο πρώτα έχουν πολύ καλή χωρική διακριτική ικανότητα ενώ τα RPC έχουν πολύ καλή χρονική διακριτική ικανότητα.

Οι **drift tubes (DT)** τοποθετούνται στο κύριο σώμα του ανιχνευτή CMS, το barrel. Στο σημείο αυτό, η ροή των σωματιδίων είναι σχετικά μικρή και το μαγνητικό πεδίο ομοιογενές. Οι DT αποτελούνται από 5 κυλίνδρους, ομοαξονικούς με το κέντρο της δέσμης και καλύπτουν όλο το barrel. Κάθε δίσκος έχει 4 σταθμούς διαφορετικής ακτίνας. Κάθε σωλήνας έχει διάμετρο 4 εκατοστά και ένα σύρμα στο κέντρο. Είναι γεμάτος με ένα μείγμα αερίου αργού και διοξειδίου του άνθρακα. Όταν ένα μιόνιο διέρχεται από το αέριο ιονίζει τα μόρια του. Τα ηλεκτρόνια που δημιουργούνται από τον ιονισμό κινούνται προς το σύρμα λόγω του ηλεκτρικού πεδίου στο κέντρο του σωλήνα. Ανιχνεύοντας το σημείο στο οποίο τα ηλεκτρόνια χτυπούν στο σύρμα και τον χρόνο που χρειάστηκε για να φτάσουν τα ηλεκτρόνια στο σύρμα, μπορεί να υπολογιστεί η διαδρομή των μιονίων.

Οι **Cathode Strip Chambers (CSC)**, τοποθετούνται στα άκρα κάθετα προς την κατεύθυνση της δέσμης όπου η ροή είναι μεγάλη και το μαγνητικό πεδίο μη ομοιόμορφο. Οι CSC έχουν πολύ μικρό χρόνο απόκρισης και έτσι μπορούν να χρησιμοποιηθούν σε περιβάλλον με τόσο υψηλή ακτινοβολία. Η αρχή λειτουργίας τους βασίζεται σε ένα πλέγμα καλωδίων. Τα καλώδια έχουν θετικά φορτισμένα σύρματα που λειτουργούν ως άνοδοι και αρνητικά φορτισμένα σύρματα που λειτουργούν ως κάθοδοι. Το πλέγμα βρίσκεται μέσα σε ένα μείγμα αερίων. Όταν ένα σωματίδιο φορτίου περνά μέσα από το αέριο, τα μόρια του αερίου ιονίζονται. Τα ηλεκτρόνια που παράγονται κατά τον ιονισμό κινούνται προς τις καθόδους παράγοντας έναν καταιγισμό. Το σήμα που παράγεται από τον καταιγισμό παρέχει πληροφορίες σχετικά με την τροχιά του σωματιδίου που τον παρήγαγε.

Οι **Resistive Plate Chambers (RPC)**, τοποθετούνται τόσο στο barrel όσο και στα καπάκια του CMS. Λειτουργούν συμπληρωματικά με τα άλλα δύο συστήματα μιονίων και παρέχουν πληροφορίες σκανδαλισμού. Αποτελούνται από δύο πλάκες, μία θετικά φορτισμένη που λειτουργέι ως άνοδος και μία αρνητικά ως που λειτουργέι κάθοδος. Ανάμεσα στις δύο πλάκες, υπάρχει ένα κενό (2 mm περίπου) γεμάτο με αέριο. Όταν ένα μιόνιο διέρχεται από το αέριο, το ιονίζει. Λόγω των φορτισμένων πλακών υπάρχει ένα ηλεκτρικό πεδίο μέσα στο αέριο που κάνει τα ηλεκτρόνια να επιταχύνουν προκαλώντας καταιγισμό. Τα σωματίδια στον καταιγισμό οδηγούνται στην άνοδο όπου συλλέγονται παρέχοντας πληροφορίες για το μιόνιο που πέρασε. Το ηλεκτρικό πεδίο μέσα στις πλάκες προκαλεί την παραγωγή του σήματος ταχύτερα από τον ρυθμό σύγκρουσης του LHC. Λόγω αυτού, ο ανιχνευτής έχει πολύ καλή χρονική απόκριση. Αυτό μαζί με την καλή χωρική ανάλυση επιτρέπει τη χρήση του RPC για σκανδαλισμό.

11.2.5 Το σύστημα σκανδαλισμού

Κατά τη λειτουργία του LHC, οι ανιχνευτές καταγράφουν δεδομένα με ρυθμό $10^9\ Hz$. Με τέτοιο ρυθμό, καθένας από τους ανιχνευτές θα παράγει 100 Terabytes δεδομένων ανά δευτερόλεπτο. Είναι αδύνατο να αποθηκευθούν όλα αυτά τα δεδομένα. Πρώτον λόγω έλλειψης αποθήκευτικού χώρου και δεύτερον επειδή δεν θα ήταν δυνατή η επεξεργασία ενός τόσο μεγάλου του όγκου δεδομένων. Επιπλέον, μόνο ένα κλάσμα αυτών των δεδομένων είναι ενδιαφέρον από άποψη φυσικής. Για την επίλυση αυτού του προβλήματος οι ανιχνευτές διαθέτουν σύστημα φιλτραρίσματος για να μειώσουν την ποσότητα των δεδομένων που καταγράφονται. Στο CMS αυτό το σύστημα ονομάζεται **TRI**gger and **D**ata **A**cquisition **S**ystem (**TRIDAS**). Η διαδικασία χωρίζεται σε δύο στάδια. Το πρώτο στάδιο ονομάζεται Level-1 Trigger (L1) και το δεύτερο είναι ο High Level Trigger(HLT).

Το Level-1 Trigger είναι το πρώτο βήμα του συστήματος σκανδαλισμού του CMS. Σε αυτό το στάδιο το σύστημα έχει συνολικό χρόνο 3,2 μs για να αποφασίσει για κάθε γεγονός εάν αξίζει να αποθηκευθεί ή όχι. Στην πραγματικότητα λόγω προβλημάτων καθυστέρησης ο χρόνος είναι περίπου 1 μs. Το κομμάτι αυτό, αποτελείται εξ ολοκλήρου από hardware. Μετά από αυτό το στάδιο η συνολική ποσότητα δεδομένων είναι περίπου 50 kHz τα οποία στη συνέχεια μεταφέρονται στο HLT για περαιτέρω φιλτράρισμα. Το HLT αποτελείται αποκλειστικά από λογισμικό. Χρησιμοποιεί αλγόριθμους παρόμοιους με αυτούς που χρησιμοποιούνται από φυσικούς στις αναλύσεις τους. Μετά την ολοκλήρωση της διαδικασίας, ο ρυθμός δεδομένων έχει πέσει περίπου στα 100 Hz. Τα δεδομένα που επιλέγονται από το σύστημα αποθηκεύονται και στη συνέχεια και χρησιμοποιούνται για ανάλυση από τους φυσικούς.

