ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

PhD THESIS

İbrahim Soner ZORBAKIR

MEASUREMENTS OF THE DIJET MASS CROSS SECTION AND COMPARISON TO FIXED ORDER THEORY PREDICTIONS AND QCD FITS WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV FOR 2016 DATA

DEPARTMENT OF PHYSICS

ADANA-2023

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MEASUREMENTS OF THE DIJET MASS CROSS SECTION AND COMPARISON TO FIXED ORDER THEORY PREDICTIONS AND QCD FITS WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV FOR 2016 DATA

İbrahim Soner ZORBAKIR

PhD THESIS

DEPARTMENT OF PHYSICS

We certify that the thesis titled above was reviewed and approved for the award of the degree of the Philosophy of Doctorate by the board of jury on 28/03/2023.

…………………………………….	……………………………...	………………………
Prof. Dr. Aysel KAYIŞ TOPAKSU SUPERVISOR	Prof. Dr. İsa DUMANOĞLU MEMBER	Prof. Dr. Eda EŞKUT MEMBER
…………………………………...	………………………………	
Assoc. Prof. Dr. Kadri ÖZDEMİR MEMBER	Asst. Prof. Dr. Yalçın GÜLER MEMBER	

PhD Thesis is written at the Department of Physics of Institute of Natural and Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Sadık DİNÇER
Director
Institute of Natural and Applied Sciences

This study is supported by Çukurova University Scientific Research Projects Unit.
Project No: FDK-2017-9147

DEDICATION

to my beloved wife…

ÖZ

DOKTORA TEZİ

CMS DENEYİ 2016 VERİLERİ İLE $\sqrt{s} = 13$ TEV'DE İKİ JET KÜTLE TESİR KESİTİ ÖLÇÜMLERİ, TEORİ TAHMİNLERİ VE KRD KARŞILAŞTIRMALARI

İbrahim Soner ZORBAKIR

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPRATMENT OF PHYSICS

Danışman : Prof. Dr. Aysel Kayış TOPAKSU
Yıl: 2023, Sayfa: 123
Jüri : Prof. Dr. Aysel Kayış TOPAKSU
: Prof. Dr. İsa DUMANOĞLU
: Prof. Dr. Eda EŞKUT
: Doç. Dr. Kadri ÖZDEMİR
: Dr. Öğr. Üyesi Yalçın GÜLER

Bu tez çalışmasında, BHÇ'de yer alan CMS detektöründen, 2016 yılında $\sqrt{s} = 13$ TeV'de alınan 35.9 fb^{-1}'lık p-p çarpışma verisi kullanılarak ikili jet kütlesi tesir kesiti ölçümü $|y| \leq 2.5$ rapidite aralığında yapılmıştır. İkili jet kütlesi tesir kesitleri ölçümleri, ikili jet kütlesi ve ikili jet sisteminin rapiditesinin fonksiyonu olarak gerçekleştirilmiştir. Detektör etkilerinden arındırılmış ikili jet sinyalleri ileri veri açma teknikleri kullanılarak elde edilmiştir. Pertürbatif KRD teorisinin NNLO hesaplamaları kullanılarak teori ve veri karşılaştırması yapılmıştır. Elde edilen ölçümler KRD analizi gerçekleştirmek için kullanılmış ve PDF'lerin belirlenmesindeki etkisi araştırılmıştır.

Anahtar Kelimeler: Yüksek Enerji Fiziği, CMS, KRD, İki jet, Tesir Kesiti

ABSTRACT

PhD THESIS

MEASUREMENTS OF THE DIJET MASS CROSS SECTION AND COMPARISON TO FIXED ORDER THEORY PREDICTIONS AND QCD FITS WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV FOR 2016 DATA

İbrahim Soner ZORBAKIR

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPRATMENT OF PHYSICS

Supervisor : Prof.Dr. Aysel Kayış TOPAKSU
Year: 2023, Pages: 123
Jury : Prof.Dr. Aysel Kayış TOPAKSU
: Prof.Dr. İsa DUMANOĞLU
: Prof.Dr. Eda EŞKUT
: Assoc. Prof. Kadri ÖZDEMİR
: Asst. Prof. Yalçın GÜLER

In this thesis, dijet jet mass cross-section measurement is performed in the $|y| \leq 2.5$ rapidity range using $35.9\ \mathrm{fb}^{-1}$ p-p collision data obtained from the CMS detector located on the LHC at $\sqrt{s} = 13$ TeV in 2016. Measurements are presented as a function of the dijet mass and the maximum rapidity of the dijet system. Detector effects were removed by using advanced data unfolding techniques. Theory and data comparisons were made using the NNLO calculations of the perturbative QCD theory. The measurement is used in a comprehensive QCD analysis in next-to-next-to-leading order to understand the effect of the measurement on the determination of the PDFs.

Keywords: High Energy Physics, CMS, QCD, Dijet, Cross Section

GENİŞLETİLMİŞ ÖZET

Yüksek enerjili çarpışmaların incelenmesi, maddenin yapısını, özellikle de Standart Model (SM) ile açıklanabilen güçlü, zayıf ve elektrozayıf etkileşimleri anlamak için önemli bir araç olmuştur. Bununla birlikte, SM tamamlanmamış bir teoridir ve kütle çekimi, karanlık madde ve nötrino kütleleri gibi bazı olayları açıklayamaz. Bu sorunları ele almak için birçok yeni fizik modeli önerilmiştir, ancak bunlar deneysel testler gerektirmektedir.

İsviçre-Fransa'daki Büyük Hadron Çarpıştırıcısı (BHÇ), özellikle kuarkların ve gluonların deneysel imzaları olan jetlerin gözlemlenmesi yoluyla, SM'in ötesindeki fiziği keşfetmek için kritik bir araçtır. Bu parçacıklara etki eden güçlü kuvvet, SM içindeki Kuantum Kromodinamiği (KRD) tarafından tanımlanır ve SM'in ötesindeki fiziği keşfetmek için KRD'nin daha iyi anlaşılması gerekir. İkili jet kütle tesir kesitinin ölçümleri, yüksek enerjilerde KRD'yi incelemek, parton dağılım fonksiyonlarını kısıtlamak ve yeni fizik araştırmalarının arka planını oluşturan SM süreçleri üzerinde deneysel kısıtlamalar sağlamak için kullanılabilir.

SM'deki parçacıklar iki kategoriye ayrılır: bozonlar ve fermiyonlar. Her kategori altı parçacık içermektedir. Temel fermiyonlar üç nesille ayrılmaktadır. Birinci nesil en kararlı ve en hafif, üçüncü nesil ise en ağır ve en kararsız parçacıkları içermektedir. Bu parçacıklar üç temel kuvvetin etkisi altındadır: güçlü, elektromanyetik ve zayıf kuvvet. Dördüncü kuvvet olan kütle çekimi, parçacık fiziğinde ihmal edilmiştir. Güçlü kuvvet, elektrik yüküne benzer, "renk" olarak bilinen bir özelliği temel almaktadır.

Bozonlar etkileşimlerin gerçekleşmesine aracılık eden parçacıklardır ve bu bozonlardan sadece gluonlar ve fotonlar kütlesizdir. SM'de meydana gelen kendiliğinden simetri kırılması, Higgs mekanizması tarafından gerçekleştirilir ve Higgs bozonu parçacıklara kütle veren yeni bir etkileşime aracılık eder. Higgs parçacığının keşfi, SM'in doğrulanmasında önemli bir adım olarak görülmektedir.

BHÇ gibi hızlandırıcılarda gerçekleşen hadron çarpışmalarını doğru bir şekilde tanımlamak için, protonun iç yapısını anlamak önemlidir. Bir hadron-hadron çarpışması süreci karmaşıktır ve başlangıç durumu radyasyonu, son durum radyasyonu, parton duşlarının oluşumu ve hadronlaşma dahil olmak üzere birkaç aşama içerir.

Hadron-hadron çarpışmaları, sert ve yumuşak alt süreçler olarak sınıflandırılabilir; sert alt süreçler, KRD'nin temel ölçeğinden çok daha büyük enerji ölçeklerinde parton düzeyindeki etkileşimlerle temsil edilir. Herhangi bir sert saçılma sürecini tanımlamak için, partonların momentum dağılımlarının iyi bilinmesi gerekir ki bu noktada PDF'ler (Parton Distribution Function) devreye girer. PDF'ler üzerindeki belirsizlikler fazladır ve yapılan ölçümlerin hassasiyetini etkiler. Jet tesir kesiti ölçümlerini sabit sıralı teori tahminleriyle karşılaştırmak, PDF'lerdeki kısıtlamaları iyileştirmek için önemli bir yöntemdir.

Bu tezde, iki jet kütle tesir kesitinin ölçümleri sunulmaktadır. Kullanılan çarpışma verileri CMS (Compact Muon Selenoid) deneyinde $\sqrt{s} = 13$ TeV'de alınan p-p çarpışmaları olup, 36.3 fb^{-1} bir ışıklılığa denk gelmektedir. Jetler, anti-k_t kümeleme algoritması kullanılarak yeniden yapılandırılmış ve jet enerjileri düzeltilmiştir.

Parçacık akış (PF) algoritması, fotonlar, elektronlar, müonlar, yüklü hadronlar ve nötr hadronlar için enerji ölçümleri dahil olmak üzere, çarpışmadaki her bir parçacığı yeniden yapılandırmak ve tanımlamak için kullanılmaktadır. Jetler daha sonra PF aday koleksiyonunda bir jet kümeleme algoritması kullanılarak yeniden yapılandırılır.

CMS detektörü, farklı prosedürler kullanan ve elektroniklere dayanan çeşitli alt sistemlerden oluşmaktadır. Ölçümler yapıldığında, ölçülen değer ile gerçek değer arasında dedektör yanıtından kaynaklı bir sapma vardır. Bu etki, sonlu dedektör çözünürlüğü, dedektörün doğrusal olmayan ve homojen olmayan tepkisi veya zaman içi ve zaman dışı yığılma (pile-up) gibi çeşitli nedenlerden kaynaklanabilir. Sonuç olarak, ölçülen spektrumun her bir bölmesinde eksik veya yinelenen olaylara yol açan komşu bölmeler arasında olay geçişleri gözlemlenebilir. Bu sorunu ele almak için bir veri açma (unfolding) işlemi gerçekleştirilir, böylece parçacık seviyesi ölçümü ile KRD teori tahminleri arasında bir karşılaştırma yapılabilir. Kütle ve rapidite binleri arasındaki geçişleri ve detektör seviyesinden parçacık seviyesi spektrumuna geçerken "yanlış" ve "sahte" girişlerin ele alınmasını hesaba katmak için iki boyutlu veri açma tekniği, Tikhonov düzeltmesi olmadan TUnfold paketiyle gerçekleştirilmiştir.

Bu analizde, bir dizi belirsizlik kaynağı incelenmiş ve ayrıntılı olarak sunulmuştur. Bu belirsizlikler, Chebyshev polinomları kullanılarak simetrikleştirilmiş ve yumuşatılmıştır. Tüm $|y|_{max}$ binlerindeki baskın belirsizlik kaynağı jet enerji ölçeğidir. Sistematik etkilerin tesir kesiti üzerindeki etkisi, nominal değer etrafında $\pm 1\sigma$ aralığında değişen deneysel parametrelerle hesaplanmıştır.

Pertürbatif KRD'deki (pKRD) teori tahminleri, NLO (next-to-leading order) ve NNLO (next-next-to-leading order) seviyesinde üretilmiştir. Nihai teorik tahmin, daha büyük mesafelerde meydana gelen etkileri hesaba katmak için matris elemanı hesaplamasını bir PDF seti ile birleştirerek elde edilir. Bununla birlikte, çok partonlu etkileşimler ve hadronizasyon gibi pertürbatif olmayan etkiler nihai sonuca dahil edilmez, bu nedenle bu süreçler için teori tahminlerinin düzeltilmesi gerekir. Pertürbatif olmayan bilgi eksikliği, doğrudan tahmine uygulanan bir düzeltme faktörüyle parametrize edilir. Pertürbatif olmayan düzeltmelerinin model bağımlılığı, çeşitli Monte Carlo olay üreteçlerinin sonuçları karşılaştırılarak hesaplanır ve tahminlere uygulanır.

TeV ölçeğindeki jet momentumları için, diferansiyel iki jet tesir kesitine yapılan elektrozayıf katkılar önemli hale gelir ve pKRD hesaplamasında ek olarak dikkate alınmalıdır. Bu etkiler, pKRD tahminine çarpımsal bir düzeltme faktörü uygulanarak düzeltilmiştir.

Mevcut ölçüm ile teorik tahminler arasındaki karşılaştırma bazı sapmalara rağmen iyi bir uyum göstermektedir.

İki jet ölçümlerinin proton dağılım fonksiyonları ve güçlü çiftlenim sabiti üzerindeki etkisini araştırmak için bir KRD analizi gerçekleştirilmiştir. Bu analiz için kullanılan veriler, iki jet kesitinin mevcut ölçümleri ile HERA (Hadron-Electron Ring Accelerator) çarpıştırıcı deneyleri olan H1 ve ZEUS'taki $e^{\pm}p$ çarpışmalarında elde edilen derin elastik olmayan saçılma ölçümlerinden oluşmaktadır. İki jet ölçümleri, özellikle gluon dağılım fonksiyonu için PDF'ler hakkında tamamlayıcı bilgiler sağlamak için kullanılmaktadır. PDF'lerdeki genel belirsizlik, parametreleştirme belirsizliği ile uyum, model ve ölçek belirsizliği katkılarının dörtlü toplanmasıyla elde edilir. İki jet ölçümlerinin dahil edilmesi, PDF belirsizliğinin genel olarak azalmasına katkı sağlamıştır. CMS verileriyle ve CMS verileri olmadan elde edilen sonuçlar büyük ölçüde uyumlu görünmektedir.

EXTENDED SUMMARY

The study of high-energy collisions has been an important tool for understanding the structure of matter, particularly the strong, weak, and electroweak (EW) interactions within the SM. However, the SM is an incomplete theory and cannot explain certain phenomena, such as gravity, dark matter, and neutrino masses. To address these issues, many new physics models have been proposed, but they require experimental tests.

The Large Hadron Collider (LHC) in Switzerland-France is a critical tool for exploring physics beyond the SM, particularly through the observation of jets as experimental signatures of quarks and gluons. The strong force acting on these particles is described by Quantum Chromodynamics (QCD) within the SM, and a better understanding of QCD is necessary to explore physics beyond the SM. Measurements of the dijet cross-section can be used to study QCD at high energies, to constrain the parton distribution functions (PDF), and to provide experimental constraints on SM processes that form the background for new physics searches.

The particles in the SM are divided into two categories: bosons and fermions, with fundamental fermions further divided into quarks and leptons. Each group contains six particles, with the first generation being the most stable and lightest, while the third generation is the heaviest and most unstable. These particles are under the influence of three fundamental forces: strong, electromagnetic, and weak force. The fourth force, gravity, is negligible in particle physics.

The strong force originates in a property known as color, which is analogous to the electric charge in Quantum Electrodynamic (QED). In gauge theory, all generators are associated with a gauge boson that acts as a mediator of the corresponding interaction. Among the bosons, only gluons and photons are massless. Spontaneous symmetry breaking, which occurs in the SM, is realized by the Higgs mechanism, with the Higgs boson mediating a new interaction that gives

particles mass. The discovery of the Higgs particle has been seen as a significant step in completing the verification of the SM.

To accurately describe hadron collisions in colliders like the LHC, it is important to understand the internal structure of the proton since the actual interaction takes place between the internal constituents of the proton. The process of a hadron-hadron collision is complex and involves several stages, including initial-state radiation (ISR), final-state radiation (FSR), the formation of parton showers, and hadronization. Products that occur due to the collision but are not related to hard processes are called underlying events.

Hadron-hadron collisions can be classified into hard and soft sub-processes, with hard sub-processes being represented by parton-level interactions at energy scales much larger than the fundamental scale of QCD. To describe any hard scattering process, the momentum distributions of the partons must be well known, which is where PDFs come into play. PDFs are a major source of uncertainty and affect the sensitivity of measurements made in the analysis. Comparing measurements with fixed-order theory predictions using jet cross-sections is a promising approach for improving constraints in PDFs.

The analysis presented here concerns measurements of the double differential dijet production cross-sections as a function of the invariant dijet mass of the two leading jets in the event, in five regions of the maximum rapidity of the dijet system up to 2.5 and for dijet mass up to 10 TeV. The data used were collected by CMS at the LHC in p-p collisions at $\sqrt{s} = 13$ TeV with an integrated luminosity of 36.3 fb^{-1}. The jets were reconstructed using the anti-k_t clustering algorithm and the data cross-sections are corrected for detector effects and compared with theoretical calculations at the NNLO in perturbative QCD (pQCD).

The PF technique is used to reconstruct and identify each particle in the collision event, including the energy measurements for photons, electrons, muons, charged hadrons, and neutral hadrons. Jets are then reconstructed using a jet clustering algorithm on the PF candidate collection, and a set of JECs (Jet Energy Corrections) is derived to correct for different sources of systematic bias that affect the reconstruction.

The CMS detector is made up of various subsystems that rely on electronics, which can lead to limitations in the precision and execution of different procedures. When measurements are performed, there is typically a deviation between the measured value and the true value of an observable, known as the detector response.

This effect can be caused by various sources such as the finite detector resolution, non-linear and inhomogeneous response of the detector, or in-time and out-of-time pile-up (PU). Consequently, event migrations can be observed among neighboring bins, leading to miss or fake events in each bin of the measured spectrum.

To address this issue, an unfolding process is performed to correct for the detector effects and unsmear the detector level spectrum, so that a comparison can be made between the particle level measurement and QCD theory predictions.

This process involves two-dimensional unfolding to account for migrations between mass and rapidity bins and the handling of miss and fake entries when going from the detector level to the particle level spectrum. TUnfold package is used with a pseudo-inversion approach that utilizes twice as many bins at the detector level as at the particle level, without Tikhonov regularization, to achieve better stability compared to plain matrix inversion.

In this analysis, a range of uncertainty sources has been examined and presented in detail. These uncertainties have been symmetrized and smoothed using Chebyshev polynomials. The dominant uncertainty source across all $|y|_{max}$ bins is the jet energy scale (JES), with a total experimental uncertainty ranging from 5% to 50%. The impact of systematic effects on the cross-section is estimated by varying experimental parameters within a $\pm 1\sigma$ interval around the nominal value.

The comparison between data and fixed-order predictions in pQCD is performed at the NLO and NNLO. The final theoretical prediction is obtained by combining the Matrix Element (ME) calculation with a PDF set to account for effects happening at larger distances. However, NP (Non-perturvative) effects like multi-parton interactions (MPI) and hadronization (HAD) are not included in the final result, so the theory must be corrected for these processes. The lack of NP information is parametrized into a correction factor that is applied directly to the prediction. The model dependence of the NP corrections is evaluated by comparing results from several Monte Carlo (MC) event generators and applied to the predictions.

For jet transverse momenta in the TeV range, EW contributions to the differential dijet cross section become important and must be considered in addition to the pQCD calculation. These effects are accounted for by applying a correction factor to the pQCD prediction.

The comparison between the detector measurement and the theory predictions shows good agreement in most phase space regions, with some deviations at the lower ends of the spectra and in the outer rapidity regions.

A QCD analysis is performed to investigate the impact of dijet measurements on determining the proton PDFs and the strong coupling constant $\alpha_s(M_Z)$. The data used for this analysis consists of measurements of deep inelastic scattering (DIS) obtained in $e^{\pm}p$ collisions at the HERA collider experiments H1 and ZEUS, along with the present measurements of the dijet cross-section. The dijet measurements are used to provide complementary information about the PDFs, especially for the gluon distribution function. The overall uncertainty in the PDFs is obtained by adding together the parametrization uncertainty and the sum in quadrature of the fit, model, and scale uncertainty contributions. The inclusion of the dijet measurements results in an overall reduction of the PDF fit uncertainty. Results obtained with and without the CMS data appear largely compatible.

ACKNOWLEDGEMENTS

First of all, I would like to express my endless thanks and respect to my advisor Prof. Dr. Aysel KAYIŞ TOPAKSU who guided me along the way.

I would like to thank Prof. Dr. Ayşe POLATÖZ, Prof. Dr. Eda EŞKUT and Prof. Dr. İsa DUMANOĞLU, whose knowledge and experience I have benefited from.

I would like to thank my dear friend Dr. Uğur KİMİNSU who is the co-analyzer of this research. We have always been side by side and supported each other in every way.

Many thanks to my family for their unwavering support and trust throughout my entire education life. My father passed away just before I finished this thesis. The loss of my father will always sting. But now, everything that I do is in honor of him. My dear father, wherever you are, you are gone but you will never be forgotten.

I cannot express enough thanks to my beloved wife Ebru USLAN ZORBAKIR, whose support I have always felt. I will always be grateful to her for always trusting me on this long and very difficult road.

Special thanks to the Turkish Atomic Energy Authority (TAEK) for financial support under the project codes 2014-TAEKCERN-A5.H6.F2-12 and 2018-TAEKCERN-A5.H6.F2-19.

TABLE OF CONTENT **PAGE**

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

GLOSSARY OF ABBREVIATIONS

ALICE : A Large Ion Collider Experiment

ATLAS : A Toroidal LHC ApparatuS

CERN : Conceil Europeenne pour la Recherche Nucleaire

CHS : Charged Hadron Subtraction

CMS : Compact Muon Selenoid

DAQ : Data Acquisition

DCB : Double Crystal Ball

DIS : Deep Inelastic Scattering

ECAL : Electromagnetic Calorimeter

EW : Electroweak

FSR : Final State Radiation

HAD : Hadronization

HCAL : Hadronic Calorimeter

HERA : Hadron-Electron Ring Accelerator

HLT : High-Level Trigger

ISR : Initial State Radiation

JEC : Jet Energy Correction

JER : Jet Energy Resolution

JES : Jet Energy Scale

LEP : Large Electron-Positron Collider

LHC : Large Hadronic Collider

LHCb : Large Hadron Collider beauty

LO : Leading Order

MC : Monte Carlo

ME : Matrix Element

MPI : Multi-Parton Interaction

NLO : Next-to-leading Order

NNLO	: Next-to-next-to-Leading Order
NP	: Non-perturbative
NSC	: Noise, Stochastic, Constant
p	: Proton
PDF	: Parton Distribution Function
PF	: Particle Flow
pQCD	: Perturbative QCD
PS	: Parton Shower
PSB	: Proton Synchrotron Booster
PU	: Pile-Up
QCD	: Quantum Chromodynamics
QED	: Quantum Electrodynamics
RM	: Response Matrix
SM	: Standard Model
SPS	: Super Proton Synchrotron
SUSY	: Super Symmetry
UE	: Underlying Event
σ	: Cross Section

1. INTRODUCTION

For a long time, high-energy collisions have been considered a particularly important method for investigating the structure of matter. These collisions have allowed a better understanding of the strong, weak, and electroweak EW interactions in the Standard Model (SM). Although SM has been remarkably confirmed by current experimental results, it is an inadequate and incomplete theory. Many questions are unexplained by the SM. For example, the model cannot explain gravity. It is impossible to explain experimentally observed results, by simply adding a graviton to the model. The theory needs some modifications to match the experimental results. It is also widely accepted that the SM is incompatible with general relativity, which explains the theory of gravity. Dark matter can be given as a second example. Observations show that the model explains about 4% of the matter available in the universe, of which about 23% should be dark matter that interacts only weakly with the SM fields. Still, the SM also falls short on dark matter candidates. Another example is neutrino masses, neutrinos are massless particles in SM. However, neutrino oscillation experiments have revealed neutrinos have mass. Mass terms for neutrinos can be added manually to the SM, but they introduce new theoretical problems.

Many new physics models have been proposed to solve these problems, but their validity must be tested. Many experimental tools are used to test the SM and its beyond. One of the most important of these tools is the LHC, which was established in Switzerland-France. In physics studies carried out in the LHC, jets, which are experimental signatures of quarks and gluons, are observed as final-state particles. The strong force acting on these last state particles is described by the QCD within the SM and a good understanding of QCD is mandatory to be able to probe the physics beyond the SM. Dijet cross-section measurements which will be presented in this thesis can be used to explore QCD at high energies, providing experimental constraints on SM processes that form a background to new physics searches and to constrain the parton distribution functions (PDFs).

Double differential measurement of the dijet mass cross section using p-p collision data recorded by the CMS at $\sqrt{s} = 13$ TeV in 2016 is presented in this thesis. Dijet mass cross section as a function of the dijet invariant mass is measured in five regions of the maximal absolute rapidity of the two leading jets. The analysis aims to improve the determination of parton distribution functions (PDFs) of the proton by including the presented measurements in fits to the dijet data together with deep-inelastic scattering data from HERA experiments. The uncertainties in the cross-section are evaluated, and the impact of the double differential dijet cross-section measurements on PDF determination is discussed.

This thesis is organized as follows: chapter 2 introduces the Standard Model. The LHC and the CMS detector are introduced in chapter 3 while event reconstruction and jets are discussed in chapter 4. Chapter 5 details dijet cross-section measurements. QCD analyses are discussed in Chapter 6 and finally, in Chapter 7 the conclusion is presented.

2. STANDARD MODEL

One of the many goals of fundamental physics is to reduce all natural phenomena into a fundamental theory that can be reproduced and predicted from experimental observations. Particle physics explains the basic building blocks of matter and the interactions between them and are defined by the SM. Particles in SM fall into two categories called "bosons" and "fermions". Bosons have integer spin and obey the Bose-Einstein statistics. Fermions, on the other hand, have half-integer spin and follow the Fermi-Dirac statistics. There are six leptons which are electron (e), muon (μ), and tau (τ) particles and their associated neutrinos (the electron neutrino (υ_e), muon neutrino (ν_μ), and tau neutrino (ν_τ), respectively). There are six types of quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). All stable matter in the universe consists of particles belonging to the first generation. Heavy and unstable particles, on the other hand, rapidly decay into the first generation to form stable particles. The first generation contains the up quark, the down quark, the electron, and the electron neutrino, the second generation contains the charm quark, the strange quark, the muon, and the muon neutrino, and the third generation contains the top quark, the bottom quark, the tau, and the tau neutrino. These particles are under the influence of three fundamental forces. These are strong, electromagnetic, and weak interactions. The fourth interaction, the gravitational force, plays a dominant role for all material in the universe, but it is negligible on the atomic scale. Each force has different interaction distances and magnitudes. Although the gravitational force is negligible in the SM, it has an infinite interaction distance. This force is responsible for physical events such as objects staying on the Earth's surface and planets rotating in orbit. The electromagnetic force has an infinite interaction range, but its intensity is vastly stronger than the gravitational force. The electromagnetic force acts on charged particles and is responsible for holding the atom together. The strong and weak forces have short interaction distances and are only effective at subatomic distances. The weak force is responsible for the decay of unstable subatomic particles such as mesons. The strong force is the strongest of the four fundamental forces. It holds the nucleus of an atom together by binding protons and neutrons through the exchange of gluons. The fermionic sector of the SM is summarized in Table 2.1.

Table 2.1. Fermions of the SM and their experimentally observed properties.

	Type	**Particle**	**Mass$[/c^2]$**	**$Q[e]$**	**Spin**
1st Generation	Quarks	u	$2.2\ MeV$	$^2/_3$	$^1/_2$
		d	$4.7\ MeV$	$-\ ^1/_3$	$^1/_2$
	Leptons	ν_e	$< 2\ eV$	0	$^1/_2$
		e^-	$0.511\ MeV$	-1	$^1/_2$
2nd Generation	Quarks	c	$1.28\ GeV$	$^2/_3$	$^1/_2$
		s	$96\ MeV$	$-\ ^1/_3$	$^1/_2$
	Leptons	ν_μ	$< 0.19\ MeV$	0	$^1/_2$
		u	$105.7\ MeV$	-1	$^1/_2$
3rd Generation	Quarks	t	$173.1\ GeV$	$^2/_3$	$^1/_2$
		b	$4.18\ GeV$	$-\ ^1/_3$	$^1/_2$
	Leptons	ν_τ	$< 18.2\ MeV$	0	$^1/_2$
		τ	$1776.9\ MeV$	-1	$^1/_2$

The SM is a gauge field theory based on the symmetry group SU(3) × SU(2) × U(1), where U(n) represents the unitary group of degree n, while SU(n) represents the special unitary group of degree n. SU(2) × U(1) describes the EW interactions. The electric charge Q is the generator of the QED gauge group U(1).

SU(3) describes QCD, the theory of strong interactions. The strong force originates in a property known as color. This property is analogous to the electric charge of electromagnetic interaction. Just as the electric charge forms the basis of the electromagnetic force, the color charge forms the basis of the strong force. Leptons, which have no color charge, are not affected by this force. On the other hand, quarks, which have a color charge, are under the influence of strong force.

In gauge theory, all the generators are associated with a gauge boson. Gauge bosons act as mediators of corresponding interactions. The gauge boson associated with the Q generator in QED is called a photon and is denoted as γ. The interaction between two charged particles in the QED is mediated by the exchange of photons. For example, the interaction between two electrons occurs when one electron emits a photon, and the other absorbs it. Similarly, in the SM there are eight gluons associated with strong interactions and four-gauge bosons W^+, W^-, Z^0 and γ associated with EW interactions. Among these bosons, only gluons and photons are massless. Because the symmetries induced by other generators are broken spontaneously. The masses of the W^+, W^- and Z^0 bosons are $m_{W^\pm} \sim 80.4$ GeV, $m_Z \sim 91.2$ GeV, respectively. These masses are quite large when considering elementary particles and are as heavy as rubidium atoms. Spontaneous symmetry breaking which occurs in SM is realized by the Higgs mechanism. The Higgs boson is regarded as the mediator of a new interaction in which the boson interacts with other particles, giving them their mass. The Higgs particle has been discovered both by ATLAS (A Toroidal LHC Apparatus) and CMS in the LHC and its mass m_H has been measured to be approximately 126 GeV. This discovery is seen as a big step in completing the verification of SM. Table 2.2 shows the summary of the gauge bosons in SM and their experimentally observed properties along with the interactions they mediate.

Table 2.2. Gauge bosons in the SM and their properties.

Gauge Boson	**Mass [GeV/c^2]**	**Interaction**	**$Q[e]$**	**Spin**
Gluons (g)	0	strong	0	1
Photon (γ)	0	electromagnetic	0	1
W	80.4	weak (charged)	± 1	1
Z	91.2	weak (neutral)	0	1
Higgs (H)	125.7	-	0	0

2.1. EW Theory

The unification of electromagnetic and weak interactions is explained by the EW theory, it was developed by Glashow, Weinberg, and Salam[1]. EW theory is based on the fact that electromagnetic and weak force are two different aspects of a single force, although they have different interaction properties. The EW theory can be defined by the SU(2) × U(1) group theory and has four gauge fields under local gauge transformation. The lagrangian of the EW theory can be written as the sum of the fermion and the gauge lagrangian:

$$\mathcal{L}_{EW} = \mathcal{L}_{fermion} + \mathcal{L}_{gauge} = \sum_{f=l,q} \bar{f}\, i\gamma^{\mu} D_{\mu} f - \frac{1}{4} W^{i}_{\mu\upsilon} W^{\mu\upsilon}_{i} - \frac{1}{4} B_{\mu\upsilon} B^{\mu\upsilon} \qquad (1)$$

where, $W^{i}_{\mu\upsilon}$ and $B_{\mu\upsilon}$ are the field tensors of the SU(2) and U(1) groups, respectively.

Gauge bosons of EW theory are γ, Z^0, and $W^{\mp}$, and they are all considered massless. However, the experimental results prove that Z^0, and $W^{\mp}$ have masses, and this shows that the symmetry is spontaneously broken. In the SM, spontaneous symmetry breaking is realized by adding a Higgs field to EW theory. The Higgs field interacts with the particles, providing mass to Z^0, and $W^{\mp}$. Fermion masses are introduced through their coupling to the Higgs field denoted as Yukawa-couplings. As the strength of Yukawa-couplings to the Higgs field increases, the particle's mass increases too.

2.2. QCD Theory

QCD is the theory that explains strong interactions between quarks and gluons, called partons, which cannot be observed as free particles in nature but only in color-neutral bound states. QCD theory emerges when the quark model is combined with the local group SU(3). The quark model classifies hadrons according to the quarks they consist of. Baryons consist of three quarks, while mesons consist of one quark and one anti-quark. An example of a baryon is a neutron consisting of one up and two down quarks and an example of a meson is π^+ which consists of one up and one anti-down quark. Although the quark model is well defined, the Pauli exclusion principle does not allow a particle consisting of three up quarks such as the Δ isobar. For such a particle to exist, a new quantum number, "color", was introduced. Quarks can have three different color states called red, green, and blue. Thus, a quark state can be given as in the following equation:

$$q(x) = \begin{pmatrix} q^{red}(x) \\ q^{green}(x) \\ q^{blue}(x) \end{pmatrix} \tag{2}$$

When the "color" state is introduced like the electric charge in electrodynamics, the QCD begins to replace the quark model. The structure of QED arises from the necessity of a local gauge invariant. If we consider that the requirement also applies to quarks with three different color charges in QCD, the Lagrangian equation for the free quark can be given as:

$$\mathcal{L}_{free} = \sum_{q=u,d,s\ldots} \sum_{colors} \bar{q}(x) \left(i\gamma_\mu \frac{\partial}{\partial x_\mu} - m_q \right) q(x) \tag{3}$$

Equation (3) should be invariant under the rotations of the quark fields in color space:

$$U = q^j(x) \rightarrow U_{jk}(x) q^k(x) \tag{4}$$

with $j, k \in \{1 \ldots 3\}$. There is only electric charge in electrodynamics and gauge transformation has only one phase factor,

$$U = exp(ia(x)) \tag{5}$$

However, there are three different color charges in QCD, and the phase factor becomes complex and turns into a 3×3 matrix,

$$U^\dagger U = UU^\dagger = 1 \tag{6}$$

$$Det(U) = 1 \tag{7}$$

These matrices represent the SU(3) group, where “3” is the number of color charges. The U matrix has eight independent elements and can therefore be represented by eight generators, T^a_{kj} with $a \in \{1 \ldots 8\}$,

$$U(x) = \exp(-i\emptyset_a(x) T^a) \tag{8}$$

Since U is x-dependent, the free quark Lagrangian in Equation 3 is not invariant under the transformation. Gauge or "gluon" fields are added to the equation to preserve gauge invariance:

$$\partial^{\mu} q^{j}(x) \rightarrow D_{kj}^{\mu} \equiv \left\{\delta_{kj}\partial^{\mu} - iA_{kj}^{\mu}(x)\right\} q^{j}(x) \tag{9}$$

$$A_{kj}^{\mu}(x) = A_{a}^{\mu} T_{kj}^{a}(x) \tag{10}$$

With all the changes, finally, the QCD Lagrangian can be written as:

$$\mathcal{L}_{QCD} = \sum_{q} \bar{q}\,(x)\big(i\gamma_{\mu}D^{\mu} - m_{q}\big)q(x) - \frac{1}{4g^{2}} trG^{\mu\upsilon}(x)G_{\mu\upsilon}(x) \tag{11}$$

where the first term describes the dynamics of quarks and their couplings with gluons, and the second term describes the dynamics of gluon fields. The strong coupling constant g is the QCD equivalent of the electric charge e in QED. The gluon field strength tensor is given as follows:

$$G^{\mu\nu}(x) = i[D^{\mu}, D^{\nu}] = \partial^{\mu}A^{\nu}(x) - \partial^{\nu}A^{\mu}(x) - i[A^{\mu}(x), A^{\nu}(x)] \tag{12}$$

2.3. Asymptotic Freedom and Confinement

Assuming that the electron is in a “vacuum”, then it is surrounded by a virtual cloud of electron-positron pairs, as is known in QED theory. This virtual cloud screens off the bare charge of electrons. But when going higher energy scales, the value of electron charge will be measured increasingly, and correspondingly of the coupling constant ($\alpha_{EM} \propto e^{2}$). This is the so-called “running of the coupling with the energy scale”. The situation is quite different in QCD. A quark in the vacuum of QCD is surrounded by a cloud of gluon and quark-antiquark pairs. Although the virtual quark-antiquark pairs reduce the color charge of the quark, the virtual gluons carrying a color charge increase it due to the self-coupling effect. This shows that the coupling constant value will decrease when going higher in energy scales. The running $\alpha_S(Q)$ as a function of the energy scale Q is shown in Figure 2.1. In the figure, the dashed red line represents the evolution of the world average value, and the black dots show the numbers obtained in the $\sqrt{s}$ = 13 TeV inclusive jet measurements (CMS Collaboration, 2017e).

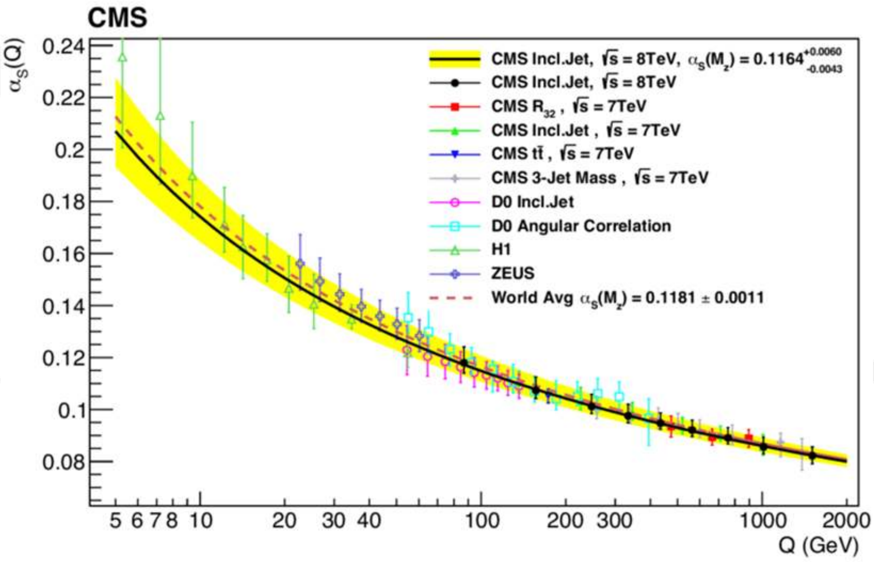

Figure 2.1. The running $\alpha_S(Q)$ as a function of the energy scale Q (CMS Collaboration, 2017e).

At short distances (high energy scale), the strong coupling constant is small enough to allow gluons and quarks to behave as free particles. This phenomenon is called "asymptomatic freedom". At long distances (low energy scale), however, the strong coupling is large enough to cause quarks and gluons to be confined within hadrons. This phenomenon is called "confinement", which means that quarks and gluons cannot behave as free particles. The major difference between QCD and

QED comes from the fact that gluons, unlike photons, carry a color charge and can interact with themselves.