Κεφάλαιο 12

Ανακατασκευή αντικειμένου

Ένα από τα πιο σημαντικά πράγματα σε ένα πείραμα φυσικής υψηλής ενέργειας είναι η σωστή ανακατασκευή των σωματιδίων που δημιουργούνται μέσα στον ανιχνευτή σαν αποτέλεσμα των συγκρούσεων. Καθώς αυτά τα σωματίδια διασχίζουν τον όγκο του ανιχνευτή, αλληλεπιδρούν με το υλικό του και παράγουν κάποιο σήμα. Συνήθως οι ανιχνευτές αποτελούνται από διαφορετικούς τύπους συστημάτων υποανίχνευσης. Τα σήματα που παράγονται από τα διάφορα σωματίδια ποικίλλουν ανάλογα με τον τύπο και τα χαρακτηριστικά των σωματιδίων καθώς και από το είδος του ανιχνευτή με τον οποίον αλληλεπιδρούν. Η ανακατασκευή αναφέρεται στη χρήση των διαφόρων υπογραφών/σημάτων που αφήνει πίσω του ένα σωματίδιο προκειμένου να προσδιοριστεί ο τύπος του καθώς και οι ιδιότητες του π.χ. ορμή, ενέργεια κ.λπ. Για να γίνει αυτό το CMS χρησιμοποιεί μια προσέγγιση που ονομάζεται Particle Flow (PF). Αντί να χρησιμοποιεί την πληροφορία από κάθε σύστημα υπο-ανίχνευσης ξεχωριστά, τις συνδυάζει για να επιτύχει καλύτερα αποτελέσματα. Στο επόμενο κεφάλαιο επεξηγούνται οι διάφορες τεχνικές και μέθοδοι που χρησιμοποιούνται για την ανακατασκευή αντικειμένων στο CMS.

12.1 Ανακατασκευή τροχιών και κορυφών

Το πρώτο βήμα της ανακατασκευής αντικειμένων είναι ο προσδιορισμός των τροχιών των σωματιδίων. Αυτό γίνεται χρησιμοποιώντας πληροφορίες από τον ανιχνευτή τροχιών. Το CMS χρησιμοποιεί έναν αλγόριθμο που ονομάζεται Combinatorial Track Finder (CTF). Είναι ένας επαναληπτικός αλγόριθμος που εκτελείται σε έξι επαναλήψεις και έχει πολύ υψηλή απόδοση. Το επόμενο βήμα είναι ο προσδιορισμός των κορυφών, Vertex. Ως κορυφή ορίζουμε ένα σημείο αλληλεπίδρασης. Για παράδειγμα, το σημείο όπου συνέβη μια σύγκρουση ή όπου ένα σωματίδιο δι-

ασπάστηκε σε άλλα σωματίδια. Ως κύρια κορυφή (Primary Vertex, PV) ορίζουμε το σημείο όπου συνέβη η σύγκρουση των πρωτονίων. Οι τροχιές επιλέγονται και ομαδοποιούνται με βάση την απόστασή τους από το σημείο αλληλεπίδρασης της δέσμης.

12.2 Θερμιδόμετρα

Τα θερμιδόμετρα χρησιμοποιούνται για την ανίχνευση και τη μέτρηση ουδέτερων σωματιδίων καθώς δεν αλληλεπιδρούν με τον ανιχνευτή τροχιών. Μπορούν να διαχωρίσουν την ενέργεια που προήλθε από ένα φορτισμένο ή ένα ουδέτερο σωματίδιο. Μπορούν επίσης να διαχωρίσουν τα ηλεκτρόνια καθώς και τα φωτόνια ακτινοβολίας πέδησης (Bremsstrahlung) επιτρέποντας έτσι τη μέτρηση με καλύτερη ακρίβεια για σωματίδια με πολύ υψηλή ενέργεια. Τα θερμιδόμετρα δεν αποτελούνται από ένα μόνο μέρος αλλά από πολά μικρότερα. Τα σωματίδια αφήνουν την ενέργεια τους συνήθως σε παραπάνω από ένα μέρη του θερμιδομέτρου συνεπώς για την μέτρηση της ολικής ενέργειας πρέπει να υπολογισθεί η συνεισφορά από όλα τα κομμάτια του θερμιδομέτρου.

12.3 Ο αλγόριθμος σύνδεσης

Τα διάφορα σήματα που αφήνονται πίσω σε κάθε σύστημα υποανιχνευτών του CMS χρησιμοποιούνται στον αλγόριθμο PF για την αναγνώριση σωματιδίων. Στην Εικόνα 4.1 μπορούν να φανούν οι διαφορετικές υπογραφές που αφήνουν διάφοροι τύποι σωματιδίων μέσα στον ανιχνευτή. Ένα από τα πιο θεμελιώδη μέρη του αλγορίθμου PF είναι η σύνδεση των χωριστών στοιχείων του ίδιου σωματιδίου μαζί αποφεύγοντας ταυτόχρονα τη μέτρηση του ίδιου στοιχείου παραπάνω απο μια φορά. Το κύριο κομμάτι αυτής της διαδικασίας είναι ο αλγόριθμος σύνδεσης. Ο αλγόριθμος αυτός μας επιτρέπει τη σύνδεση των διάφορων εναποθέσεων στον ανιχνευτή μεταξυ τους με σκοπό την ταυτοποίηση ενος σωματιδίου.

12.3.1 Μιόνια

Ο αλγόριθμος PF ανακατασκευάζει τα μιόνια με βάση τις πληροφορίες από τον ανιχνευτή τροχιών και το σύστημα μιονίων. Ένα ανακατασκευασμένο μιόνιο από το σύστημα μιονίων ονομάζεται standalone muon, ενώ ένα που δημιουργείται χρησιμοποιώντας τις πληροφορίες του ανιχνευτή τροχιών ονομάζεται tracker muon. Συνδυάζοντας τις πληροφορίες και από τα δύο συστήματα ανίχνευσης ανακατασκευάζεται

ένα global muon.

12.3.2 Ηλεκτρόνια και φωτόνια

Τα ηλεκτρόνια και τα φωτόνια ανακατασκευάζονται συνδυάζοντας πληροφορίες από τον ανιχνευτή τροχιών και το ECAL. Καθώς διασχίζουν το υλικό του ανιχνευτή τροχιών, τα ηλεκτρόνια μπορούν να εκπέμψουν φωτόνια μέσω της ακτινοβολίας πέδησης. Τα φωτόνια αυτά μπορούν στη συνέχεια να μετατραπούν σε ζεύγη ηλεκτρονίων / ποζιτρονίων (e^+e^-), τα οποία μπορούν επίσης να εκπέμπουν φωτόνια. Για το λόγο αυτό η ανακατασκευή φωτονίων και ηλεκτρονίων συμβαίνει μαζί.