2.4. Perturbative QCD (pQCD)

In hadron colliders such as the LHC, it is important to understand the internal structure of the proton to describe the resulting hadron collisions since the actual interaction takes place between the internal constituents of the proton. The complex picture of a hadron-hadron collision is illustrated in Figure 2.2. When two hadrons collide, the hard process occurs first. The emissions after the collision are called ISR. The partons that emerge after the collision can also radiate, resulting in new emissions called FSR. Later, quark-antiquark pairs begin to form parton shower (PS). These quarks and antiquarks may combine to form primary hadrons. The creation of hadrons due to the confinement phenomenon is called HAD. These primary hadrons are made up of a multitude of additional partons, and these partons can interact with the same hadron-hadron collision as well, resulting in interactions called MPI. Products that occur because of the collision but are not related to the primary vertex of the hard processes such as radiation, hadron remnants, and particles form MPI, etc. are called underlying events (UE).

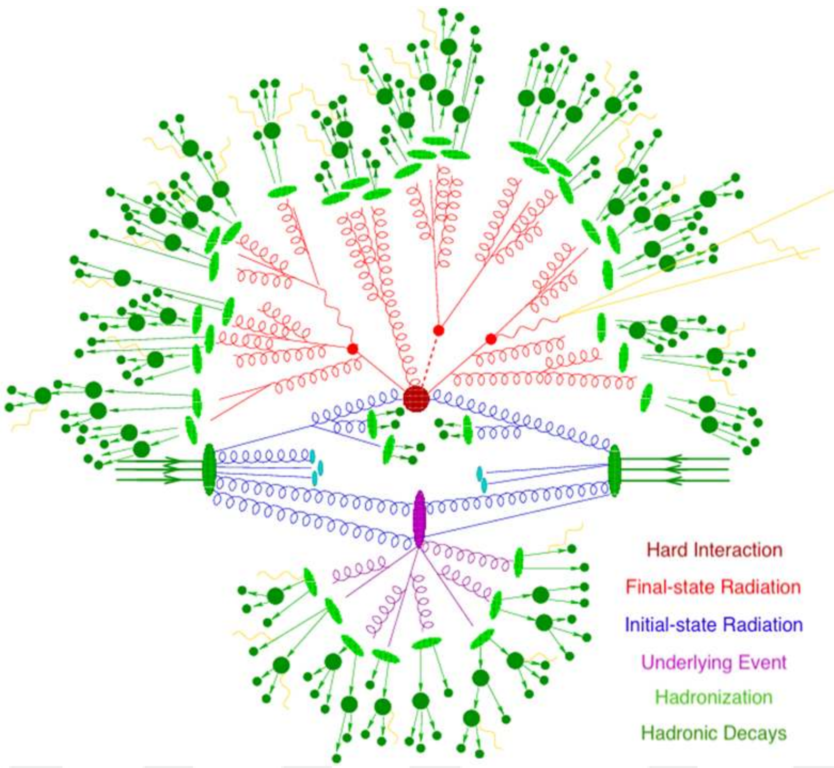

Figure 2.2. Illustration of the different stages of a hadron-hadron collision (https://www.desy.de/~reuter/downloads/vancouver_09.pdf).

Hadron-hadron collisions can be classified into two different sub-processes, called hard and soft. Soft sub-process can be defined with NP QCD, while hard sub-process representing parton level interactions, at energy scales much larger compared to the fundamental scale of the QCD, can only be defined with fixed-order pQCD. Fixed order perturbation theory provides the parton level predictions at leading order (LO), NLO and NNLO for α_S.

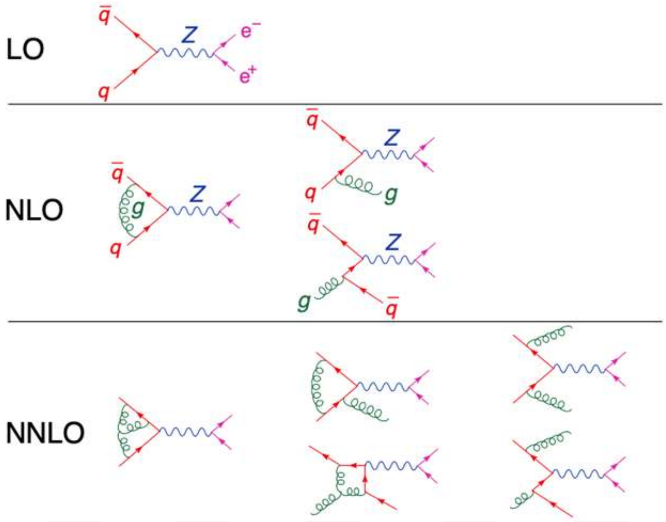

Figure 2.3. Feynman diagrams contributing to Z boson production at LO, NLO and NNLO level.

Various scattering amplitudes are required when performing pQCD calculations at the LO, NLO, and NNLO levels. Considering the inclusive production of a Z boson process, these amplitudes are shown in Figure 2.3. At the LO level, three amplitudes are required and only one Feynman diagram is to contribute to the $q\bar{q} \rightarrow Z \rightarrow e^{+}e^{-}$. At the NLO level, one-loop amplitudes contribute to virtual corrections, such as a one-loop correction to $q\bar{q} \rightarrow Z$. Additionally, these virtual corrections must be combined with real radiation i.e., tree amplitudes having one additional parton in the final state. Subprocesses of the Z boson production are $q\bar{q} \rightarrow Zg$, $qg \rightarrow Zq$ and $\bar{q}g \rightarrow Z\bar{q}$. At the NNLO level, there are three types of terms, two-loop virtual corrections, mixed virtual and real corrections from one-loop amplitudes with one additional parton, and three amplitudes with two additional partons. As the number of final state partons increases, theoretical computation gets more complicated for every order in α_S.

2.5. PDFs

The momentum distributions of the partons must be well known to describe any hard scattering process. This distribution, $f(x, Q^2)$, is called PDF. PDFs give the probability density that a certain parton with momentum fraction x will be found inside the proton at the energy scale of the interaction Q^2. Since the colliding protons are composite particles, the valence quarks they contain interact with each other through gluon exchange. At the same time, the gluons that allow the interaction to occur can temporarily create virtual quark-antiquark pairs. Thus, the valence quarks, gluons, and the virtual pairs they create become an element of the collision with an energy scale Q. PDFs are determined in many experiments studying DIS at different energy scales Q. DGLAP evolution equations (V. N. Gribov and L. N. Lipatov, 1972 and Y. L. Dokshitzer, 1977) are used to describe the evolution of PDFs measured from low-energy scales to high-energy scales. These equations allow the determination of the change in PDFs as the energy scale Q variates. PDFs obtained from measurements at fixed-target DIS and proton-nucleon experiments can be used as a foundation for the physics analyses at the LHC.

3. EXPERIMENTAL SETUP

3.1. CERN and LHC

CERN (Conseil Européen pour la Recherche Nucléaire), established in Geneva, is the world's largest high-energy research laboratory. At CERN, about ten thousand scientists from six hundred institutes in a hundred countries are working together to take our knowledge of the basic structure of matter deeper. Nowadays, our understanding of matter has advanced further than the nucleus and different theories have been put forward. These theories are tested using particle accelerators and detectors set up for specific purposes at CERN. As briefly mentioned in the Introduction section, there are some issues that SM falls short of, and these issues are currently being studied at CERN as well.

The LHC, built on the border of France and Switzerland, is designed to seek answers to crucial questions about the universe. The accelerator is placed inside the tunnel where the LEP (The Large Electron-Positron Collider) was previously located. The tunnel is 100 m below the ground and its circumference is 27 km. Protons collide head-on trying to obtain the maximum center of mass energy to form new particles. Protons move in opposite directions in the LHC at 99.999999% of the speed of light and are kept in a high vacuum in different pipes and they are held in the accelerator ring by a very strong magnetic field, which is achieved using superconducting electromagnets. The targeted center of mass energy is 14 TeV. In everyday terms, 14 TeV is not a huge amount of energy, and it is equal to the kinetic energy of a flying mosquito. What makes LHC so special is that it has focused this much energy on a point that is a hundred trillion times smaller than a speck of dust. With this energy focused on this point, the conditions in the LHC are reflecting the conditions ten to twenty-five seconds after the Big Bang, and thus, it is aimed to solve the mystery of the formation of the universe.

The acceleration of protons in LHC starts with Linac2, a linear accelerator, to an energy of 50 MeV. Then PSB (The Proton Synchrotron Booster) increases the energy up to 1.4 GeV and sends the protons to PS (Proton Synchrotron). PS increases the energy of protons to 26 GeV and transfers them to SPS (Super Proton Synchrotron). In SPS, when the protons have an energy of 450 GeV, they are finally sent to the LHC. Protons collide head-on at four points on the ring. These collision points are surrounded by large detectors called ATLAS, ALICE (A Large Ion Collider Experiment), CMS, and LHCb (Large Hadron Collider beauty). ATLAS and CMS are general-purpose detectors while others are special-purpose detectors. Figure 3.1 shows the CERN accelerator complex and the detectors on the LHC.

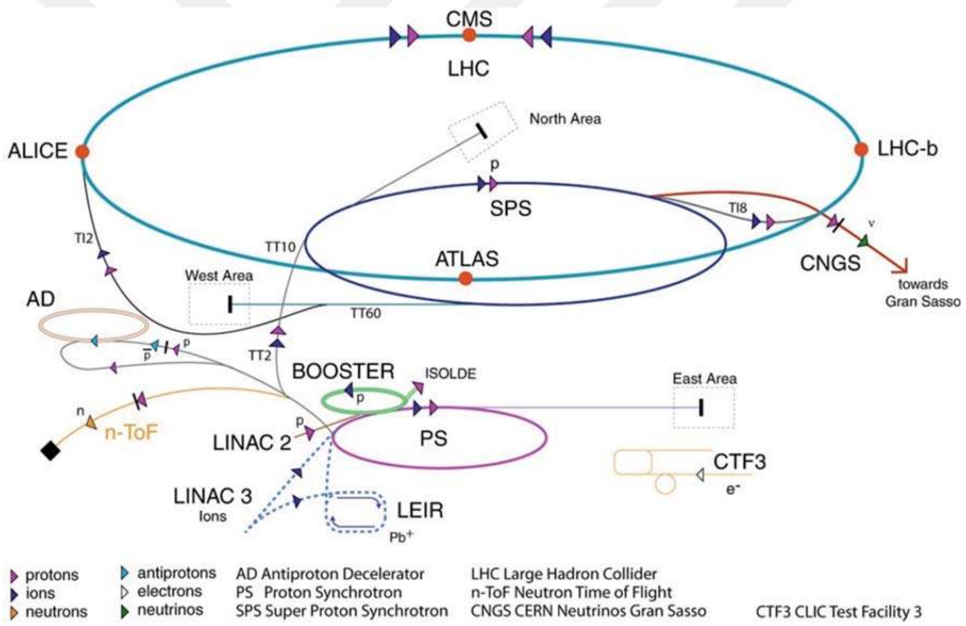

Figure 3.1. The CERN accelerator complex and the four large detectors located on the LHC ring: ALICE, ATLAS, CMS, and LHCb (Mangano, 2010).

3.2. Detectors at CERN

Events arising from the collisions in the LHC are studied using large detectors that are capable of reconstruction and can manage the high collision rate. These detectors can be thought of as large, three-dimensional cameras with millions of sensors that can take fourthy million pictures per second. Detectors have several layers, and each layer has a different functionality.

The heavy particles expected to be produced in collisions in the LHC have very short lifetimes and decay into lighter, known particles in a very short time. Particles such as electrons, muons, photons, protons, and neutrons that emerge after the collisions interact with the detector at speeds close to the speed of light. Detectors reconstruct these light particles to find the presence of new and heavy particles.

The orbits of the charged particles are bent using the magnetic field in the detectors, and the curve radius of their orbits is used to calculate their momentum. The greater the kinetic energy of the particle, the larger the trajectory it draws. Therefore, to accurately determine the radius of the curve, the trajectories of particles with high kinetic energy must be precisely defined. Another important part of the detector is the calorimeter, which is used to measure the energies of particles. Calorimeters must be built large enough to absorb as much particle energy as possible. This principle can answer the question "why are the detectors in the LHC so big?". The detectors are built in a hermitic structure to precisely measure momentum and energy and reconstruct each event in detail. The first point of interaction where particles collide head-on is also at the center of this hermitic structure. After the information from all layers of the detector is combined, the identity of the particle that leaves a track on the detector can be determined. Charged particles such as electrons, protons, and muons leave tracks on the detector through ionization. Electrons are very light particles and therefore lose their energy very quickly. Protons, on the other hand, can travel further in the detector because they are heavier than electrons. Photons do not leave tracks, but when they arrive at the calorimeter they are transformed into an electron and a positron. Then, using this information, the energies of the photons can be measured in the detector. Muons are the only particles in the detector that can reach the outermost layers because they are minimum ionized particles and therefore lose energy more slowly. Figure 3.2 shows the pattern of tracks of various particles on the detector.

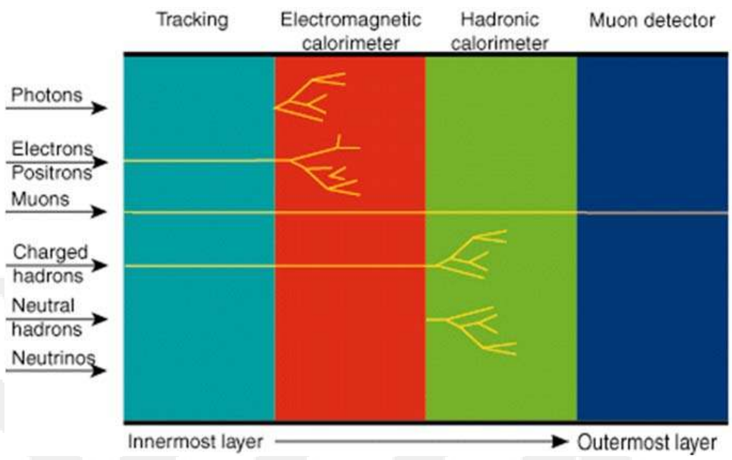

Figure 3.2. Particles and how they interact with the detector. (www.lhccloser.es/img/subidas/5_8_1_3.gif)

Each part of the detector is connected to an electronic reading system with hundreds of thousands of wires. Simply put, as soon as a signal is detected in the detector, the system records the exact location and time of the signal and instantly sends it to the computer. In this way, several hundred computers work simultaneously, and all the information is gathered. A trigger system decides in a very short time what event is relevant or not. Many different criteria in the trigger system will select potential events that have physical meaning. With these criteria, data containing one hundred million events is reduced to a few hundred thousand events for further investigation.

3.2.1. The ATLAS Detector

ATLAS is one of two general-purpose detectors available at LHC. A wide spectrum of physics is explored in the detector, such as the Higgs boson, extra dimensions, and the particles that are supposed to form dark matter. In ATLAS, the trajectories, momentum, and energies of the particles are measured by six different sub-detectors placed in layers around the collision point. The orbits of the particles are bent by a very large magnet so that the momentum of these particles can be measured. Figure 3.3 shows the ATLAS detector. The ATLAS detector is 44 m long, 25 m high, and 25 m wide and weighs 7000 tons. More than 3,000 scientists from 172 institutes in 38 countries are working on the experiment.

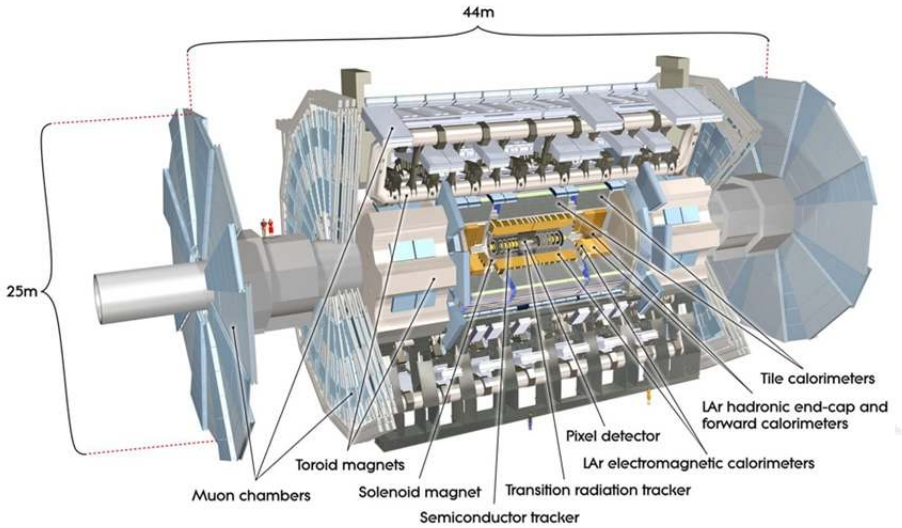

Figure 3.3. The ATLAS detector. (fisica.unipv.it/ricerca/img/ATLAS.jpg)

3.2.2. The ALICE Detector

All matter in the universe is made up of atoms. Every atom, except hydrogen, has a nucleus of protons and neutrons surrounded by a cloud of electrons. Protons and neutrons are made up of quarks and gluons that bind quarks together. Quarks have not been observed as free particles because of color confinement. In the LHC, the temperature is approximately 100.000 times higher than the center of the sun. At certain times of each year, lead ions collide in the LHC to create conditions right after the Big Bang. Under these conditions, protons and neutrons begin to melt, and the quarks and gluons they contain are released from their bonds and create quark-gluon plasma. The existence and properties of quark-gluon plasma are of great importance for understanding QCD phenomena such as color confinement and chiral-symmetry. ALICE is 16 m high, 26 m long, 16 m in diameter, and weighs 10.000 tons. The research results in the ALICE experiment are expected to provide more detailed information about the quark-gluon plasma structure and to answer the question of how plasma transformed into the particles that make up the present universe. Figure 3.4 shows the ALICE detector.

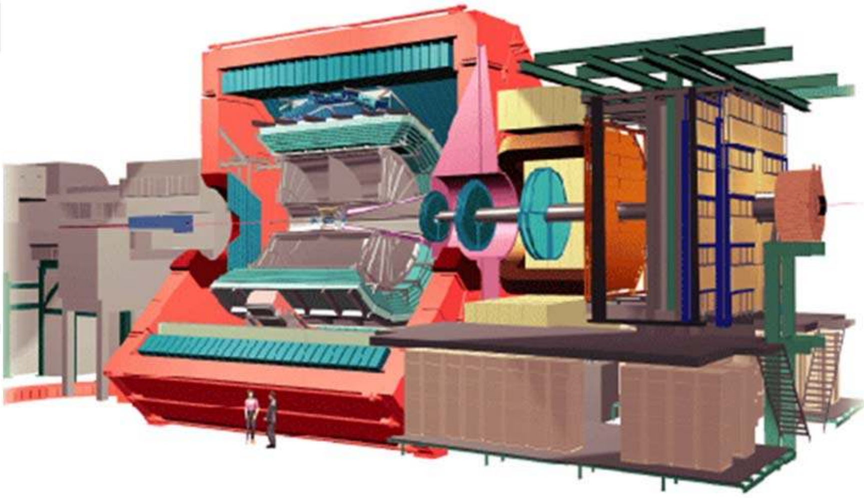

Figure 3.4. The ALICE detector (public.web.cern.ch/public/lhc/ALICE-en.html).

3.2.3. The LHCb Detector

The LHCb experiment was designed to investigate the difference between matter and antimatter by examining the interactions of the b-quark. Unlike other detectors where the collision point is surrounded by subdetectors, several sub-detectors placed in a row are used to detect particles thrown forward in a single direction in the LHCb experiment. In LHC, quarks are produced in abundance in different varieties before forming hadrons. A specially designed tracker system is used in the LHCb to capture b-quarks. The LHCb detector is 21 m long, 10 m high, and 13 m wide and weighs 5600 tons. Figure 3.5 shows the LHCb detector.

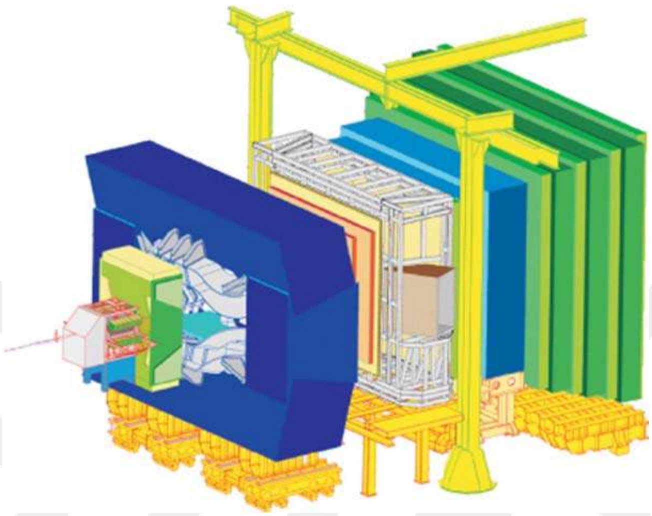

Figure 3.5. The LHCb detector (https://public.web.cern.ch/public/en/lhc/LHCb-en.html).

3.2.4. The CMS Detector

CMS is designed to uncover different signatures of new physics by precise measurements of the particles. It is not possible to accomplish this task with a single device. For this reason, sub-detectors, each of which has a different task, are designed. The point where the particle beams collide head-on, or so-called the main interaction point, is located in the center of the CMS and the subsystems are placed around the main interaction point. The main feature of the CMS is the strong magnetic field created by the superconducting coil. An electromagnetic calorimeter, a hadronic calorimeter, and a tracker are placed inside the selenoid coil. An iron yoke surrounds the coil, and muon chambers are then spaced apart to be embedded in the yoke. The CMS detector is 28.7 m long, 15 m high, and 15 m wide and weighs 14000 tons. More than 3300 scientists from 189 institutes in 41 countries are working on this experiment. Figure 3.6 shows the CMS detector and its subsystems.

Particles that are detected by the CMS are electrons, photons, hadrons, muons, neutrinos, or neutrino-like particles. While electrons and photons deposit their energies in the electromagnetic calorimeter, they do not interact with the hadronic calorimeter. Strongly interacting charged and uncharged particles deposit their energies in both electromagnetic and hadronic calorimeters. Muons, on the other hand, travel through the entire detector but deposit very little energy in the calorimeters. Therefore, muons can be identified by the signals they leave in muon chambers. Neutrinos or neutrino-like particles, such as the "neutralinos" predicted in supersymmetric theories, interact only weakly with the CMS, and cannot be detected directly. Such particles can be observed indirectly and they are determined through loss of energy in the measurements. The lost energy can be calculated by the vectorial sum of the energy found in the calorimeter.

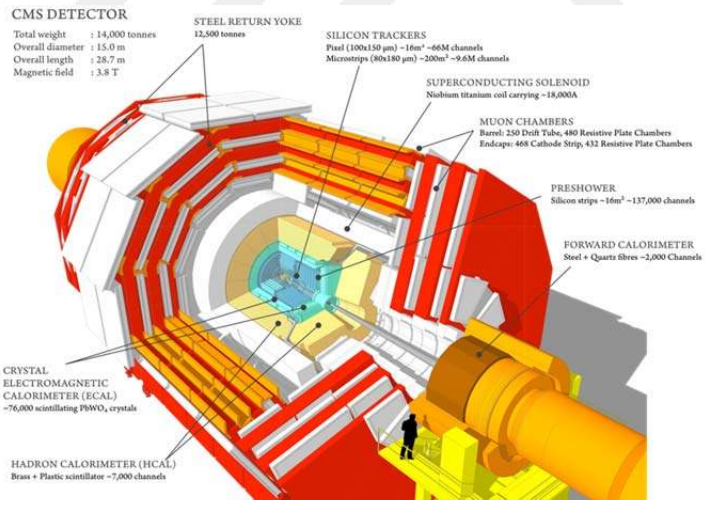

Figure 3.6. The CMS detector and its subsystems (CMS TDR, 2006).

One of the most important parts of the CMS detector is the triggering and data collection system. Millions of collisions occur within the detector per second, and only a small fraction of them have physical meaning. For this reason, events have to be selected instantly according to their characteristics and this task is conducted with the trigger system. The data collection system, on the other hand, is responsible for accumulating large amounts of data from selected events and permanently recording this data for further analysis.

The collision point is accepted as the origin in the coordinate system adopted by the CMS. The y-axis points vertically upwards and the x-axis points radially in the radius direction to the center of the LHC. The z-axis shows the beam direction. The azimuthal angle is measured from the x-axis in the $x - y$ plane, and the polar angle θ is measured from the z-axis. Pseudo-rapidity can be defined as:

$$\eta = -\ln(\theta/2) \tag{13}$$

The transverse momentum and energy expressed by p_T and E_T respectively are measured in the detector. The energy loss in the vertical plane is expressed as E_T^{miss}. A schematic drawing of a quadrant of the CMS detector is shown in Figure 3.7 (CMS TDR, 2006).

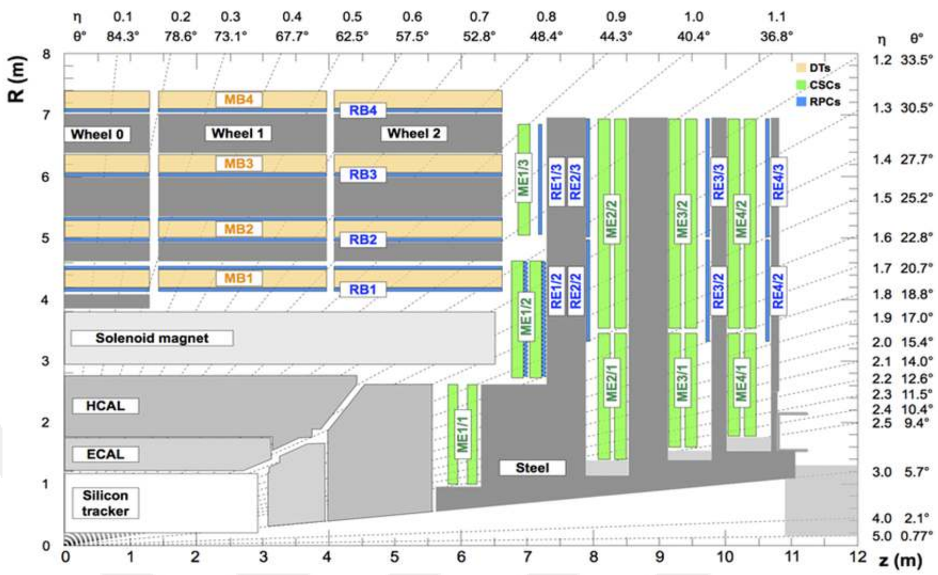

Figure 3.7. Schematic drawing of a quadrant of the CMS detector. (www.mdpi.com/2218-1997/5/1/18)

3.2.4.1. CMS Tracking System

The tracking system is the first sub-detector that particles from the collision point will encounter. The goal of the tracker is to determine the tracks of the charged particles bent by the magnetic field so that the momentum of the particles can be calculated from the curvature of their trajectories, and the electric charge of the particle can also be determined. The tracker occupies a cylindrical volume with a radius of 2.4 m and a length of 5.4 m and consists of two parts. These parts are the Silicon Pixel Detector and the Silicon Strip Detector. The two systems are made of silicon and therefore work as ionization detectors.

The Silicon Pixel Detector is located in the innermost part, close to the point of interaction. High resolution is provided through silicon blocks with the help of the pixels. The main purpose of the Silicon Pixel Detector is to determine the primary and secondary interaction points of heavy and relatively long-lived particles. The Silicon Pixel Detector also complements the track finder system in the Silicon Strip Detector. The geometric representation of the pixel detector is shown in figure 3.8.

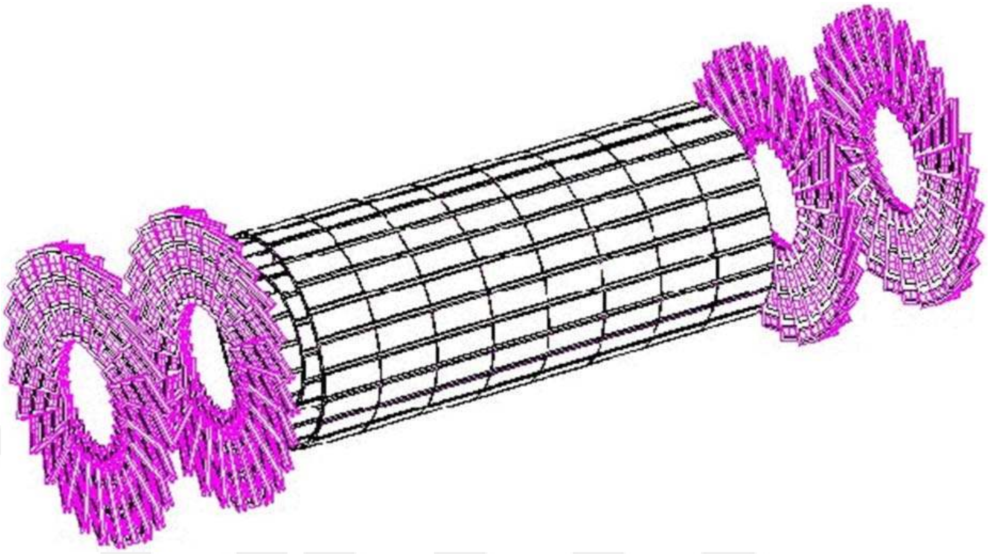

Figure 3.8. The Silicon Pixel Detector (https://relativity.phy.olemiss.edu/HEP/cms/pixel/index.html).

The Silicon Strip Detector surrounding the Silicon Pixel Detector consists of silicon micro-strips, a mechanical support structure, and reading units connected to the sensors. In the Silicon Strip Detector, thousands of optical cables are used to read the data. At the end of the detector, there are radiation resistant APV (Analogue Pipeline Voltage) chips. Because of the high radiation and low noise requirement, the tracker has to operate at -100° C. The Silicon Strip detectors are divided into the inner barrel part (TIB), the inner disks (TID), the outer barrel (TOB), and the outer endcaps (TEC+ and TEC-). The layout of the Silicon Strip Detector substructure is sketched in figure 3.9.

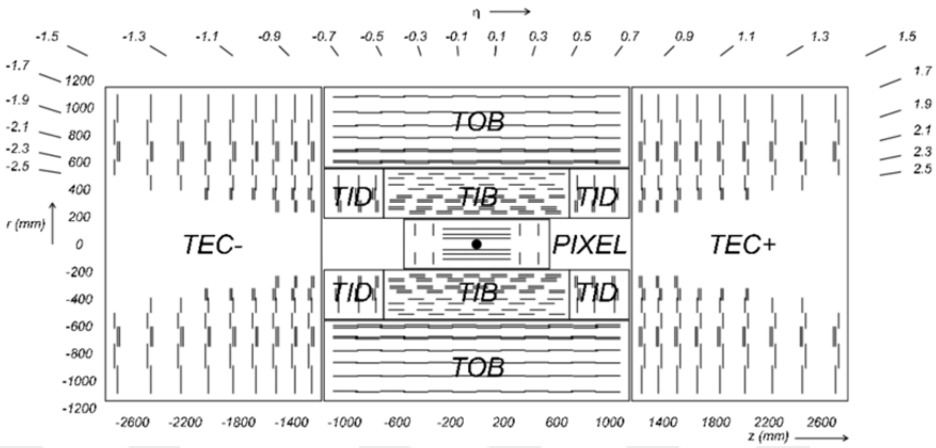

Figure 3.9. The geometric representation of the Silicon Strip Detector surrounding the Silicon Pixel Detector (https://www.researchgate.net/figure/Schematic-overview-of-the CMS-tracking-system-78-Thesiliconpixel-detector-in-the_fig7_44234277).

3.2.4.2. The Electromagnetic Calorimeter (ECAL)

Electrons and photons leave all their energy in the ECAL. These particles interact with matter through bremsstrahlung radiation and electron-positron pair formation. Lead tungsten crystals are used in ECAL, and they are made of metal and heavier than stainless steel. They are highly transparent and “scintillates” when electrons and photons pass through them. The calorimeter consists of a central barrel zone (EB) and two endcaps (EE+, EE-). The pseudo rapidity range of EB extends to $|\eta| < 1.48$ and the pseudo rapidity range of EE to $|\eta| = 3.0$. Lead tungsten crystals are used in ECAL for two reasons. The first reason is that they are very dense. Thus, electromagnetic showers occur early and can be detected in the entire calorimeter. The second reason is that they are fast enough to detect the light absorbed during the shower. To gain extra spatial precision, a pre-shower detector is placed in front of the endcaps. These help CMS to distinguish between single high-energy photons and the less interesting close pairs of low-energy photons. ECAL and its subdetectors are shown in figure 3.10.

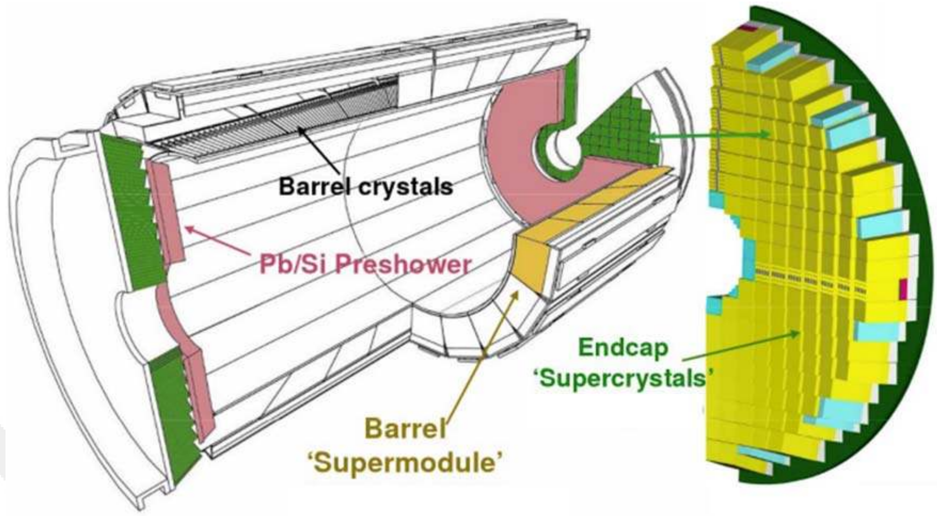

Figure 3.10. The ECAL and its subdetectors (CMS TDR, 2006).

3.2.4.3. The Hadronic Calorimeter (HCAL)

Protons, neutrons, and pions are strongly interacting particles, and the HCAL is designed to measure the energies of these particles interacting strongly. HCAL has an especially important place in the measurements of hadronic jets, neutrinos, and exotic particles that can be determined by loss of energy in the measurements. Therefore, jet resolution and precise measurement of the direction of the jets and the loss of energy are of significant importance for HCAL. In the CMS detector, hadronic showers are formed after electromagnetic showers and have longer longitudinal lengths. Hence, HCAL is thicker than ECAL and encloses it. The HCAL is a sampling calorimeter consisting of brass and stainless-steel absorbers and plastic scintillators. The signals in the scintillators are read by the photo sensors in the barrel and endcap area via wavelength-shifting optical cables. Calorimeter covers $|\eta| < 5.42$ pseudo rapidity region and consists of four subdetectors. These subdetectors are hadronic barrel (HB), hadronic endcap (HE), hadronic outer (HO), and hadronic forward (HF) calorimeters. When viewed from the interaction point, HB and HE are placed right behind the ECAL. The HB is placed between the outer radius of ECAL and the inner radius of the magnetic coil.

This caused the HB to be insufficient to fully absorb hadronic showers. Therefore, the HO is placed just behind the magnetic coil to complement the HB. After $|\eta| = 3$, HF is placed at a distance to extend the pseudo rapidity range up to $|\eta| = 5.2$. A quarter slice of HCAL is shown in Figure 3.11.

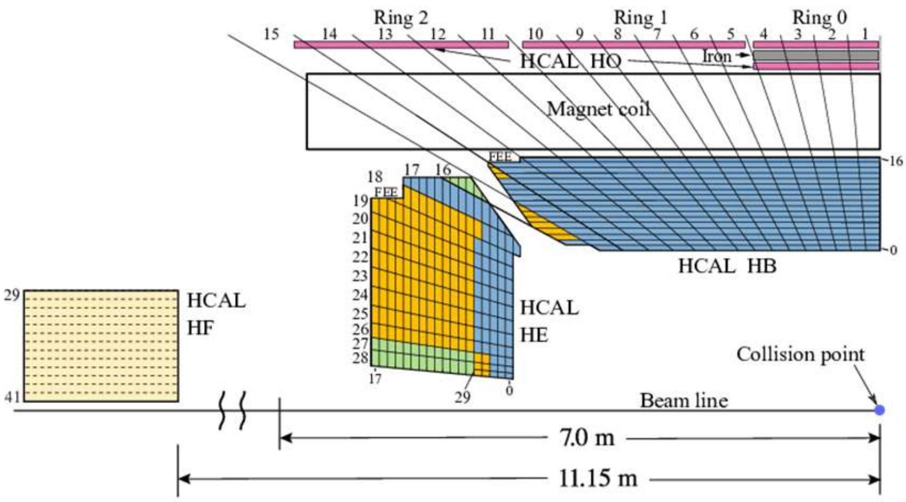

Figure 3.11. A quarter slice of the HCAL.
(https://cds.cern.ch/record/2691403/files/Figure_001.png)

3.2.4.4. Superconducting Solenoid

Momentum measurements of charged particles depend on the bending of their trajectories in the magnetic field. To obtain a strong magnetic field, a superconducting solenoid coil is used in CMS. This coil is the main part of the experiment and produces a magnetic field of 4 T. The magnetic coil is 12.5 m long and 6.3 m in diameter. The magnet used in CMS is the largest superconducting magnet ever built. It weighs 12000 tons and is cooled down to −268.5 °C. The magnetic field produced by the magnet is 100.000 times stronger than the Earth's magnetic field and has enough energy to melt 18 tons of gold. Figure 3.12 shows the magnet placed during the construction phase of the CMS. The tracker and the calorimeters are place inside the solenoid.

Figure 3.12. Superconducting Solenoid. (cms.web.cern.ch/news/superconducting-magnet).

3.2.4.5. Muon Chambers

Muon chambers are placed between the iron yokes, just outside the magnetic coil, as shown in Figure 3.13. In Figure 3.13, the iron yokes are represented in red. MS1, MS2, MS3, and MS4 indicate the muon stations. A muon can be detected by the tracks it leaves on muon stations. The tracks left by the muon in many layers in each station are followed and combined with the information coming from the tracker and hence, the trajectory of the muon through the detector is determined. By determining this trajectory, the momentum of the particle can be calculated. Thanks to the powerful magnet coil in the CMS, even trajectories of muons with very large momentum can be bent and their momentum can be measured.

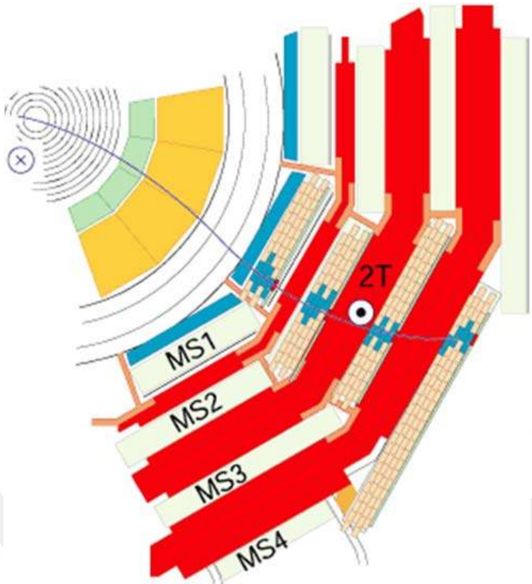

Figure 3.13. Representation of the muon that leaves a track on the muon stations placed between the iron yokes (http://cms.web.cern.ch/news/muon-detectors).