12.3.3 Jets

Τα κουάρκ και τα γκλουόνια δεν ανιχνεύονται μεμονωμένα αλλά σχηματίζουν πιο σύνθετα σωματίδια μέσω της διαδικασίας αδρονοποίησης που περιγράφηκε προηγουμένως. Αυτά τα σύνθετα σωματίδια στη συνέχεια διασπώνται αφήνοντας τις πειραματικές υπογραφές που βλέπουμε στους ανιχνευτές. Οι πίδακες είναι ομάδες σωματιδίων που παράγονται κατά τη διάρκεια αυτής της διαδικασίας. Λόγω του μεγάλου αριθμού αυτών των σωματιδίων δεν είναι δυνατή η ανακατασκευή τους μεμονωμένα και έτσι σχηματίζονται οι δομές που ονομάζονται πίδακες. Δεν είναι επίσης πολύ σημαντικό να ανιχνευθούν τα προϊόντα των σωματιδίων που παράγονται από τη διάσπαση, αλλά μάλλον μέσω των ιδιοτήτων του πίδακα να μελετήσουμε το αρχικό σωματίδιο από το οποίο προήλθε ο πίδακας. Ο κύριος σκοπός είναι να μετρηθεί η ενέργεια του πίδακα. Οι πίδακες ανακατασκευάζονται χρησιμοποιώντας αλγόριθμους που τρέχουν πάνω στα διάφορα αντικείμενα που ανιχνεύονται στα θερμιδόμετρα προκειμένου να τα ομαδοποιήσουν. Καθώς οι πίδακες είναι σύνθετα αντικείμενα, ο ορισμός τους δεν είναι μοναδικός και εξαρτάται από τον αλγόριθμο που χρησιμοποιήθηκε για τη δημιουργία τους.

12.3.4 Ελλείπουσα εγκάρσια ενέργεια (MET)

Το CMS έχει σχεδιαστεί για να είναι ερμητικά κλειστό όπως αναφέρθηκε προηγουμένως. Αυτό γίνεται επειδή χρειάζεται να ανιχνεύσει όλα τα σωματίδια που παράγονται από τις συγκρούσεις των πρωτονίων. Ωστόσο, υπάρχουν σωματίδια όπως τα νετρίνα και άλλα θεωρητικά σωματίδια που δεν αλληλεπιδρούν με τον ανιχνευτή. Αυτά τα σωματίδια ανιχνεύονται έμμεσα με τη μέτρηση της ανισορροπίας στην εγκάρσια ενέργεια. Αυτή η ανισορροπία οφείλεται στην παραγωγή σωματιδίων ενός από αυτούς τους τύπους. Η ελλείπουσα εγκάρσια ενέργεια (Missing Transverse Energy (E_T) MET)

ορίζεται ως το αρνητικό διανυσματικό άθροισμα της εγκάρσιας ορμής (p_T) όλων των σωματιδίων που ανακατασκευάζονται με τον αλγόριθμο ροής σωματιδίων.

Κεφάλαιο 13

Συστήματα αυτόματου ελέγχου

Αυτό το κεφάλαιο εστιάζει στην περιγραφή των εννοιών που χρησιμοποιούνται στο σχεδιασμό και την κατασκευή συστημάτων ελέγχου, όπως το σύστημα αυτόματου ελέγχου που χρησιμοποιείται στον ανιχνευτή CMS (Detector Control System (DCS)), καθώς και τα εργαλεία που αναπτύχθηκαν και χρησιμοποιούνται στο CERN για τη δημιουργία συστημάτων ελέγχου.

13.1 Συστήματα SCADA

Συστήματα όπως το DCS που χρησιμοποιούνται στο CMS ανήκουν σε μια ευρύτερη κατηγορία συστημάτων που ονομάζονται (SCADA, Supervisory Control And Data Acquisition). Το SCADA ως όρος χρησιμοποιείται ευρέως για την περιγραφή συστημάτων που παρακολουθούν και ελέγχουν απομακρυσμένες διαδικασίες. Ο όρος λήψης δεδομένων στο πλαίσιο ενός SCADA δεν πρέπει να συγχέεται με αυτόν που χρησιμοποιείται σε ένα πείραμα φυσικής (βλ. Ενότητα 3.3.6), το οποίο αναφέρεται στη λήψη δεδομένων για το σκοπό ενός πειράματος φυσικής. Υποδηλώνει την μέτρηση διαφόρων τιμών με σκοπό την παρακολούθηση του συστήματος, η οποία στις περισσότερες περιπτώσεις γίνεται με πολύ χαμηλό ρυθμό σε σύγκριση με ένα σύστημα λήψης δεδομένων (Data AcQuisition system, DAQ). Αυτά τα συστήματα χρησιμοποιούνται εκτενώς στη βιομηχανία καθώς και σε πολλούς άλλους τομείς εκτός από τα πειράματα φυσικής. Γενικά ένα σύστημα αυτόματου ελέγχου αποτελείται απο 3 μέρη:

- Το **supervisory layer** το οποίο είναι υπεύθυνο για την απεικόνιση του όλου συστρήματος και για τον έλεγχο του. Είναι επίσης το κομμάτι του συστήματος με το οποίο ο χρήστης αλληλεπιδρά.

- Το **front end layer** είναι το κομμάτι του εξοπλισμού από το οποίο αποτελείται η εκάστοτε διαδικασία που πρέπει να παρακολουθηθεί

- Το **communication layer** είναι το ενδιάμεσο τμήμα μεταξύ των άλλων δύο και επιτρέπει την επικοινωνία τους. Μετατρέπει τα σήματα που έρχονται από το hardware σε μορφή ώστε να μπορεί αν είναι κατανοητή από τον χρήση και το ανάποδο.

Τα συστήματα SCADA συνήθως δεν είναι πλήρη συστήματα ελέγχου, αλλά είναι μια σειρά από εργαλεία τα οποία μπορούνα να χρησιμοποιηθούν για την ανάπτυξη ενός συστήματος ελέγχου.

13.2 WinCC_OA SCADA Toolkit

Στο CERN τα διάφορα συστήματα αυτόματου ελέγχου που χρησιμοποιούνται στο εργαστήριο όπως για παράδειγμα για τον έλεγχο του επιταχυντή LHC ή των 4 πειραμάτων που αναφέρθηκαν προηγουμένως έχουν κατασκευαστεί με ένα σύστημα SCADA. Μετά από μια εκτενή έρευνα αγοράς, το CERN διάλεξε το Wincc_OA σαν το σύστημα που θα χρησιμοποιηθεί. Στη συνέχεια του κεφαλάιου θα γίνει μια περιγραφή του συστήματος αυτού.

13.2.1 Κύρια χαρακτηριστικά

Ένα από τα κύρια χαρακτηριστικά του WinCC που έκανε το CERN να το επιλέξει ως εργαλείο για την ανάπτυξη του συστήματος SCADA του ήταν η ικανότητά του να επεκτείνετε. Πολλές από τις άλλες διαθέσιμες λύσεις SCADA έχουν ένα όριο στον αριθμό των στοιχείων που μπορούν να ελέγξουν. Το WinCC δεν έχει τέτοιο όριο αφού προσφέρει τη δυνατότητα ανάπτυξης ενός κατανεμημένου συστήματος. Προσφέρει εγγενώς τη δυνατότητα σύνδεσης διαφορετικών συστημάτων. Με αυτόν τον τρόπο ένα μεγάλο σύστημα μπορεί να χωριστεί σε πολλά μικρότερα που μαζί αποτελούν το σύστημα ελέγχου. Αυτό είναι πολύ σημαντικό καθώς επιτρέπει την κατανομή του υπολογιστικού φορτίου καθιστώντας δυνατή τη σχεδίαση και ανάπτυξη πραγματικά μεγάλων συστημάτων όπως αυτό που απαιτούνται για τη λειτουργία του LHC και των ανιχνευτών του.