3.2.4.6. CMS Trigger and Data Acquisition (DAQ) System

p-p and heavy ion collisions occur at high interaction rates in LHC. The time interval between colliding beams is 25 ns. Depending on the luminosity, several p-p interactions occur in each colliding beam. At the expected luminosity value of $3.0 \times 10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$, approximately 60 p-p interactions take place every 25 ns. It is not possible to store and process this huge amount of data, with a size of about 7.4 MB per event. Therefore, a strong limitation must be applied to the data. This task is carried out with the trigger system. The trigger system used for data selection in CMS has two subsystems. These subsystems are called L1T (Level-1 Trigger) and HLT (High-Level Trigger).

The L1T system searches very quickly for events that may have physical meaning. A reader system that scans the main headlines of newspapers can be given as an example of its operation of it. In this way, the event count per second is reduced to thousands from billions. In the HLT system, information from different systems of the detector is processed synchronously to reconstruct a whole event. This information is then sent to the computer systems, which are the last link of the trigger system. Computers in this system work in less than a tenth of a second and looks for events that represent complex physical processes and select a hundred out of an average of one hundred thousand events, ignoring the rest. The remaining hundred events constitute data that may have physical meaning and will be stored and analyzed in more detail. Despite the existence of the trigger system in CMS, more than 600 petabytes of data (600 million gigabytes), equivalent to over 20.000 years of 24/7 HD video recording is stored.

3.2.4.7. Physics Studies at CMS

There are many unanswered questions in particle physics as stated before. The most important among these questions is the hierarchy of particle masses and how particles have these masses. According to SM, particles gain mass by the Higgs mechanism. However, the mass of the Higgs particle in the mechanism cannot be predicted by the SM. Depending on the mass of the Higgs boson, different decay channels appear in the experiments. Figure 3.14 shows simulated Higgs events. The first one shows the channel where the Higgs boson decays into two photons. In this channel, while no tracks are detected by the tracker, two high-energy clusters are formed in ECAL. The middle one shows four lepton channels. In this channel, the tracks of four leptons are seen both in the tracker and in the muon chambers, but no significant energy can be measured in the calorimeter. In the third one, the Higgs boson decays into two electrons detectable in the tracker and ECAL, and a dijet is detected in HCAL.

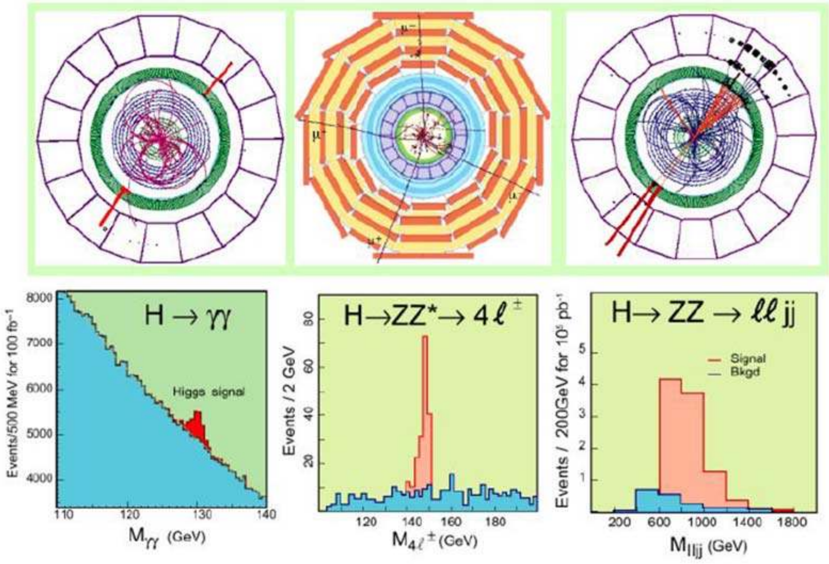

Figure 3.14. Higgs decay channels (Wulz,2008).

The SUSY theory is one of the most important candidates for beyond SM and is thought to be able to unify fundamental forces. There is an extremely broad spectrum of particles in the SUSY studied in the CMS. In addition to these particles, there are 5 different Higgs bosons.

The experiment also seeks answers to cosmological questions. One of these questions is matter-antimatter asymmetry. Another question is about dark matter, which makes up 23% of the universe. SUSY can answer these questions and studies on it continue in CMS. Apart from these, the search for massive vector bosons, extra dimensions, and heavy ion physics is within the scope of CMS's physics research.

4. EVENT RECONSTRUCTION

Once an event has passed through the trigger system, it is stored for later analysis. This stored event contains the information recorded from the sub-detectors of the CMS such as the tracks of the particles in the tracker and muon systems and the energy deposits in the calorimeters while they are passing. For the event to be analyzed, this information must be reconstructed into physics objects such as electrons, photons, muons, and jets.

4.1. Track Reconstruction

Despite not being physical objects themselves, tracks serve as the foundation for reconstructing all physical objects in the CMS. There are five steps in the reconstruction of a track. Reconstructing the tracks found within the strip and pixel detectors comes first. This is done by converting signals from the tracker electronics into detector hits via a method known as local reconstruction. The global reconstruction process is made up of the final four steps. In order to predict the charged particle tracks, the global reconstruction uses the hits obtained from the local reconstruction. Vertex reconstruction, which has a major impact on the reconstruction of physics objects, is critically dependent on the reconstruction of charged particles. The Combinatorial Track Finder (CTF) technique is used for charged particle reconstruction. The CTF operates by periodically adding an extra hit to the trajectory estimations. Iterative updates are made to the estimates until they are as precise as possible.

The global track reconstruction is accomplished in four stages. These are seed generation, trajectory construction, trajectory cleaning, and trajectory smoothing. Seeds are initially determined by double or triple hits when reconstruction begins. Once the seeds are chosen, they are spread outward by adding newly found hits in each layer. The track parameters are updated, and the trajectory is determined at each layer. Because the trajectory is affected by particle energy and charge, additional hits are added while accounting for Gaussian energy loss and multiple scattering. When no more hits can be added, all hits are fitted to produce the best estimate.

The fit procedure begins with the innermost hit and progresses outward by repeatedly extending the trajectory. The following phase is trajectory cleaning. This is accomplished by eliminating hits that contribute to inefficiency. After the trajectory has been created and cleaned, the trajectory characteristics are obtained by incorporating only the final hit. The parameters used are d_0, z_0, ϕ, $\cot\theta$, and p_T of the track. d_0 and $z_0$are the coordinates of the point of interaction along the radial and z-axis. ϕ and θ are the azimuthal and polar angles of the momentum vector associated with the corresponding track. Following the determination of the parameters, a final fit is done to achieve a more optimum estimate. It is accomplished by utilizing the results of the preceding fit and fitting from the farthest hit to the beamline. This is known as trajectory smoothing.

4.2. Primary Vertex Reconstruction

The Primary Vertex (PV) is the first point of interaction when the p-p collision occurs. PV reconstruction is carried out in three stages. These include, in order, selecting reconstructed tracks, grouping tracks from the same vertex, and fitting the tracks to establish the vertex position. Tracks are initially chosen based on how they align with the beam spot, the amount of hits from the pixel and strip detectors, and the applied fit's normalized χ^2 value. These selected trajectories are then projected toward the beamline interaction point.

A three-dimensional vertex fit is applied to PV candidates with at least two tracks. The adaptive vertex fit algorithm (Fruhwirth R., Waltenberger W. and Vanlaer P., 2007) is developed from the Kalman filter technique (Pei Y., Biswas S., Fussell D. S., Pingali K, 2019). This fit is used to calculate the location of the PV. Secondary vertexes refer to tracks that do not originate in the interaction point or are separated into more than one track. Because tracks with small p_T are more likely to arise from PU events, the vertex associated with the tracks from the hard interaction in an event is chosen by picking the PV with the total of the tracks with the largest p_T^2.

4.3. Photon Reconstruction

As a result of the electromagnetic showers, high-energy photons deposit almost all of their energy in ECAL. Because of the existence of the magnetic field, charged particles in electromagnetic showers are normally dispersed along a path. It is critical to gather the bremsstrahlung photons and electron-positron pairs that come from this dispersion in order to precisely determine the energy of the photon. CMS reconstructs the energy dispersion by applying clustering techniques that group energy depositions. After that, the combined energy depositions are regrouped to create superclusters (SCs). CMS has created two approaches for grouping electromagnetic showers. These are hybrid superclustering and multi5x5 superclustering, which are employed in the barrel and endcaps, respectively.

The hybrid superclustering algorithm aims to use showers spread in the ϕ direction to group energy depositions in the small η window. The algorithm first starts with the selection of the seed crystal whose energy threshold is both greater than $E_{T,hybseed}(1\text{ GeV})$ and has the local highest energy. After seed selection, a 3x1 dimensional rectangle is formed around the seed in the $\eta - \phi$ plane. If the center crystal has higher energy than the threshold $E_{wing}(0\text{ GeV})$, then a 5x1 dimensional rectangle is formed with the seed crystal in the center. The algorithm can be run in multiple steps in the $\eta - \phi$ plane through the crystal, and with each step, it continues to add neighboring rectangles if their energies are higher than the threshold. These adjacent rectangles are then grouped along ϕ direction to form the local maximum if the seed rectangle has an energy higher than $E_{seed}(0.35\text{ GeV})$. This local maximum is then tagged as "basic clusters".

The multi5x5 method has fewer parameters than the hybrid clustering method. The idea behind this clustering is similar in that they both group the energy deposits that make up SCs. But the algorithm works differently than the hybrid algorithm. In the hybrid algorithm, SCs are first formed, and simple clusters are obtained by decomposing the SCs, while in the multi5x5 algorithm, first simple clusters are formed and then these clusters are grouped to obtain SCs. Multi5x5 clustering starts with seed crystals that are local maxima with the highest energy deposition compared to the neighboring crystals. In addition, these crystals must have higher energy than the threshold $E_{T,EEseed}(0.18\text{ GeV})$. 5x5 crystals are formed around each seed starting from the highest energy to form clusters. To group clusters into an SC, the energy of each cluster must be greater than the threshold $E_{T,EEcluster}(1\text{ GeV})$.

4.4. Electron Reconstruction

Because electrons are charged particles, they leave hits in the tracker and deposit energy in ECAL because of electromagnetic showers. As a result, in electron reconstruction, a supercluster (SC) is generated by grouping the energy depositions in ECAL with the hits found in the silicon tracker. Section 4.3 explored the techniques used in ECAL to cluster energy deposits. Once the SCs have been reconstructed, the tracker seeds derived from two or three tracker hits are found to begin creating electron orbitals. To find these seeds, ECAL-driven or tracker-driven seeding methods are applied.

Tracker-driven seedling uses charged particle tracks, predicts ECAL surface trajectories, and binds to SCs. The classic Kalman filter track reconstruction approach for charged particles is used to reconstruct the tracks. Many hits may be missed because electrons might lose a significant amount of energy because of Bremsstrahlung. As a result, electrons are subjected to a unique track reconstruction technique. The algorithm in ECAL-guided seeding begins with an energy cluster in ECAL that propagates in the direction because of bremsstrahlung. The algorithm predicts collision vertex trajectories and picks those that correspond to the isolated high-energy electrons in the pixel detector. Following the creation of electron seeds, tracks are constructed using the Kalman Filter approach.

4.5. Muon Reconstruction

Muons are reconstructed using information from the inner tracker and muon systems. Muons can also deposit part of their energy in calorimeters. Muons are classified as "global muons" or "tracker muons" based on their reconstruction approach. In muon systems, global muons are seeded from individual tracks, whereas tracker muons are seeded from tracker hits. Track fitting is done iteratively by employing a Kalman filter approach on all available tracks. Once the individual tracks from muon system and the tracker tracks have been reconstructed, the global muon fitter looks for matching tracker tracks for each individual track. To reconstruct global muons, all of the hits in the linked tracks are fitted using a Kalman filter approach.

4.6. Jet Reconstruction

Following the reconstruction of individual charged hadrons, neutral hadrons, and photons, they are merged into PF jets by using jet algorithms. There are multiple approaches to jet reconstruction, including calorimeter jets and PF jets. Signals from the ECAL and HCAL towers are used to reconstruct colorimeter jets. The energy deposited in the calorimeters is used to identify particle clusters, which are subsequently merged to generate jets. In contrast, PF jets are reconstructed using a technique known as PF, which attempts to identify and quantify the characteristics of individual particles in the detector, such as charged particles, neutral particles, and photons. The PF methodology reconstructs individual particles and their attributes by combining information from the tracker, calorimeters, and muon detectors. Using jet clustering methods, these particles are subsequently grouped into jets.

Nearly 65% of the jet energy is carried by charged hadrons and 25% by photons, which are measured very precisely by the tracker and ECAL. The remaining 10% contribution comes from neutral hadrons which can be measured with HCAL only. The HCAL resolution is very poor compared to the resolution of ECAL and the tracker. Hence, jet energy is measured more accurately with the PF than with the jets from calorimeter towers. PF jets are typically considered to be more precise than calorimeter jets since they rely on more detailed measurements of the properties of individual particles in the detector.

Jet clustering algorithms, which apply certain rules to calorimeter towers or related particles, are used to reconstruct jets. These algorithms are created by considering the positional proximity of the jets. In simple terms, a good jet clustering algorithm should fulfill the following conditions:

- It should be fully defined to include any pre-clustering and decays in detail,
- It should be simple to implement in experimental analysis and be independent of the detector structure,
- It should be easy to use in theoretical calculations,
- It must provide a finite cross-section at any order of the perturbation theorem,
- It must be collinear and infrared safe.

Being collinear and infrared-safe is a basic requirement for jet clustering algorithms. Infrared safe means that jet clustering algorithm should not changed when a gluon is added to the jet clustering algorithm. Collinear safe means that the result of the jet clustering algorithm should not changed when one parton of the jet structure is split into two partons. Figure 4.1 shows the configuration of infrared and collinear safety.

In the CMS, jets are reconstructed using several different algorithms that are designed to identify collimated sprays of particles produced in high-energy collisions. These algorithms attempt to group individual particles into jets based on their kinematic properties and the geometry of the detector.

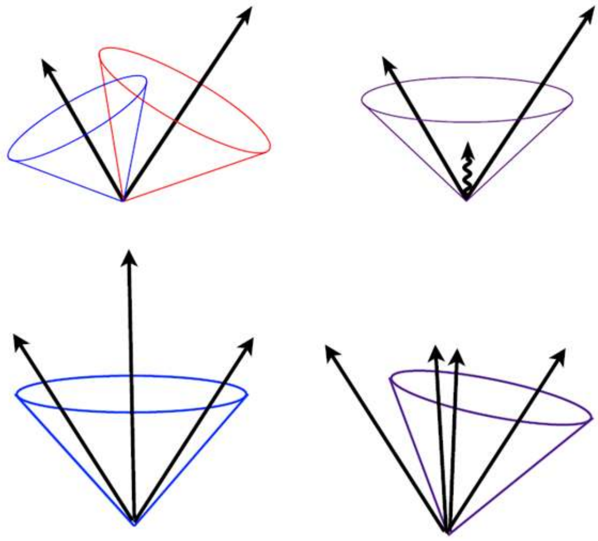

Figure 4.1. Configuration of infrared (top) and collinear (bottom) safe.

4.6.1. Iterative Cone (IC) Algorithm

IC is a simple cone-based jet clustering algorithm. It is used in the HLT system in CMS because it has a short execution time. In this algorithm, the seed particles with an energy greater than 1 GeV are defined sequentially. The basis of cone algorithm is based on finding the highest energy particle and taking all the particles remaining in the cone with a radius ΔR_{ij} and accepting it as a jet. ΔR_{ij} is calculated as follows:

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} < R \quad (14)$$

Here y_i and ϕ_i are the rapidity and azimuth angle of particle i, respectively. This calculation continues until a stable cone is found. The dimensionless parameter R is the radius of the jet.

When a stable cone is found, the jet is identified, and all its components are deleted from the remaining inputs. Although the algorithm does not have infrared and collinear safety, it is used for cones with a radius $R = 0.5$ in CMS. Figure 4.2 shows the operating configuration of the IC.

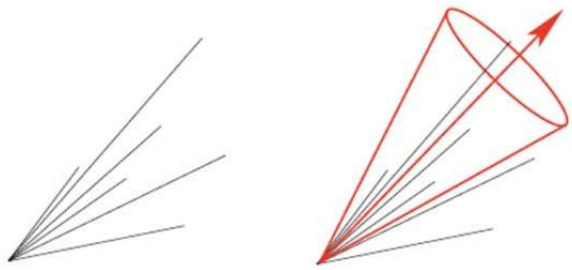

Figure 4.2. IC operating configuration.

4.6.2. Seedless Infrared Stable Cone (SISCone) Algorithm

Algorithms such as IC consider all particles in an event as seeds and look for stable cones. When a light particle is inserted between two heavy particles, this algorithm approaches that particle like a seed and starts looking for a third cone. This issue is known as infrared insecurity. To solve this problem in IC, a seedless search on all particles is proposed and this proposal is named as Midpoint Cone Algorithm (Blazey, 2000). According to this proposal, after a stable cone with a true seed particle is found, a false seed is inserted between the two stable cones and a new cone is sought. But this process takes approximately $N \times 2^N$ (10^{17} years for 100 particles) to find the jets among N particles. This time problem is solved by moving the process to two dimensions. In one dimension, there is only one degree of freedom (y or η) to determine the position of a cone. In two dimensions, there are two degrees of freedom (y and ϕ) to determine the position of the cone containing a pair of particles. Figure 4.3 illustrates this approach.

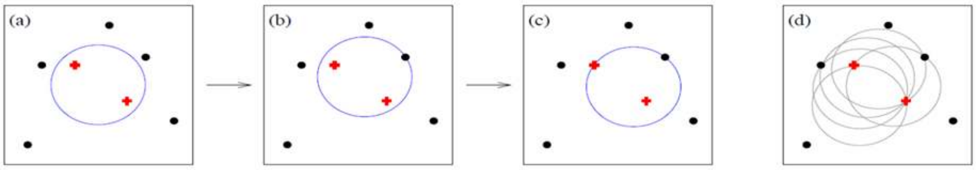

Figure 4.3. Determining the position of the circle containing a particle pair in two dimensions (Salam, 2007).

Referring to Figure 4.3, two starting points are first selected (a). The circle is rotated in different directions until one of these starting points comes into contact with circle (b). When the first point touches the circle, the circle is fixed at that point and continues to rotate around that point until the second point touches (c). Finally, when both points touch the circle, all other circles where both points touch are determined (d). This implementation takes approximately $N^2 . ln\ 2$ times and provides acceptable infrared safety. The problem-free final version of the algorithm is called the SISCone algorithm. The SISCone algorithm is not recommended by the CMS due to its low CPU efficiency and inaccurate results in events involving PU.

4.6.3. The anti-k_T Algorithm

The anti-k_T algorithm is both infrared and collinear secure and is used as the default jet clustering algorithm in CMS. The algorithm is a special case of the k_T algorithm. The k_T algorithm is based on pairwise recombination. In this algorithm, if the momentum of these paired particles is less than the value defined in the algorithm, these two particles are combined. The distance d_{ij} between particles i and j and the distance d_{iB} between the beam and particle i are defined as:

$$d_{ij} = min(p_{Ti}^2, p_{Tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{15}$$

$$d_{iB} = p_{Ti}^2 \tag{16}$$

where R is the radius of the cone. The k_T algorithm works by following the steps below:

- A list of all particles is made.
- d_{ij} and d_{iB} are calculated.
- If d_{ij} is the smallest, a new particle is created by combining i and j particles and returns to the first step.
- If d_{iB} is smallest, thread i is deleted from the list and returns to step first.
- These processes are continued until no particles remain.

Distance calculations can be generalized as follows:

$$d_{ij} = min(p_{Ti}^{2p}, p_{Tj}^{2p})\frac{\Delta R_{ij}^2}{R^2} \tag{17}$$

$$d_{iB} = p_{Ti}^{2p} \tag{18}$$

where p parameter is taken as 1, k_T algorithm is used, and “soft” particles are clustered first. If the parameter p is taken as -1, the anti-k_T algorithm is used, and the "hard" particles are clustered first. The anti-k_T algorithm gives the best jet shapes and is recommended by CMS to use in jet physics analysis. Figure 4.4 shows the behavior of different jet clustering algorithms.

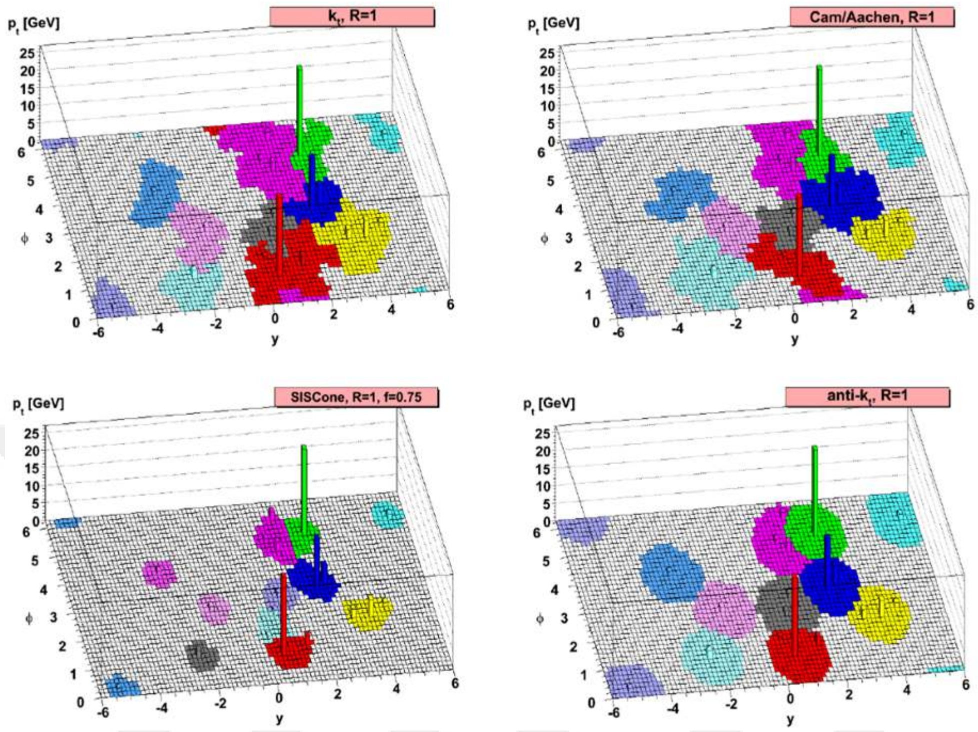

Figure 4.4. The behavior of different jet clustering algorithms (Salam, 2010).

5. CROSS SECTION MEASUREMENTS

Measurements using pp collisions that occur in hadron accelerators such as the LHC are sensitive to the internal structure of the proton, which is described by PDFs. PDFs are one of the biggest sources of uncertainty and reduce the sensitivity of measurements made in the analysis. Therefore, in particle physics, more precise measurements of PDFs play a particularly key role. A promising approach to improve constraints in PDFs consists of comparing measurements with fixed-order theory predictions. Jet cross-sections are particularly suitable for this purpose due to abundant jet production over a large phase area in hadron colliders. This is further supported by theory predictions at NNLO in pQCD.

Inclusive jets and dijets in pp collisions have been extensively measured by both the ATLAS and CMS Collaborations at several center-of-mass energies. The analysis conducted by CMS (CMS Collaboration, 2011) corresponds to an integrated luminosity of 5.0 fb^{-1}. The measurements cover the jet p_T range from 0.1 TeV to 2 TeV, and the dijet-mass range from 0.3 TeV to 5 TeV in five rapidity bins up to $|y| = 2.5$. The measured cross-sections agree with the predictions of pQCD at the NLO obtained with five different PDF sets. The analysis conducted by ATLAS (ATLAS Collaboration, 2011a) corresponds to an integrated luminosity of 17 nb^{-1}. The cross sections are extended into previously unmeasured kinematic regimes. For inclusive jets, the double differential cross section has been measured for jets with $|y| < 2.8$ and $p_T > 60$ GeV. The p_T distribution extends up to 600 GeV. For dijet events, containing a jet with $p_T > 30$ GeV in the same rapidity region, the cross-section has been measured as a function of the dijet mass. The dijet mass distribution is extended up to nearly 2 TeV. For both inclusive and dijet measurements, the theory agrees well with the data, validating the pQCD approach in a new kinematic regime.

Dijet mass cross-section measurement has been performed recently using $35.9\ fb^{-1}$ p-p collision data from the CMS detector on the LHC at $\sqrt{s} = 13$ TeV center of mass energy in 2016. Cross-sections of the dijet jet mass were extended to 7 TeV. Theory and data comparisons were made using the LO, NLO, and NNLO calculations of the pQCD theory. The CT14 PDF set was used to reweight the theory cross sections (Kiminsu, U., 2022).

The analysis presented here includes an extended version of the analysis conducted by Kiminsu et al. and concerns measurements of the double differential inclusive dijet production cross sections as a function of the invariant dijet mass of the two leading jets in the event. The key differences from the previous analysis are as follows: Effective trigger ranges (trigger turn-ons) are calculated according to inclusive jet p_T instead of invariant dijet mass. The 2D unfolding technique is used and the mitigations between adjacent mass bins as well as adjacent rapidity bins are considered to correct data more precisely. Theory prediction calculations have been expanded to include more PDF sets and the QCD analysis, which will be discussed in the next section, has been performed.

The covered phase space is divided into five regions of the maximum rapidity of the dijet system $|y|_{max} = max(|y_1|, |y_2|)$ up to 2.5 and for dijet mass $m_{1,2} = \sqrt{(E_1 + E_2)^2 + (\vec{p}_1 + \vec{p}_2)^2}$ up to 10 TeV, where subscripts 1 and 2 represent the leading and the subleading jets in the event respectively. The data were collected, in 2016, by CMS at the LHC in p-p collisions at $\sqrt{s} = 13$ TeV and correspond to an integrated luminosity of 36.3 fb^{-1}. The jets were reconstructed using the anti-k_T clustering algorithm (M. Cacciari, G. P. Salam, and G. Soyez, 2008), with the distance parameter R = 0.4. The data cross-sections are corrected for detector effects, while the final results are compared with theoretical calculations at NNLO derived from pQCD.

A thorough explanation of the analysis process is given in the following sections. The reconstruction of jets and event selection are covered in section 5.2

after a brief introduction to the relevant data and simulation samples (section 5.1). The explanation of the trigger studies and the method for combining data based on trigger selections to cover the whole phase space is covered in section 5.3. The measurement is carried out as explained in section 5.4. Then, a two-dimensional unfolding method is applied to the measured cross-sections (Section 5.5). Section 5.6 presents the primary sources of uncertainty that need to be taken into account. After unfolding, detector-independent estimates are compared to fixed-order theory predictions. (Section 5.7).

5.1. Data and MC Simulation Samples

The studies presented in this thesis are based on p-p collision data recorded by the CMS detector in 2016 at a center-of-mass energy of 13 TeV. Integrated luminosity for the entire data set is 36.3 fb^{-1}. Luminosity blocks, which correspond to a data-taking period of 25 nanoseconds, are used to further partition runs. After data collection, a procedure known as data certification is used to confirm that all detector components are in good functioning condition. This process verifies each run and luminosity block. Only verified runs and luminosity blocks supplied by the CMS Collaboration centrally are used in this analysis. To better understand the impact of shifting detector conditions, data gathered over longer time scales are separated into multiple data-taking periods (frequently referred to as eras). In Table 5.1, there are seven such run periods, denoted by the letters B to H, each of which represents a different phase of the 2016 data collection.

The GEANT4 package (GEANT4 Collaboration, 2003) is used to achieve a realistic detector simulation. Since there is a finite detector resolution for the dijet mass measurement, generator results are used for the resolution studies. Furthermore, the information provided by MC event generators is used to unfold the experimental results from the detector level to the stable particle level and perform a comparison between them. 2016 Moriond17 PU campaign MC samples for the Pythia 8 generator (T. Sjostrand, S. Mrenna, and P. Z. Skands, 2008) with CUETM1 tune (CMS Collaboration, 2016a) are used. There are 15 different slices with respect to the generator variable $\hat{p}_T$, ranging from 15 GeV and extending up to infinity. Table 5.2 shows the Pythia slices along with their respective luminosity values. To compare the results coming from Pythia and perform tests to determine the need for an additional systematic uncertainty due to the model dependence of the algorithms inside MC event generators, the MadGraph (J. Alwall et al., 2014.) sample is also used. The difference between the MadGraph sample to the standalone Pythia is that Pythia is only used for the generation of the ME, while MadGraph combined with Pythia is used to simulate the NP processes like PS, HAD, and multi-parton interactions (MPI). MadGraph slices with respect to H_T are presented in Table 5.3. The corresponding DAS names of the samples are:

- Pythia sliced: /QCD_Pt_*to*_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv2PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6*/MINIAODSIM.
- MadGraph + Pythia sliced: /QCD_HT*to*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM.

Table 5.1. Data samples.

Run Period	DAS Name	Integrated Luminosity (fb^{-1})
B	$/JetHT/Run2016B-17Jul2018-ver2-v1/MINIAOD$	5.8
C	$/JetHT/Run2016C-17Jul2018-v1/MINIAOD$	2.6
D	$/JetHT/Run2016D-17Jul2018-v1/MINIAOD$	4.3
E	$/JetHT/Run2016E-17Jul2018-v1/MINIAOD$	4.1
F	$/JetHT/Run2016F-17Jul2018-v1/MINIAOD$	3.1
G	$/JetHT/Run2016G-17Jul2018-v1/MINIAOD$	7.7
H	$/JetHT/Run2016H-17Jul2018-v1/MINIAOD$	8.7

Table 5.2. Pythia $\hat{p}_T$ slices.

$\hat{\mathbf{p}}_{\mathbf{T}}$ **slices**	$\mathcal{L}$ $(\mathbf{pb^{-1}})$
30 → 50	0.07081
50 → 80	0.5183
80 → 120	5.126
120 → 170	26.78
170 → 300	126.2
300 → 470	2,872
470 → 600	6,109
600 → 800	72,360
800 → 1000	609,900
1000 → 1400	1,045,000
1400 → 1800	3,410,000
1800 → 2400	17,240,000
2400 → 3200	145,900,000
3200 → Inf	2,368,000,000

Table 5.3. MadGraph H_T slices.

$\boldsymbol{H}_{\mathbf{T}}$ **slices**	$\mathcal{L}$ $(\mathbf{pb^{-1}})$
50 → 100	0.01719
100 → 200	2.988
200 → 300	33.54
300 → 500	155.3
500 → 700	1,980
700 → 1000	2,298
1000 → 1500	12,610
1500 → 2000	98,330
2000 → Inf	238,400

5.2. Jet Reconstruction and Event Selection

Particles produced by high-energy proton collisions traverse the CMS detector, creating distinct signals in the many subdetectors. Particles created in a collision event are reconstructed at CMS via the PF algorithm (CMS Collaboration, 2017c), which combines information from multiple subdetectors. This results in several so-called PF candidates from multiple subdetectors.

5.2.1. Jet Reconstruction

PF candidates consist of reconstructed four-vectors associated with the particles as well as various other information depending on the particle type. The PF technique reconstructs and identifies each particle by combining information from all the subdetectors. The energy of the photons is derived directly from the ECAL measurements. The energy of the electrons is determined from a combination of the corresponding ECAL cluster energy of the track momentum at the main interaction vertex and the energy sum of all bremsstrahlung photons attached to the tracks. The energy of muons is derived from the corresponding track momentum. The energy of the charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energy and is calibrated for the non-linear response of the calorimeters. Finally, the energy of the neutral hadrons is derived from the corresponding calibrated ECAL and HCAL energy. The primary interaction peak is defined as the highest sum of the squared transverse momenta of the charged tracks associated with it.

Jets are then reconstructed by running a jet clustering algorithm on the PF candidate collection. The anti-k_T algorithm performs the clustering of the candidate particles by combining them into jets with a given distance parameter R. Currently, there are two official cone sizes suggested by CMS for the MINIAOD data sample, $\mathrm{R} = 0.4$, and $\mathrm{R} = 0.8$.

5.2.2. Charged Hadron Subtraction (CHS) Method

One of the outstanding challenges of the LHC is the huge amount of instantaneous luminosity, which will result in many additional p-p collisions in each event called PU. In such a high PU environment, the accurate reconstruction of jets will become increasingly important.

CHS is a technique used to reduce the effect of in-time PU (originating from the same bunch crossing in the event) on reconstructed physics objects. In this approach, charged hadrons unambiguously associated with PU vertices are removed from the event and the remaining PF candidates are allowed to cluster to form jets.

5.2.3. Jet Energy Corrections (JECs)

Accurate measurement of jet energy is an essential part of jet analyses. To correct the jet energy for various sources of systematic bias that affect the reconstruction, a set of JECs is derived (CMS Collaboration, 2017a). Figure 5.1 shows schematically the order in which each different correction is applied to the jets. Each correction that comes in the form of a multiplicative factor is applied separately and treats both data and MC samples. The data corrections are represented on the upper side of Figure 5.1, while the MC corrections are represented on the lower side. Jets contained inside the data samples are referred to, as PFCHSJets which are jets reconstructed by the PF and anti-k_T algorithm and have the impact of detector-related effects. On the other hand, jets contained in MC samples are referred to as GenJets, which are jets generated by MC simulation and reconstructed by the anti-k_T clustering algorithm therefore not having the impact of the detector effects. Apart from the CHS PU corrections, additional PU corrections are applied concerning neutral particles which do not originate from the hard scattering and are related to MPIs that contaminate the energy measurement of the event. A series of corrections are applied next for the non-linear and non-uniform response of the detector to transverse momentum (p_T) and eta (η). Table 5.4 shows a summary of the jet selection criteria. The newest version of JECs is applied as recommended by the CMS JetMET group (twiki.cern.ch/twiki/bin/viewauth/CMS/ JetMET).

5.2.4. Jet Energy Resolution (JER)

The detector has a finite JER. Consequently, each time the detector measures the energy of a jet, there is a deviation between the value of the recorded energy and the true energy of the jet. The better the resolution of the measurement, the lower the deviation between the two values. Simulated samples are smeared from the generator level to obtain the reconstructed level, thus containing the full simulation of the detector. The smear process on simulated samples is performed with an optimistic value than the real detector resolution. Therefore, an additional correction must be applied to the simulated samples to correct the resolution contained in the MC to resemble the one found in the data. A brief description is given in the following subsections to describe the procedure used in the resolution studies.

Table 5.4. Jet selection criteria and JEC and JER versions applied to the analysis.

Leading jet selection	$p_{T_1} \geq 100$ GeV, $\lvert y_1 \rvert \leq 2.5$
Sub-leading jets selection	$p_{T_2} \geq 50$ GeV, $\lvert y_2 \rvert \leq 2.5$
JEC version	$Summer16\_07Aug2017\_V11$
JER version	$Summer16\_25nsV1$
Jet ID	$TightLepVeto$

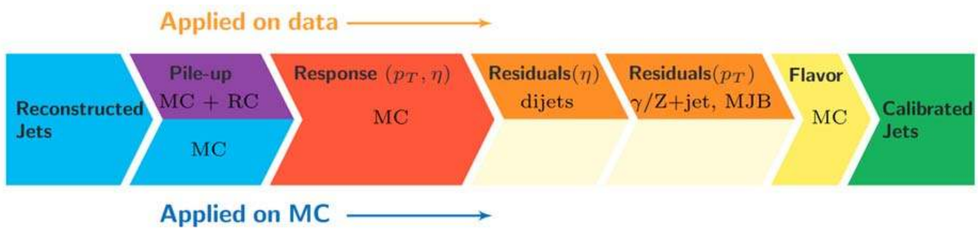

Figure 5.1. Various stages of JEC for PFCHSJets.

5.2.4.1. Response Distributions

The phase space is divided according to three reconstructed level quantities, the jet transverse momentum p_T^{rec}, the jet absolute pseudo-rapidity $|\eta^{rec}|$ and the soft activity ρ, a quantity related to pileup. The bin edges can be found in Table 5.5. Each ρ bin is divided into 19 $|\eta^{rec}|$ bins, each of which is then subdivided into 62 p_T^{rec} bins.

Table 5.5. List of binning schemes used in the resolution studies, in terms of p_T^{rec}, $|\eta^{rec}|$ and ρ.

p_T^{rec} (GeV)	$\lvert\eta^{rec}\rvert$	ρ
15, 18, 21, 24, 28	0.000, 0.261	0,
32, 37, 43, 49, 56	0.522, 0.783	6.69,
64, 74, 84, 97, 114,	1.044, 1.305,	12.39,
133, 153, 174, 196, 220,	1.566, 1.740,	18.09,
245, 272, 300, 330, 362,	1.930, 2.043,	23.79,
395, 430, 468, 507, 548,	2.172, 2.322,	29.49,
592, 638, 686, 737, 790,	2.500, 2.650,	35.19,
846, 905, 967, 1032, 1101,	2.853, 2.964,	40.9,
1172, 1248, 1327, 1410, 1497,	3.139, 3.489,	∞
1588, 1684, 1784, 1890, 2000,	3.839, 5.191,	–
2116, 2238, 2366, 2500, 2640,	–	–
2787, 2941, 3103, 3273, 3450,	–	–
3637, 3832, 4037	–	–

The response is then calculated according to:

$$\Delta = \frac{p_T^{rec} - p_T^{gen}}{p_T^{gen}} \quad (19)$$

The resulting distributions are then fitted by a Double Crystal-Ball (DCB) function. The DCB is a very good candidate for the fit procedure since it combines a gaussian function to parametrize the core of the response distributions with an exponential form to parametrize the tails. Figure 5.2 shows the fits for a central ρ bin ($18.09 < \rho < 23.79$). The fit aims to extract the width of the response distribution which represents the desired resolution.

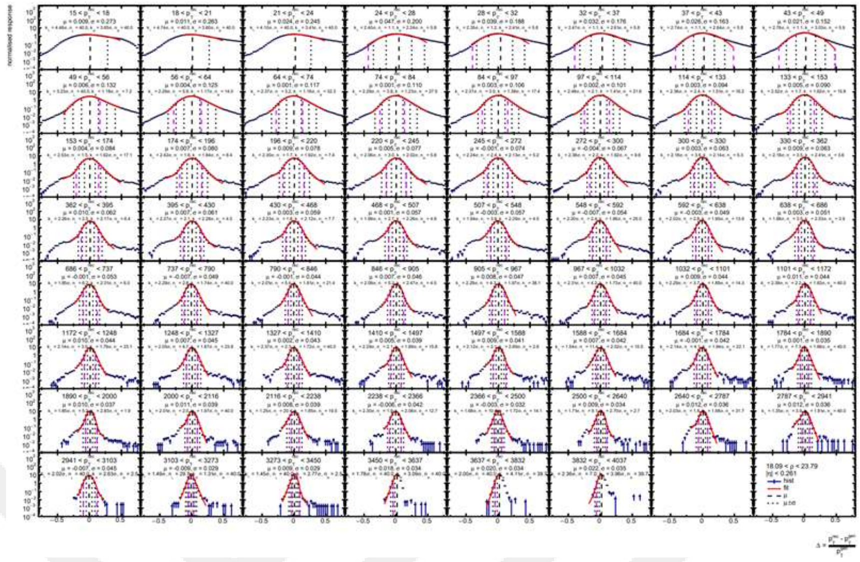

Figure 5.2. Response distributions across the whole p_T^{rec} range in slice $18.09 < \rho < 23.79$, $|\eta^{rec}| < 0.261$.