Το WinCC περιέχει μια βάση δεδομένων που επιτρέπει στον χρήστη να ομαδοποιήσει τα διάφορα σήματα τα οποία έρχονται απο το hardware. Η ομαδοποίηση αυτή επιτρέπει στην πολύ γρήγορη προσπέλαση των δεδομένων της βάσης αλλά και στην

καλύτερη οργάνωση του συστήματος. Με αυτόν τον τρόπο μπορεί να επαναχρησιμοποιηθούν κομμάτια κώδικα. Η λογική μοιάζει με αυτήν που χρησιμοποιείται στον Αντικειμενοστραφή Προγραμματισμό.

Ένα άλλο πλεονέκτημα του WinCC είναι ότι μπορεί εύκολα να επεκταθεί μέσω της γλώσσας προγραμματισμού που διαθέτει. Αυτή η γλώσσα προγραμματισμού ονομάζεται ConTRoL (CTRL) και έχει χρησιμοποιηθεί στο CERN για την ανάπτυξη ενός συνόλου βιβλιοθηκών (Βλ. 5.6) που μπορούν να χρησιμοποιηθούν ως βάση για το σχεδιασμό και την ανάπτυξη συστημάτων ελέγχου. Η CTRL είναι μια γλώσσα γενικού τύπου, σαν την C και χρησιμοποιείται για τη δημιουργία κώδικα αλλά και γραφικών. Το πλεονέκτημα αυτής της γλώσσας είναι ότι είναι φιλική προς το χρήστη και έχει εύκολη εκμάθηση καθώς ακολουθεί τους γενικούς κανόνες μιας γλώσσας προγραμματισμού. Με αυτόν τον τρόπο είναι εύκολο για τους νέους χρήστες να εξοικειωθούν με το σύστημα και να μπορούν να συνεισφέρουν γρήγορα ακόμα και σε μεγάλα συστήματα. Ένα άλλο σημαντικό πλεονέκτημα του WinCC, σε αντίθεση με τα περισσότερα προϊόντα SCADA, είναι ότι μπορεί να χρησιμοποιηθεί σε διάφορα λειτουργικά συστήματα. Μπορεί να προσφέρει την ίδια λειτουργικότητα τόσο σε Windows όσο και σε Linux.

13.3 Joint COntrols Projet (JCOP)

Το Joint COntrols Projet (JCOP), δημιουργήθηκε για να παρέχει τα εργαλεία για όλα τα πειράματα στο LHC για την ανάπτυξη συστημάτων αυτομάτου ελέγχου χρησιμοποιώντας μια κοινή βάση. Κατά τη διάρκεια των πειραμάτων του LEP κατέστη σαφές ότι υπήρχε έλλειψη τυποποίησης τόσο από πλευράς hardware όσο και software. Αυτό κατέστησε δύσκολη την ενσωμάτωση και τη συντήρηση των διαφόρων εργαλείων που δημιουργήθηκαν από τα διάφορα πειράματα. Ως αποτέλεσμα, αποφασίστηκε ότι το CERN θα βασιστεί όσο το δυνατόν περισσότερο σε εμπορικά προϊόντα για hardware και software. Στόχος ήταν να δημιουργηθεί ένα σύνολο εργαλείων ώστε η προσπάθεια για ανάπτυξη και διατήρηση και ενσωμάτωση διαφόρων εργαλείων να περιοριστεί στο ελάχιστο. Ως αποτέλεσμα δημιουργήθηκε ένα framework. Περιέχει ένα σύνολο από βιβλιοθήκες, γραφικά καθώς και άλλα στοιχεία που μπορούν να επεκταθούν από τα διάφορα πειράματα. Όλα τα βασικά εργαλεία για τη δημιουργία ενός συστήματος ελέγχου είναι ήδη διαθέσιμα και δεδομένου ότι όλα τα συστήματα κατασκευάζονται χρησιμοποιώντας την ίδια βάση, μπορούν εύκολα να ενσωματωθούν.

Κάποια απο τα βασικά χαρακτηριστικά του είναι ότι διαθέτει βιβλιοθήκες και γραφικά για τη χρήση όλως των κοινών ειδών hardware που χρησιμοποιούνται απο κοινού για τα 4 πειράματα. Διαθέτει έναν εύκολο τρόπο για αλληλεπίδραση με την

εσωτερική βάση του WinCC, τόσο για προσπέλαση των ήδη υπάρχοντων στοιχείων αλλά και για τη δημιουργία καινούργιων. Επιτρέπει την μοντελοποίηση του ανιχνευτή σαν μια μηχανή πεπερασμένων καταστάσεων κάνοντας έτσι πολύ πιο εύκολη την παρακολούθηση αλλά και τον έλεγχο του. Διαθέτει εργαλεία που επιτρέπουν την επικοινωνια με εξωετρικές βάσεις δεδομένων, κυρίως της ORACLE. Δίνει την δυνατότητα εύκολης παρακολούθησης και διαχείρησης των διαφόρων επιμέρους υποσυστημάτων του κυρίως συστήματος. Επίσης περιέχει βιβλιοθήκες που επιτρέπουν τον ελέγχου της πρόσβασης στο σύστημα. Αυτό γίνεται με τη δημιουργία χρηστών, ομάδων και ρόλων. Επιπλέον είναι δυναττόν να καθοριστεί ακριβώς τι δικαιώματα έχει κάθε χρήσστης μέσα στο σύστημα. Η λειτουργία αυτή επιτρέπει την εφαρμογή προ-τοκόλλων ασφαλείας στη λειτουργία του ανιχνευτή κάτι που είναι πολύ σημαντικό για τη διασφάλιση της ομαλής λειτουργίας του πειράματος.

Κεφάλαιο 14

Σύστημα αυτόματου ελέγχου του ανιχνευτή CMS

Το σύστημα αυτόματου ελέγχου του ανιχνευτή CMS είναι η κύρια διεπαφή μεταξύ των χειριστών και ανιχνευτή. Χρησιμοποιείται για την παρακολούθηση και τον έλεγχο του ανιχνευτή καθώς και άλλων περιφερειακών συστημάτων του πειράματος CMS. Στο κεφάλαιο αυτό παρουσιάζεται η σχεδίαση και ο τρόπος με τον οποίο υλοποιήθηκε το σύστημα αυτό.

Ο κύριος σκοπός του συστήματος αυτόματου ελέγχου του ανιχνευτή είναι να φέρει τον ανιχνευτή CMS σε μια κατάσταση όπου μπορεί να καταγράψει δεδομένα φυσικής με ασφαλή και ελεγχόμενο τρόπο. Για να επιτευχθεί αυτό, εκατομμύρια παράμετροι παρακολουθούνται κάθε χρονική στιγμή. Αυτές οι παράμετροι καταγράφονται και χρησιμοποιούνται ακόμα και όταν ο ανιχνευτής δεν συλλέγει δεδομένα. Μπορούν να χρησιμοποιηθούν ως μια πρώτη ένδειξη εκτίμησης της ποιότητας των δεδομένων που καταγράφηκαν. Εάν για παράδειγμα ένα τμήμα του ανιχνευτή δεν λειτουργούσε ή παρουσίαζε κάποια βλάβη κάποια χρονική στιγμή, τότε τα δεδομένα που καταγράφηκαν εκείνη τη στιγμή ενδέχεται να μην είναι κατάλληλα για μελέτη κάποιου φυσικού φαινομένου. Μπορούν ωστόσο να χρησιμοποιηθούν για τον προσδιορισμό της αιτίας του προβλήματος αλλά και για τη βαθμονόμηση και τη μέτρηση της απόδοσης άλλων τμημάτων του ανιχνευτή. Επιπλέον, η συλλογή δεδομένων από το DCS χρησιμοποιείται για την ανακατασκευή των διαφόρων σωματιδίων. Κάθε υποανιχνευτικό σύστημα συμπεριφέρεται διαφορετικά ανάλογα με τις περιβαλλοντικές συνθήκες. Αυτές παρακολουθούνται και καταγράφονται από το DCS και οι τιμές τους χρησιμοποιούνται για τον προσδιορισμό και τη βαθμονόμηση της απόκρισης του ανιχνευτή.