5.2.4.2. JER Fits

A fit on the resolution is performed for each ρ and $|\eta^{rec}|$ bins. Some examples of the fit details are given in Figure 5.3 for the same central ρ bin ($18.09 < \rho < 23.79$). Each cell in Figure 5.3 corresponds to a different $|\eta^{rec}|$ bin and covers the whole p_T^{rec} range. The blue points illustrate the width of the response distributions as extracted from Figure 5.2, while the fit of the resolution is displayed by the red curve. It can be observed that the resolution takes lower values towards higher energies and becomes better. The standard "NSC" (Noise, Stochastic, Constant) function shown in Equation 20 is used to fit the resolution which is the current recommendation given by the JetMET group.

$$JER = \sqrt{p_0 \frac{|p_0|}{p_T^2} + \frac{p_1^2}{p_T^3} + p_2^2} \tag{20}$$

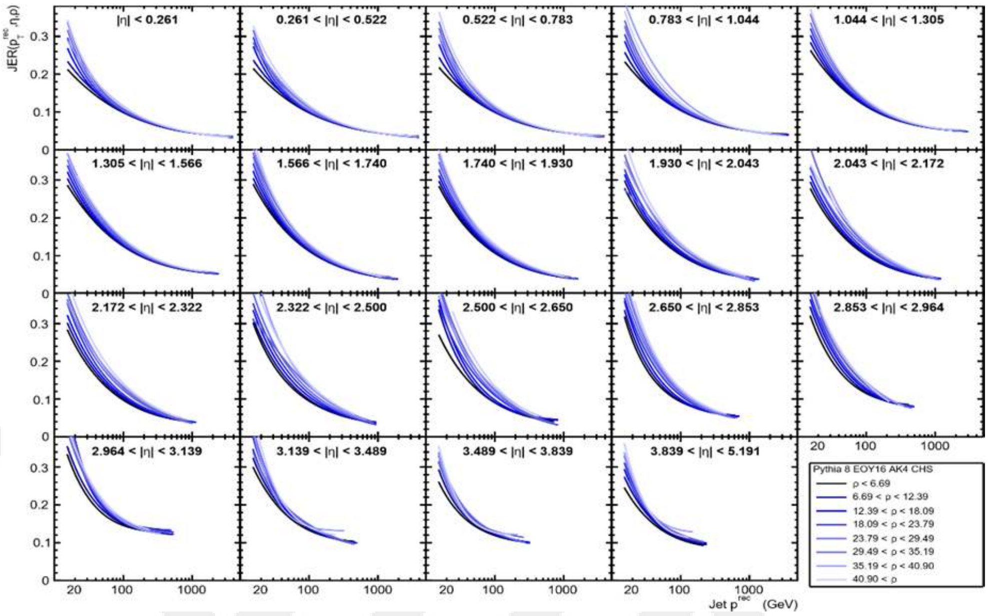

Figure 5.3. The JER as a function of the transverse momentum.

5.2.5. JER Smearing

When making use of simulated samples, care must be taken to ensure that both the detector response and the resolution is modeled accurately to reflect the actual measurement conditions. Since these can only be evaluated reliably until after data taking, simulated samples typically exhibit a much better (i.e., smaller) JER than the measurements. There are two official approaches developed by JetMET, the scaling method, and the stochastic method, to smear the resolution in simulated samples. With the scaling method, the corrected four-momentum of a reconstructed jet is rescaled with a factor derived as follows:

$$c_{JER} = 1 + (s_{JER} - 1)\frac{p_T - p_T^{ptcl}}{p_T} \quad (21)$$

where p_T is its transverse momentum, p_T^{ptcl} is the transverse momentum of the corresponding jet clustered from generator-level particles, and s_{JER} is the data-to-simulation core resolution scale factor. This method only works if a well-matched particle level jet is present and can result in a large shift of the response otherwise. The following requirements are imposed for the matching:

$$\Delta R < R_{cone}/2\,, \left|p_T - p_T^{ptcl}\right| < 3\sigma_{JER}p_T \qquad (22)$$

where R_{cone} is the jet cone size parameter (for instance, 0.4 for AK4 jets) and σ_{JER} is the relative p_T resolution as measured in simulation.

Stochastic smearing, which does not require the presence of a matching particle level jet, is an alternative to the scaling method. In this case, corrected jet four-momentum is rescaled with a factor derived as follows:

$$c_{JER} = 1 + \mathcal{N}(0, \sigma_{JER})\sqrt{max(s_{JER}^2 - 1, 0)} \qquad (23)$$

where σ_{JER} and s_{JER} are the relative p_T resolution in simulation and data-to-simulation scale factors, and $\mathcal{N}(0, \sigma_{JER})$ denotes a random number sampled from a nominal distribution with a zero mean and variance σ^2.

A combination of the two approaches is recommended by the JetMET group. Comparisons between different smearing options are done for both the dijet mass spectrum, shown in Figure 5.4, and the inclusive jet spectrum, shown in Figure 5.5. The hybrid method with a fixed core at one sigma is used in the analysis since it displays the best agreement with the stochastic approach.

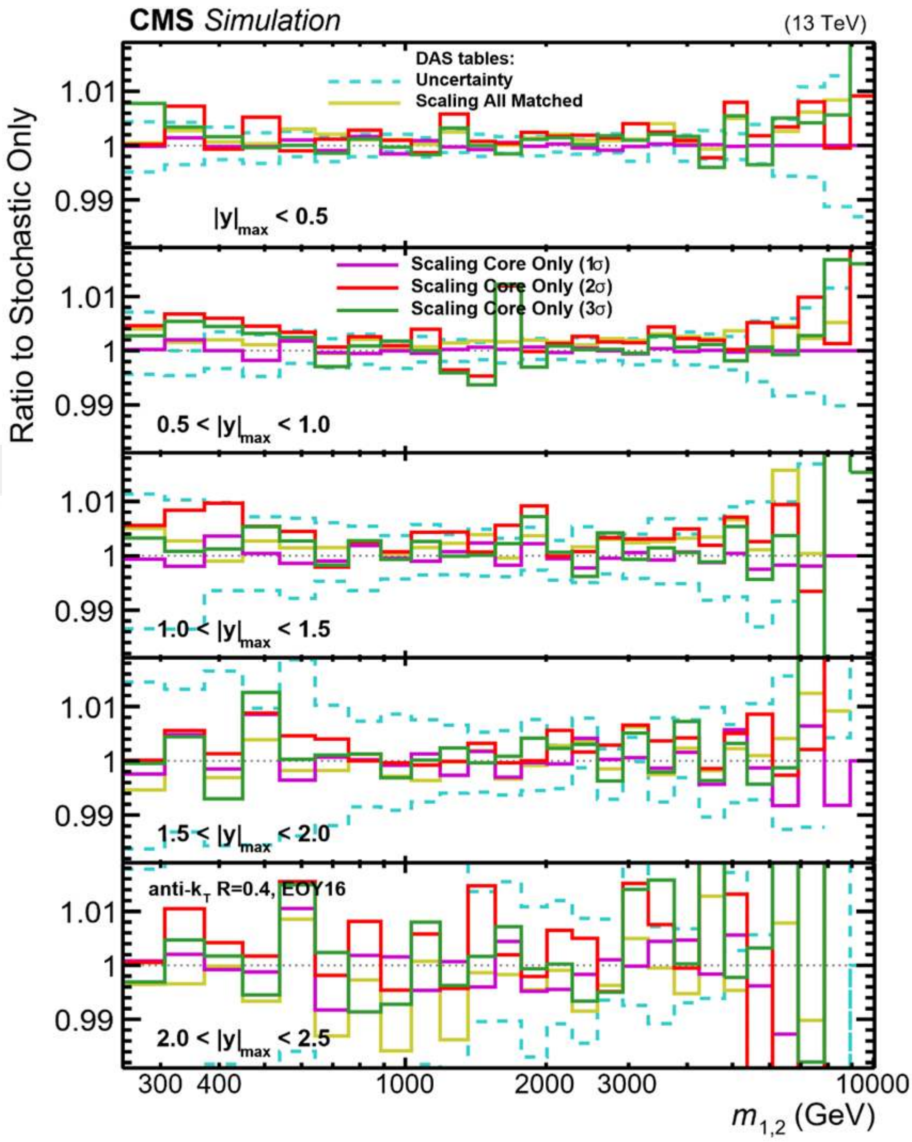

Figure 5.4. The ratio of cross-sections for the dijet mass spectrum, after smearing. Each curve represents a comparison between different smearing methods. Methods are compared to the stochastic method (denominator) by changing the numerator. Each cell in the plot corresponds to a different rapidity bin.

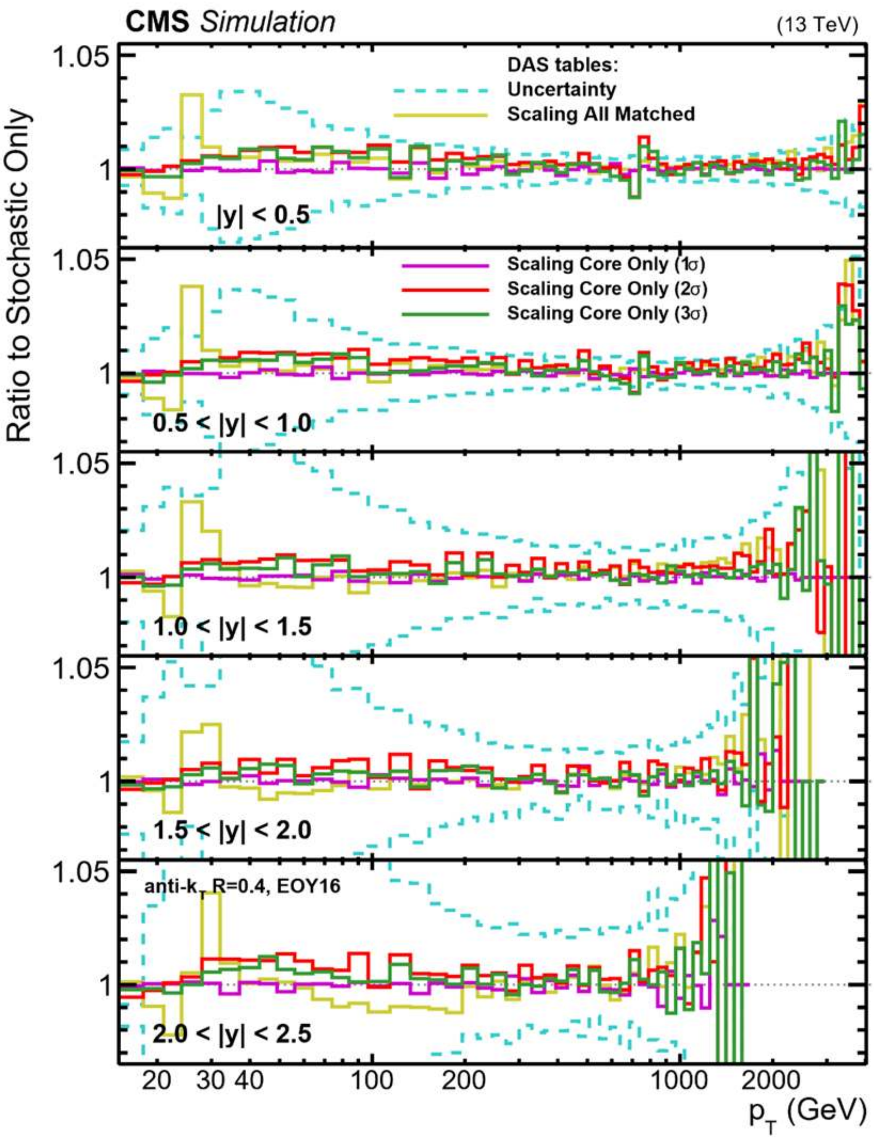

Figure 5.5. The ratio of cross-sections for the inclusive jet spectrum, after smearing. Each curve represents a comparison between different smearing methods. Methods are compared to the stochastic method (denominator) by changing the numerator. Each cell in the plot corresponds to a different rapidity bin.

5.2.6. PU Profile Reweighting

Another correction that should be considered for the simulated samples is PU profile reweighting. Two approaches are widely used for reweighting. These are the trigger-dependent method (per trigger PU profile reweighting) and the trigger-independent method (global PU profile reweighting). In the trigger-dependent method, PU profiles are created separately for each trigger path. In the trigger-independent method, a single profile is created for all trigger paths. Trigger-dependent method can introduce biases due to the specific trigger requirements used, while trigger-independent method is less susceptible to such biases but can have larger statistical uncertainties. The choice of method depends on the specific analysis and the available data and simulation samples. The results of both methods can be seen for the dijet mass spectrum in Figure. 5.6 and the inclusive jet spectrum in Figure 5.7. In the figures, the red color represents the comparison between the calibrated cross-section using the trigger-independent method and the raw cross-section, while in blue color represents the same comparison but this time using the trigger-dependent method. The cyan color dashed band indicates the uncertainty coming from the variation of the trigger-dependent method. The trigger-dependent approach is followed for the correction of the simulated samples.

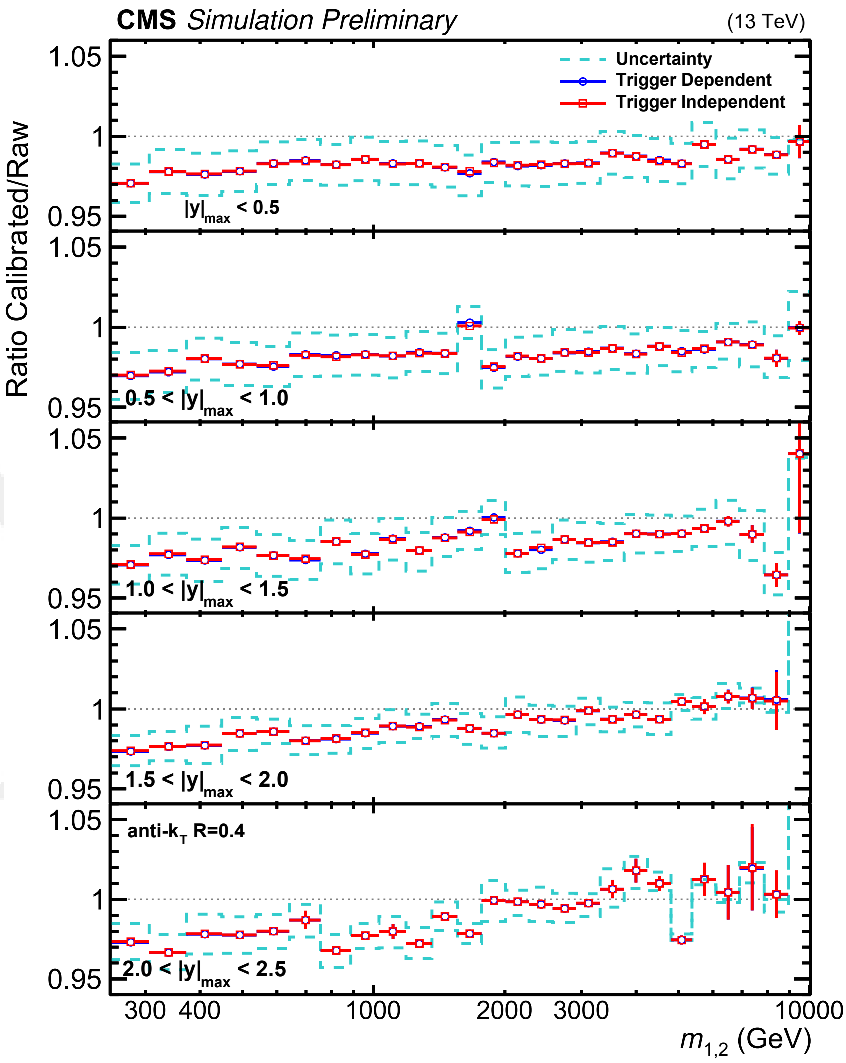

Figure 5.6. The ratio between the calibrated (corrected) and the raw (uncorrected) simulated data with respect to the dijet mass.

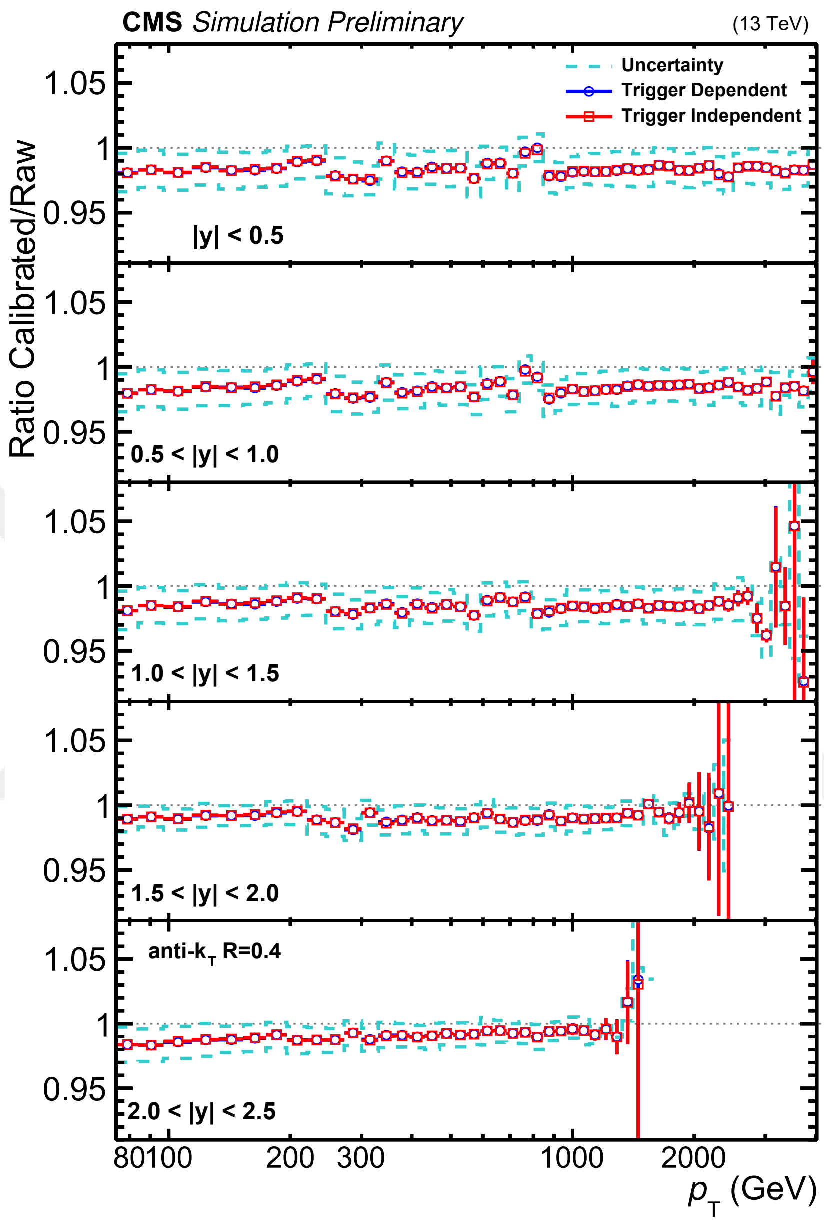

Figure 5.7. The ratio between the calibrated (corrected) and the raw (uncorrected) simulated data with respect to the inclusive jet spectrum.