Σε αυτό το μέρος είναι σημαντικό να διευκρινιστεί ότι το DCS δεν είναι σύστημα

ασφαλείας. Αν και η λέξη ασφαλής αναφέρθηκε πριν, ο ρόλος του δεν είναι να προστατεύσει τον ανιχνευτή. Παρόλο που παρέχει πληροφορίες που μπορούν να χρησιμοποιηθούν για τη λήψη προληπτικών ενεργειών από τους χειριστές, δεν είναι σύστημα ασφαλείας. Σε άλλα πειράματα του LHC το Σύστημα Ασφάλειας του Ανιχνευτή (Detector Safety System DSS) και το DCS είναι ενσωματωμένα μαζί. Αυτό δεν συμβαίνει στο πείραμα CMS όπου υπάρχει ένα ξεχωριστό ειδικό σύστημα το οποίο ασχολείται με την ασφαλεία του ανιχνευτή.

Εννοιολογικά το DCS υλοποιείται ως κατανεμημένο σύστημα. Αποτελείται από περίπου 30 υποσυστήματα το καθένα με συγκεκριμένο ρόλο. Μερικά είναι γενικής φύσεως και προσφέρουν λειτουργικότητα που χρησιμοποιείται από όλα τα υποσηστήματα. Για παράδειγμα ο έλεγχος εξοπλισμού ο οποίος είναι κοινός για όλα τα μέρη του πειράματος ή για τον μηχανισμό ελέγχου πρόσβασης. Άλλα έχουν συγκεκριμένο ρόλο μέσα στο πείραμα όπως για παράδειγμα τον έλεγχος του συστήματος της υψηλής τάσης του ηλεκτρομαγνητικού θερμιδόμετρου. Καθένα από αυτά τα συστήματα υλοποιείται σε δύο αντίγραφα τα οποία είναι πανομοιότυπα και επικοινωνούν μεταξύ τους ανά πάσα στιγμή. Το ένα είναι ενεργό ενώ το άλλο είναι σε κατάσταση αναμονής. Σε περίπτωση που το ενεργό σύστημα παρουσιάσει βλάβη, το δεύτερο παίρνει τη θέση του ενεργού στο σύστημα. Για την υλοποίηση του συστήματος χρησιμοποιούνται 60 servers, δύο για κάθε σύστημα. Ένα τρέχει το ενεργό σύστημα και το άλλο το παθητικό. Εκτός από το ότι βρίσκονται σε διαφορετικούς, servers και οι servers αυτοί βρίσκονται επίσης γεωγραφικά σε διαφορετικά σημεία. Αυτό διασφαλίζει ότι σε οποιαδήποτε κατάσταση ένα από τα 2 συστήματα θα είναι λειτουργικό.

Χρησιμοποιώντας το παράδειγμα του JCOP, το σύστημα ελέγχου στο CMS δεν είναι μια ενιαία εφαρμογή. Είναι ένα σύνολο μικρών διακριτών κομματιών λογισμικού τα οποία μπορούν να εγκατασταθούν ξεχωριστά και ονομάζονται components. Το καθένα από αυτά παρέχει συγκεκριμένες λειτουργίες. Καθένα από τα 30 συστήματα που αναφέρονται παραπάνω ξεκινά από το ίδιο ακριβώς σημείο. Στη συνέχεια, ανάλογα με τον ρόλο του στο σύστημα εγκαθίστανται ένα σύνολο components δίνοντάς του την τελική του μορφή. Το DCS του CMS αποτελείται από περίπου 200 components. Ορισμένα είναι γενικής φύσεως και εγκαθίστανται παντού, ενώ άλλα παρέχουν συγκεκριμένη λειτουργικότητα και εγκαθίστανται σε ένα από σύστημα.

Κεφάλαιο 15

Μέτρηση διαφορικής ενεργού διατομής

Αυτό το κεφάλαιο περιέχει πληροφορίες σχετικά με τη ανάλυση φυσικής που έγινε στο πλαίσιο αυτής της διατριβής. Η ανάλυση αφορά τη μέτρηση της διαφορικής ενεργού διατομής του ζεύγους $t\bar{t}$ υψηλής εγκάρσιας ορμής όπου και τα δύο top κουάρκ διασπώνται αδρονικά. Εισάγεται και εξηγείται η μέθοδος που χρησιμοποιήθηκε και τέλος παρουσιάζονται και σχολιάζονται τα αποτελέσματα. Αυτή η ανάλυση είναι μια μέτρηση ακριβείας που επιτρέπει να εξεταστούν οι προβλέψεις της QCD και πιθανόν να περιοριστούν περαιτέρω οι διάφορες παράμετροι της. Όπως αναφέρθηκε προηγουμένως, δεδομένου ότι αυτή η συγκεκριμένη διαδικασία αποτελεί υπόβαθρο για πολλές επεκτάσεις του SM, ο ακριβής υπολογισμός της είναι απαραίτητος για τον έλεγχο αυτών των μοντέλων. Στη συγκεκριμένη ανάλυση η τελική κατάσταση αποτελείται από σωματίδια που έχουν υψηλή εγκάρσια ορμή (p_T). Οι συνθήκες κάτω από τις οποίες πραγματοποιήθηκαν οι συγκρούσεις στο LHC επιτρέπουν μια τέτοια μέτρηση, καθώς η ενέργεια είναι αρκετά υψηλή ώστε να μπορούν να παραχθούν σωματίδια με τόσο υψηλη ορμή. Επίσης, ο υψηλός ρυθμός συγκρούσεων καθιστά τον LHC ένα εργοστάσιο παραγωγής top κουάρκ έτσι ώστε αρκετά γεγονότα να έχουν κινηματικά χαρακτηριστικά σε αυτό το κομμάτι του φασικού χώρου ώστε να είναι εφικτή αυτή η ανάλυση.

Η μεγάλη πρόκληση αυτής της ανάλυσης είναι η ανακατασκευή των προϊόντων διάσπασης του top κουάρκ. Όπως αναφέρθηκε στο κεφάλαιο 2 τα top κουάρκ δεν ανιχνεύονται άμεσα, αλλά αυτό που βλέπουμε είναι τα προϊόντα διάσπασής τους. Σε προηγούμενες ανάλυσεις, η μελέτη γινόταν σε ένα χώρο έως 500 GeV στον οποίο τα προϊόντα διάσπασης μπορούν να ανακατασκευαστούν ξεχωριστά. Σε υψηλότερα p_T, $(p_T/m \approx 1)$ τα προϊόντα διάσπαση είναι αρκετά κοντά μεταξύ τους και δεν

μπορούν να ανακατασκευαστούν ξεχωριστά. Τα σωματίδια που μας ενδιαφέρουν ανακατασκευάζονται σαν πίδακες μεγάλης ακτίνας. Επιλέγεται η πλήρως αδρονική τελική κατάσταση επειδή είναι ευκολότερη η ανακατασκευή των προϊόντων διάσπασης σε αυτό το κανάλι σε σχέση με τα άλλα.

Όπως συζητήθηκε προηγουμένως στο κεφάλαιο 2 τα top κουάρκ διασπώνται σε ένα μποζόνιο W και ένα b κουάρκ. Εφόσον αυτή η ανάλυση εξετάζει μόνο την αδρονική διάσπαση του μποζονίου W, τα γεγονότα που μας ενδιαφέρουν θα περιέχουν άλλους δύο πίδακες που προέρχονται από τη διάσπαση του W. Η τελική κατάσταση θα έχει τουλάχιστον 6 σωματίδια. Μπορεί ωσστόσο να περιέχει και περισσότερα καθώς επιπλέον σωματίδια μπορούν να ακτινοβοληθούν κατά τη διάρκεια της όλης διαδικασίας. Η εικόνα 7.2 δείχνει τους πιθανούς τρόπους διάσπασης του ζεύγους $t\bar{t}$ με ένα να αποσυντίθεται αδρονικά και ένα λεπτονικά. Στην περίπτωσή μας και τα δύο top κουάρκ θα πρέπει να διασπαστούν αδρονικά. Τα παρτόνια στην τελική κατάσταση θα ανακατασκευαστούν ως πίδακες μεγάλης ακτίνας. Ο ορισμός του σήματος για αυτήν την ανάλυση είναι η παρουσία δύο τέτοιων πιδάκων.

Ως υπόβαθρο προσδιορίζουμε διαδικασίες που μπορεί να έχουν παρόμοια υπογραφή στον ανιχνευτή με τη διαδικασία που μας ενδιαφέρει. Αυτά τα γεγονότα μπορεί να περάσουν τα κριτήρια επιλογής μας και να αναγνωριστούν λανθασμένα ως σήμα $t\bar{t}$. Στην ανάλυσή μας, η κύρια διαδικασία υποβάθρου είναι η παραγωγή QCD πιδάκων. Αυτό σημαίνει ότι μπορούμε να έχουμε γεγονότα που μέσω της ισχυρής αλληλεπίδρασης παράγουν άλλους πίδακες υψηλής ενέργειας που περνούν τα κριτήρια επιλογής μας και ανακατασκευάζονται ως σήμα. Αυτοί οι πίδακες δεν προέρχονται από μια συγκεκριμένη αλληλεπίδραση αλλά μπορεί να είναι αποτέλεσμα οποιασδήποτε διαδικασίας που μπορεί να εκπέμπει γκλουόνια υψηλής ενέργειας. Αυτή η διαδικασία υποβάθρου μοντελοποιείται μέσω μιας μεθόδου που βασίζεται στα δεδομένα. Αυτό σημαίνει ότι χρησιμοποιούμε πληροφορία που προέρχονται από τα δεδομένα για να προσδιορίσουμε τη συνολική συνεισφορά γεγονότων αυτού του είδους προκειμένου να την αφαιρέσουμε.

Καθώς η μέτρηση που εκτελείται σε αυτή την ανάλυση είναι διαφορική, γίνεται σε ως συνάρτηση κάποιας μεταβλητής. Πριν αναφερθούν οι μεταβλητές, χρειάζεται να γίνει κάποια εισαγωγή. Στο πείραμα CMS τα ανακατασκευασμένα αντικείμενα ταξινομούνται με βάση την εγκάρσια ορμή τους (p_T). Όταν αναγερόμαστε σε ένα αντικείμενο ως leading αναφερόμαστε στο αντικείμενο με την υψηλότερη εγκάρσια ορμή (p_T). Αυτό ισχύει για όλες τις μεταβλητές όχι μόνο για την ίδια την p_T. Για παράδειγμα, η μάζα του leading πίδακα, σημαίνει τη μάζα του πίδακα με την υψηλότερη εγκάρσια ορμή από όλους τους ανακατασκευασμένους πίδακες.

Οι μεταβλητές που χρησιμοποιούνται είναι η εγκάρσια ορμή του leading και του sub-leading πίδακα. Καθώς και η ωκύτητα (y) του leading και του sub-leading πί-

δακα. Η ωκύτητα ορίζεται από τη σχέση 7.1. Επίσης χρησιμοποιούβται οι μεταβλητές m_{JJ}, $p_{T_{JJ}}$ και y_{JJ} που είναι το άθροισμα της μάζας, της εγκάρσιας ορμής και της ωκύτητας όλων των πιδάκων του γεγονότος αντίστοιχα.

Χρησιμοποιούνται μια σειρά από κριτήρια επιλογής με σκοπό την επιλογή γεγονότων που περιέχουν ζεύγη $t\bar{t}$ υψηλής εγκάρσιας ορμής που διασπόνται αδρονικά. Τα κριτήρια αυτά εφαρμόζονται αφού γίνει κάποια βασική προεπιλογή στα ανακατασκευασμένα αντικείμενα από τον ανιχνευτή. Στον πίνακα 7.10 εμφανίζονται τα βασικά κριτίρια της ανάλυσης. Αυτή είναι η βάση και είναι κοινή σε όλες τις περιοχές που θα οριστούν αργότερα. Απαιτεί τουλάχιστον δύο πίδακες στα γεγονότα και ένα βέτο στα λεπτόνια, δηλαδή να μην υπάρχουν καθόλου λεπτόνια. Κινηματικά, οι πίδακες απαιτείται να έχουν εγκάρσια ορμή 450 και 400 GeV για τον leading και τον subleading πίδακα αντίστοιχα. Απαιτείται και οι δύο να έχουν απόλυτη ψευδοωκύτητα (Εξίζωση 7.2) μικρότερη από 2,4. Η μάζα τους απαιτείται να είναι στην περιοχή από 50 έως 300 GeV. Τέλος, η μάζα m_{jj} του συστήματος απαιτείται να είναι μεγαλύτερη από 1000 GeV.

Εκτός από τα κριτήρια επιλογής που χρησιμοποιούνται στην ανάλυση για την επιλογή γεγονότων, αναπτύχθηκε και χρησιμοποιήθηκε ένα Ενισχυμένο Δέντρο Απόφασεών (Boosted Decision Tree, BDT). Είναι ένας αλγόριθμος ο οποίος ανήκει στον τομέα της μηχανικής μάθησης. Μπορεί να εκπαιδευτεί με σκοπό τον διαχωρισμό γεγονότων που προέρχονται από τη διάσπαση $t\bar{t}$ και άλλων που προέρχονται από γεγονότα υποβάθρου. Επίσης χρησιμοποιήθηκε ένας αλγόριθμος ο οποίος παρέχεται κεντρικά από το πείραμα CMS ο οποίος χρησιμοποιείται για b-tagging, να μπορέσει δηλαδή να ξεχωρίσει πίδακες οι οποίο περιέχουν μέσα τους ένα b κουάρκ.

Συνδυάζοντας τα όλα μαζί, ορίζονται οι ακόλουθες περιοχές σε αυτήν την ανάλυση. Η περιοχή σήματος (Signal Region, SR) είναι η περιοχή στην οποία εκτελείται η μέτρηση. Περιέχει γεγονότα που ικανοποιούν τα βασικά κριτήρια επιλογής (Πίνακας 7.10). Απαιτεί και οι δύο πίδακες να έχουν μέσα τους ένα μικρότερο πίδακα ο οποίος να είναι b-tagged και να περάσουν την επιλογή του BDT. Τέλος, πρέπει να περάσουν το κριτήριο μάζας όπου η μάζα τους πρέπει να είναι στην περιοχή (120, 220) GeV. Η επόμενη περιοχή είναι η εκτεταμένη περιοχή σήματος (SR_A). Είναι παρόμοια με την SR, αλλά έχει μια διαφορά στο κριτήριο της μάζας η οποία πρέπει να είναι στο διάστημα (50, 300) GeV. Αυτή η περιοχή χρησιμοποιείται για τη διαδικασία καθορισμού του υποβάθρου QCD. Η επόμενη περιοχή είναι η περιοχή ελέγχου, που χρησιμοποιείται και αυτή για τον υπολογισμό του υποβάθρου (Control Region, CR). Περιέχει γεγονότα που περνούν τα βασικά κριτήρια καθώς και το BDT. Απαιτείται επίσης να έχουν μάζα μεταξύ 120 και 200 GeV. Οι πίδακες σε αυτήν την περιοχή απαιτείται να μην έχουν περάσει το κριτήριο του b-tagging. Αυτή η περιοχή χρησιμοποιείται για να ορίσει την κατανομή της συνεισφοράς από την QCD και στη

συνέχεια να την αφαιρέσει. Η τελική περιοχή είναι η εκτεταμένη περιοχή ελέγχου (CR_A). Είναι παρόμοια με την CR αλλά έχει ένα πιο χαλαρό κριτήριο για τη μάζα των πίδακων, (50, 300) GeV.

Όπως αναφέρθηκε προηγουμένως, τα πραγματικά σωματίδια που θέλουμε να μετρήσουμε δεν μπορούν να ανιχνευθούν άμεσα αφού διασπώνται πολύ γρήγορα. Αυτό που στην πραγματικότητα ανιχνεύεται είναι τα προϊόντα διάσπασης τους. Αλλά ακόμη και τότε μερικές φορές τα προϊόντα αυτά διασπώνται πριν μπορέσουν να ανιχνευθούν. Κατά τη διάρκεια της διαδικασίας του πειράματος μέχρι την ανίχνευση των σωματιδίων μπορούμε να ορίσουμε διάφορες περιοχές, Εικόνα 7.9.

Η πρώτη περιοχή που ορίζουμε είναι το επίπεδο των παρτονίων (parton level). Περιέχει σωματίδια που θέλουμε να μετρήσουμε, στην περίπτωσή μας το ζεύγος $t\bar{t}$. Η επόμενη περιοχή είναι το επίπεδο των σωματιδίων (particle level). Μετά την διάσπαση, την αδρονοποίηση και άλλες διαδικασίες παράγονται τα προϊόντα που θα ανιχνευθούν από τον ανιχνευτή. Το επίπεδο αυτό περιέχει τη συλλογή των σωματιδίων που θα ανιχνευθούν. Τέλος έχουμε το επίπεδο του ανιχνευτή που περιέχει αυτό που μετράται από τον ανιχνευτή. Οι περιοχές που ορίσαμε προηγουμένως αφορούσαν το επίπεδο του ανιχνευτή. Με τον ίδιο τρόπο ορίζουμε την περιοχή του σήματος στο επίπεδο παρτονίων Πίνακας 7.14 αλλά και στο επίπεδο σωματιδίων Πίνακας 7.14.

Όπως συζητήθηκε παραπάνω, η κυρίαρχη συνεισφορά υποβάθρου σε αυτή την ανάλυση προέρχεται από την παραγωγή πιδάκων μέσω της QCD. Είναι πιθανό οι πίδακες που προέρχονται από την ακτινοβολία ενός παρτονίου να μιμούνται τον πίδακα που παράγεται από ένα top κουάρκ. Η περιοχή της SR αποτελείται από τρία μέρη. Το πραγματικό μέρος $t\bar{t}$ που θέλουμε να διατηρήσουμε, το τμήμα QCD καθώς άλλες συνεισφορές από άλλες διαδικασίες. Αυτό μπορεί να αναπαρασταθεί από την εξίσωση 7.6.

Η μέθοδος που χρησιμοποιείται για τον υπολογισμό της συνεισφοράς του υποβάθρου βασίζεται στην υπόθεση ότι η αντιστροφή του κριτιρίου για το b-tagging μπορεί να μας δώσει την ίδια κατανομή με τη συνεισφορά της QCD. Αυτό σημαίνει ότι η περιοχή που παίρνουμε έχει το ίδιο ¨σχήμα¨ καθώς και ότι είναι καθαρή ότι δηλαδή περιέχει μόνο πίδακες QCD και όχι άλλων ειδών. Οι παρακάτω εικόνες χρησιμοποιούνται για τον έλεγχο της υπόθεσης αυτής, 7.10 - 7.15. Δείχνουν τη σύγκριση της κατανομής των QCD πιδάκων χρησιμοποιώντας τα κριτήρια επιλογής SR και CR όπως περιγράφηκαν προηγουμένως. Τα διαγράμματα δείχνουν μια αρκετά καλή συμφωνία που σημαίνει ότι η CR μπορεί να χρησιμοποιηθεί για να υπολογίσει κανείς την κατανομή της συνεισφοράς της QCD στην SR. Σε ορισμένες από τις μεταβλητές όπως η μάζα του leading πίδακα και του p_T των πίδακα όπου φαίνεται να υπάρχει λίγο μεγαλύτερη ασυμφωνία στα σχήματα. Για να διορθωθεί αυτό, χρησιμοποιείται ένας συντελεστής διόρθωσης.

Παρόλο που η αντιστροφή της απαίτησης για b-tagging μας δίνει την κατανομή της QCD στο SR, πρέπει να δούμε εάν εφαρμόζοντας αυτά τα κριτήρια επιλογής επιλέγεται μόνο QCD και όχι σήμα. Εάν δεν συμβαίνει αυτό, χρειάζεται να γίνει κάποια διόρθωση. Στην εικόνα 7.16 φαίνεται ότι στην περιοχή ελέγχου QCD υπάρχει μια συνεισφορά της τάξης του 10% από σήμα το οποίο πρέπει να αφαιρεθεί πριν γίνει η μέτρηση.

Χρησιμοποιώντας την CR λαμβάνεται μια κατανομή για τη συνεισφορά της QCD, αλλά πρέπει να προσδιοριστεί ο αριθμός των γεγονότων, ο παράγοντας κανονικοποίησης. Για το λόγο αυτό πραγματοποιείται ένα fit. Αυτό γίνεται στα δεδομένα και συγκεκριμένα στην περιοχή SR_A. Αυτή η περιοχή επιλέγεται καθώς είναι μεγαλύτερη και περιλαμβάνει περισσότερο σημεάι που επιτρέπει την καλύτερη μοντελοποίηση των διαφόρων κατανομών. Το fit πραγματοποιείται στην κατανομή της μάζας του leading πίδακα. Χρησιμοποιείται η εξίσωση 7.7. Από τη διαδικασία αυτή, μας ενδιαφέρει ο παράγοντας κανονικοποίησης της QCD, ο οποίος δίνεται από τον όρο N_{qcd}.

Η εκτέλεση μιας μέτρησης περιέχει πάντα κάποιο σφάλμα. Λόγω της στατιστικής φύσης της σωματιδιακής φυσικής υπάρχει κάποια εγγενής αβεβαιότητα στη μέτρηση η οποία εκφράζεται ως στατιστικό σφάλμα. Αυτό ονομάζεται τυχαίο σφάλμα καθώς η εκτέλεση του ίδιου πειράματος ξανά και ξανά θα έχει ως αποτέλεσμα μερικές φορές, τιμές μεγαλύτερες από την πραγματική τιμή και άλλες φορές χαμηλότερες τιμές. Συνολικά όμως θα μας δώσει μια καλή εκτίμηση της πραγματικής τιμής. Λόγω της αδυναμίας πρόβλεψης του πραγματικού αποτελέσματος του πειράματος αυτό το σφάλμα ονομάζεται τυχαίο. Υπάρχει ένα άλλο είδος σφάλματος που ονομάζεται συστηματικό σφάλμα. Αυτό το σφάλμα μπορεί να είναι πολλών μορφών αλλά πάντα θα μας δίνει λάθος αποτέλεσμα προς την ίδια κατεύθυνση σε σύγκριση με τη σωστή τιμή. Για παράδειγμα, η μέτρηση μιας θερμοκρασίας με ένα θερμόμετρο το οποίο είναι εσφαλμένα βαθμονομημένο κατά 1 βαθμό όλα τα αποτελέσματά μας θα είναι συστηματικά λανθασμένα κατά 1 βαθμό. Η εικόνα 7.31 δείχνει μια σχηματική εννοιολογική αναπαράσταση των δύο τύπων σφαλμάτων.

Στη συγκεκριμένη ανάλυση οι συστηματικές αβεβαιότητες χωρίζονται σε δύο κατηγορίες, στις πειραματικές και τις θεωρητικές. Οι πειραματικές αβεβαιότητες αναφέρονται σε αβεβαιότητες που εισάγονται στην ανακατασκευή των αντικειμένων όπου τα δεδομένα και οι προβλέψεις προσομοίωσης δεν ταιριάζουν. Οι θεωρητικές αβεβαιότητες επηρεάζουν κυρίως τη μοντελοποίηση των διαφόρων διεργασιών και επηρεάζουν τη μέθοδο που χρησιμοποιείται για την εξαγωγή των τελικών αποτελεσμάτων από το επίπεδο του ανιχνευτή στο επίπεδο του παρτονίων και των σωματιδίων. Για τις πειραματικές αβεβαιότητες χρησιμοποιήθηκαν οι παρακάτω: φωτεινότητα, Pileup, b-tagging, jet energy scale, jet energy resolution, προβλεψη του υποβάθρου QCD, trigger. Για τις θεωρητικές χρησιμοποιήθηκαν οι Parton Distribution Functions,

Renormalisation and Factorisation scales, strong coupling constants, matrix elements, Underlying event tune, Initial and Final state radiation, color reconnection.

Πριν τη μέτρηση της διαφορικής ενεργού διατομής στο επίπεδο των παρτονίων και των σωματιδίων, η μέτρηση πραγματοποιείται στο επίπεδο του ανιχνευτή. Αυτό περιέχει όλα τα κριτήρια επιλογής που αναφέρθηκαν παραπάνω. Η διαφορική ενεργός διατομή δίνεται από την εξίσωση 7.10. Εκτός από τη διαφορική ενεργό διατομή υπολογίζεται η κανονικοποιημένη διαφορική ενεργός διατομή. Είναι ουσιαστικά η κατανομή της διαφορικής ενεργού διατομής. Δίνεται από την εξίσωση 7.11. Τα απότελέσματα της μέτρησης αυτή δίνονται στις εικόνες 7.32 - 7.38.

Η μέτρηση της διαφορικής διατομής γίνεται με τη μέθοδο του Unfolding τόσο για τον παρτονικό χώρο όσο και για τον σωματιδιακό. Είναι μια μαθηματική διαδικασία που χρησιμοποιείται για την επίλυση αυτών που ονομάζονται αντίστροφα προβλήματα. Αυτά είναι προβλήματα όπου έχουμε τα αποτελέσματα και θέλουμε να αντιστρέψουμε την αιτία αυτών και να γυρίσουμε στην αρχική κατάσταση. Μπορεί να περιγραφεί από τη σχέση, 7.12, όπου y είναι η παρατήρηση ή η μέτρηση και x η πραγματική πραγματική τιμή. Συνήθως δεν μπορούμε ποτέ να μετρήσουμε την ακριβή αρχική τιμή για διάφορους λόγους όπως η ανάλυση του οργάνου που χρησιμοποιείται για τη μέτρηση. Αυτό μπορεί να μοντελοποιηθεί από το A. Όταν λύνουμε ένα αντίστροφο πρόβλημα προσπαθούμε να αντιστρέψουμε το αποτέλεσμα του A για να πάρουμε πίσω το x. Στην περίπτωση μας, το x είναι η πραγματική διαφορική ενεργός διατομή και το y είναι αυτό που μετράται από τον ανιχνευτή. Η εξίσωση που χρησιμοποιείται για να προκύψει το τελικό αποτέλεσμα είναι 7.13.

Οι μετρήσεις που παράγονται σε αυτή τη διατριβή συγκρίνονται επίσης με τρία θεωρητικά μοντέλα. Αυτά είναι τα POWHEG σε συνδυασμό με το PYTHIA8, aMC@NLO [FxFx] σε συνδυασμό με το PYTHIA8 και POWHEG σε συνδυασμό με το HERWIG. Η σύγκριση παρουσιάζεται ως τιμές χ^2.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα της μέτρησης της διαφορικής ενεργού διατομής. Αρχικά παρουσιάζεται ένας πίνακας με τις τιμές χ^2 και p 7.21 - 7.24 και στη συνέχεια τα διαγράμματα της μέτρησης 7.53 - 7.66.

Από τα αποτελέσματα φαίνεται οτι υπάρχει μια απόκλιση της τάξης του 35% μεταξύ της θεωρίας και τουυ πειράματος για όλες τις μεταβλητές. Η απόκλιση αυτή είναι σταθερή για όλα τα μοντέλα που χρησιμοποιήθηκαν για τη σύγκριση. Φαίνεται οτι η θεωρία προβλέπει μεγαλύτερο αριθμό σωματιδίων από αυτά τα οποία μετρώνται. Αντίθετα για την κανονικοποιημένη διαφορική ενεργό διατομή φαίνεται ότι υπάρχει καλή συμφφωνία μεταξύ πειράματος και θεωρίας. Η μέτρηση αυτή είναι σε συμφωνία με παλαιότερες μετρήσειες που πραγματοποιήθηκαν από τα πειράματα ATLAS και CMS. Η μέτρηση αυτή θα πρέπει να ξανα πραγματοποιηθεί με ακόμα περισσότερα γεγονότα για να βελτιωθεί η ακρίβεια της αλλά και να κατανοηθεί από ποιό κομμάτι

της θεωρίας πηγάζει αυτή η ασυμφωνία.