5.2.7. L1 ECAL Prefiring Correction

Another correction applied to the data during the offline analysis is related to an issue spotted at the L1 trigger ECAL primitives. The L1 ECAL primitives were prematurely firing (prefiring) due to a misadjusted time offset which was originally introduced to negate the already existing effect of ECAL. The problem originates from the high radiation environment inside CMS, which causes ECAL crystals to be prone to radiation damage. For this reason, calibration over time is needed to mitigate the prefiring effect.

At the end of the 2016 data taking the prefiring issue caused some jets in the forward region ($2.0 < |\eta| < 3.0$) to be wrongly considered by the L1 trigger as belonging to the previous bunch crossing. As a result, an additional inefficiency is introduced in the measurement which should be treated. There are two main approaches, either to apply new weights in the simulation samples to reproduce the prefiring effect or to apply the inverted weights in the data samples to correct the effect. In the context of this analysis, the second approach is followed, while the weights for the jets are taken by maps with respect to p_T and η as provided by JetMET. Furthermore, a basic systematic uncertainty equal to 20% of the prefiring rate is considered.

JetMET provides two kinds of correction tables. On one side there are dedicated maps for each run period (era), called maps per era, while on the other side, there are maps averaged over all eras, called average maps. Upon use, both cases carry advantages and disadvantages over one another. The main advantage of maps per era is that they account for the time dependence of the prefiring effect but contain fewer statistics since they are split among the different eras, and therefore are provided in a coarser binning scheme leading to a loss of resolution. On the contrary, average maps contain more statistics and are provided in a finer binning scheme. Figure 5.8 displays the ratio between the data before applying the prefiring correction (raw) over the data after applying the correction (calibrated). Red points illustrate the use of maps per era while blue points illustrate the use of average maps. The cyan dashed line indicates the resulting uncertainty after varying the nominal value of average maps. The results are compatible if uncertainties for both cases are considered. In the context of this analysis, the average maps are used to compensate for the loss of efficiency caused by the prefiring effect.

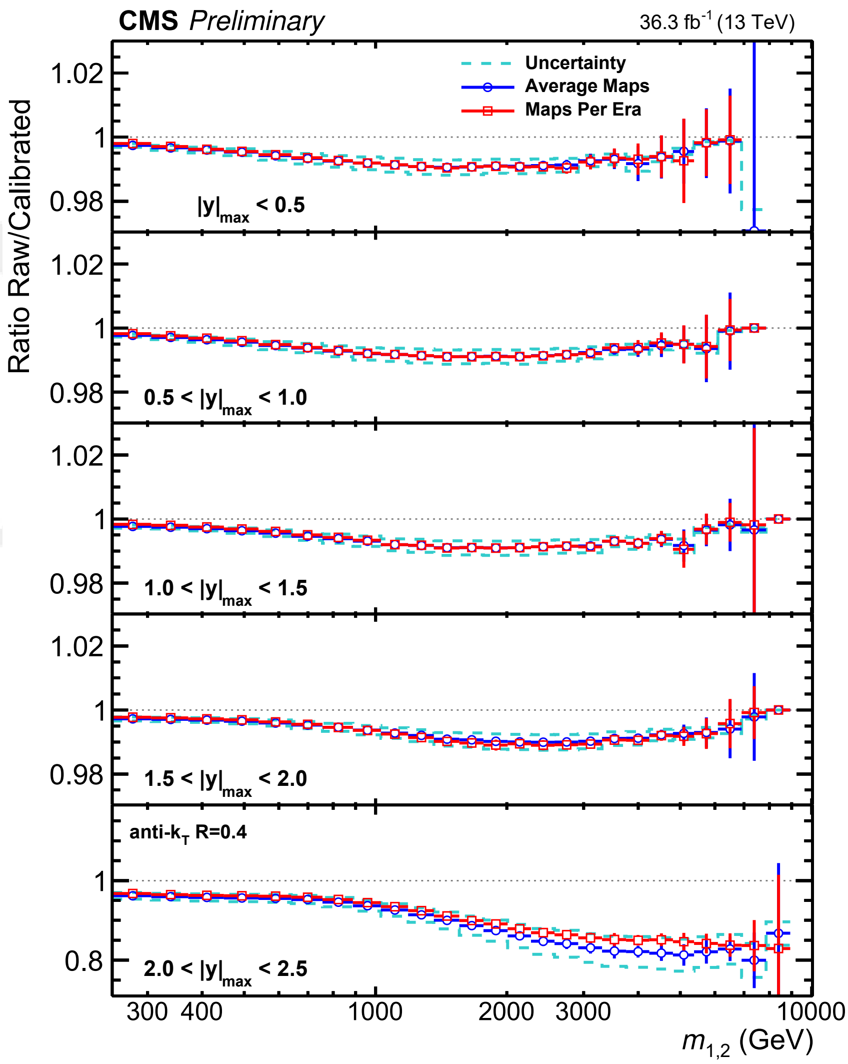

Figure 5.8. The ratio between the raw (uncorrected) and the calibrated (corrected) experimental data.

5.3. Trigger Studies

The trigger system is divided into two parts: a hardware-driven Level 1 Trigger (L1) and a software-driven HLT. The aim of these triggers in the CMS data collecting process is to detect events that are valuable for scenarios such as physics analysis or calibration. The event rate was reduced from the usual 40 MHz at which proton bunch collisions occur to 100 kHz after the L1 trigger, and eventually to 1 kHz after the HLT by just recording out these events and rejecting the rest. Object reconstruction methods are used on the trigger level to identify events of interest. To guarantee that the processing is rapid enough, this reconstruction is simplified and simply provides an approximation of observable quantities.

5.3.1. L1 Trigger

At the L1 stage, so-called trigger primitives are reconstructed by performing a coarse-grained readout of the detector, to identify the detector regions in which significant activity occurs. This is followed by the reconstruction of trigger objects, which represent potential particles or jet candidates. The trigger decision is then taken based on the kinematic properties of the trigger objects, such as transverse momentum p_T or η. The rate is reduced to 100 kHz after the L1 trigger.

5.3.2. HLT System

Events accepted by the L1 trigger are sent to the HLT. This trigger level is totally software-implemented and consists of a series of trigger paths, each of which provides a binary accept/reject response based on a sequence of reconstruction and filtering phases. The vast majority of HLT paths are designed to detect a physical object or collection of objects in a specific configuration or kinematic range. Most generally, this indicates a low p_T threshold, which means that only trigger objects with transverse momentum greater than this threshold qualify for an event to be acknowledged by the trigger path.

In this analysis the High-Level single PF Jet Triggers of the HLT trigger system are used, where each trigger has a different p_T threshold. This means that triggers record only events that contain at least one well-reconstructed jet and require the p_T of the leading jet of the event to be larger than the trigger threshold. The AK4 analysis was performed using the HLT_PFJet_X type of triggers where X stands for the value of the p_T threshold of triggers. Note that the online jet reconstruction, happening inside the trigger system, is a fast and simplified version when compared to the one performed during the offline analysis. This prior reconstruction is needed for a fast calculation of some basic jet-related quantities, such as the jet p_T, to help the triggers make the final decision.

5.3.3. Prescale Factor

Another property of the triggers is the prescale factor k, which answers the question of why different trigger thresholds are used in the CMS. The hypothetical case of using one trigger makes this trigger solely responsible for covering the hole phase space of the measurement. But in this way, it would be impossible to keep up with the high-rate production of events. For this reason, different triggers are used to cover different phase space ranges of the measurement. Also, because the cross-section spectrum is steeply falling with respect to the dijet mass, regions at lower energies have event rates larger than those for higher energy regions. So, the prescale factor k is introduced to maintain the rate reduction coming from lower energy regions. For example, if a trigger has a prescale factor $k = 1000$, this means that for every 1000 events only 1 will be recorded and the selected event is randomly chosen among these 1000 events. Table 5.6 contains the available trigger paths along with the corresponding effective luminosity recorded by each trigger. All triggers are prescaled, except for the last two triggers which are un-prescaled ($k = 1$), therefore recording all passing events. The highest prescale factor is assigned to the trigger with the lowest threshold while for the other triggers the prescale factor gets lower values in descending order. Since different triggers record different sub-ranges of the phase space, which information to extract must be carefully selected from triggers and the ranges during the offline analysis. Events from different triggers are then combined to reconstruct the final spectrum.

Table 5.6. Trigger paths and respective effective luminosities for AK4 analysis.

$Trigger\ Path$	$\mathcal{L}$ ($\mathbf{pb^{-1}}$)
HLT_PFJet_40	0.267036
HLT_PFJet_60	0.726258
HLT_PFJet_80	2.775889
HLT_PFJet_140	24.1946
HLT_PFJet_200	103.827
HLT_PFJet_260	593.764
HLT_PFJet_320	1,772.50
HLT_PFJet_400	5,194.33
HLT_PFJet_450	36,321.9
HLT_PFJet_500	36,321.9

5.3.4. Efficiency Curves

The purpose of the trigger studies is to calculate the efficiency curves for each trigger and their corresponding turn-on point. A turn-on point is defined as the point above which the efficiency of a trigger is higher than 99.5%. Using the turn-on points, it is possible to determine which trigger makes the maximum contribution to which range of the spectrum. The emulation method is followed for the actual calculation. For example, the efficiency curve of the HLT_PFJet_80 trigger can be calculated by following the steps given below:

1. Filling a histogram with all events recorded by the trigger with a lower p_T threshold (HLT_PFJet_60) and labeled as a reference.
2. Accessing L1 and HLT object of the reference histogram.

3. Comparing which of these objects fulfill the criteria to be recorded by the trigger of which its efficiency curve be calculated (HLT_PFJet_80).
4. Filling a second histogram with the events that have passed step 3. This histogram is called an emulated histogram.
5. Dividing the reference histogram with the emulation histogram.

After step 5, the efficiency curve of the HLT_PFJet_80 is obtained, and the mathematical representation is given below:

$$eff_{HLT\_PFJet80} = \frac{emulated_{HLT\_PFJet80}}{reference_{HLT\_PF\_Jet60}} \tag{24}$$

Since the purpose of the analysis is to measure the dijet production cross-section with respect to the invariant mass of the two leading jets in the event, it would be logical that the calculation of the turn-on points should also be with respect to the dijet mass. Since the HLT_PFJet_X has varying p_T thresholds for each trigger, the calculation of the turn-on points should be done with respect to the leading jet p_T. Consequently, the trigger study is conducted two times, once with respect to $m_{1,2}$ and once with respect to p_T.

5.3.4.1. Turn-on Point Calculations with Respect to Dijet Invariant Mass

Because the observable is sensitive to both the dijet invariant mass $m_{1,2}$ and the rapidity y, the turn-on point of each trigger are calculated for each different rapidity region. It should be noted that the efficiency curve of the HLT_PFJet_40 trigger cannot be calculated, since it has the minimum available p_T threshold and the emulation method requires a trigger with a lower threshold for the calculation. The turn-on points are found by performing a fit with an error function. While this is true for most of the cases, it can be observed that triggers HLT_PFJet_60 and HLT_PFJet_80 have slightly different behavior at lower energies. This is due to a lack of statistics, and the behavior of the error function for these triggers cannot be parametrized. For these cases, where the fit cannot be performed, turn-on points are defined as the first bin edge after which the efficiency of the trigger remains constant at 100%. Indeed, it is ensured that the triggers are 100% efficient inside the bins chosen to extract the information from. So, calculated values are mitigated to the next bin edge and that bin is used as the actual turn-on point. By following this strategy, slightly larger statistical uncertainties are introduced, but in return, unnecessary inefficiencies are avoided in the final measurement. Table 5.7 shows the migrated turn-on points, which are those used during the offline AK4 analysis. The values are presented in GeV. The turn-on points are not contained for the HLT_PFJet_500 trigger in Table 5.7 since it is not used.

Table 5.7. Mitigated turn-on points using the emulation method with respect to the dijet mass $m_{1,2}$. The values of the turn on points are in GeV.

Trigger Path	$(\mathbf{0.0, 0.5})$	$(\mathbf{0.5, 1.0})$	$(\mathbf{1.0, 1.5})$	$(\mathbf{1.5, 2.0})$	$(\mathbf{2.0, 2.5})$
HLT_PFJet_40	–	160	249	449	887
HLT_PFJet_60	160	200	306	539	1029
HLT_PFJet_80	200	249	449	756	1187
HLT_PFJet_140	372	449	756	1187	1769
HLT_PFJet_200	539	641	1029	1769	2915
HLT_PFJet_260	641	887	1361	2273	3754
HLT_PFJet_320	887	1029	1556	2915	4805
HLT_PFJet_400	1029	1361	2008	3306	5374
HLT_PFJet_450	1187	1556	2572	4244	6908

5.3.4.2. Turn-on Point Calculations with Respect to Leading Jet p_T

The turn-on points of all trigger paths and for each rapidity region are calculated by following the same strategy but this time with respect to the transverse momentum of the leading jet in the event.

The difference in the calculations with respect to dijet invariant mass comes from using only one set of turn-on points, which remains the same for all rapidity regions. This is done because p_T is a physical quantity that is not sensitive to the rapidity, therefore, the values of the turn-on points among different rapidity regions have small divergence given the same trigger path. Whereas the dijet invariant mass $m_{1,2}$ is sensitive to the rapidity causing slight inefficiencies. Therefore, turn-on points with respect to the leading jet p_T are used in this analysis. Figure 5.9 shows the trigger efficiency curves as a function of p_T. Each figure contains five plots corresponding to rapidity bins, where in each plot all 9 trigger efficiency curves are presented with a different marker style and color. The corresponding values for the turn-on points can be found in Table 5.8.

Table 5.8. Calculated turn-on points using the emulation method (tag&probe method for HLT_PFJet_40) with respect to the transverse momentum p_T. The same set of turn on points is used for all rapidity bins.

$\boldsymbol{Trigger\ Path}$	$\boldsymbol{Turn-on\ points}$
HLT_PFJet_40	74
HLT_PFJet_60	97
HLT_PFJet_80	133
HLT_PFJet_140	220
HLT_PFJet_200	272
HLT_PFJet_260	362
HLT_PFJet_320	430
HLT_PFJet_400	548
HLT_PFJet_450	592

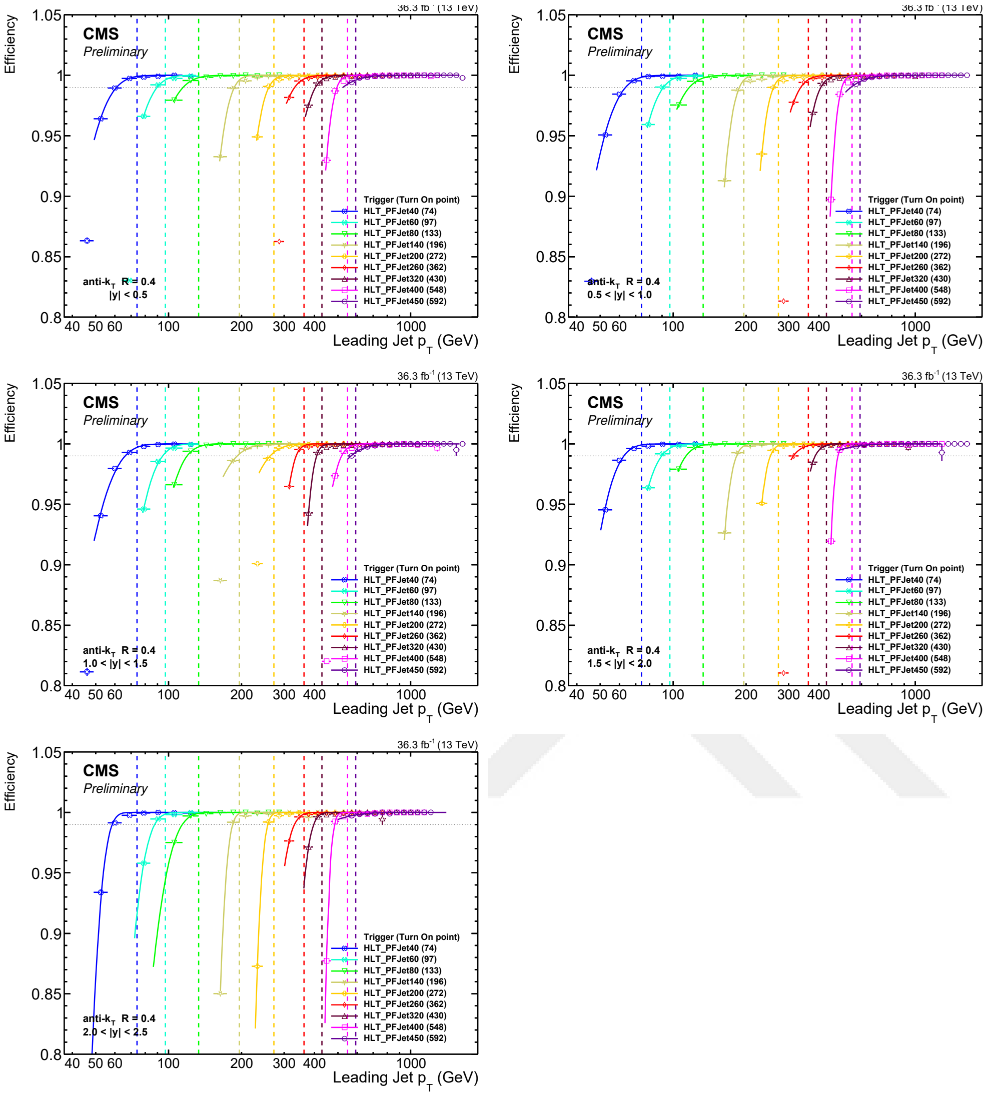

Figure 5.9. Trigger efficiency curves as a function of the leading jet p_T. Each plot corresponds to a different rapidity region and contains the efficiency curves of all 9 trigger paths represented with a different marker style and color. Lines are placed on top of the markers to visually differentiate each trigger threshold. Vertical dashed lines indicate the calculated turn-on point of each trigger, while the values are given in GeV inside the parenthesis of the legend.

5.4. Measurement at the Detector Level

Construction of the detector level spectrum can be done after obtaining the necessary information from the trigger studies. Table 5.9 shows the binning scheme used for AK4 analysis. The same binning, as presented in the table, is used for all five rapidity bins.

Table 5.9. Detector level $m_{1,2}$ binning scheme for AK4 analysis. The same binning scheme is used for all five rapidity bins.

Binning scheme (**GeV**)		
$[160-200\rangle$	$[200-249\rangle$	$[259-306\rangle$
$[306-372\rangle$	$[372-449\rangle$	$[449-539\rangle$
$[539-641\rangle$	$[641-756\rangle$	$[756-887\rangle$
$[887-1029\rangle$	$[1029-1187\rangle$	$[1187-1361\rangle$
$[1361-1556\rangle$	$[1556-1769\rangle$	$[1769-2208\rangle$
$[2208-2273\rangle$	$[2273-2572\rangle$	$[2572-2915\rangle$
$[2915-3306\rangle$	$[3306-3754\rangle$	$[3754-4244\rangle$
$[4244-4805\rangle$	$[4805-5374\rangle$	$[5374-6094\rangle$
$[6094-6908\rangle$	$[6908-7861\rangle$	$[7861-8929\rangle$
$[8929-10050\rangle$		

The spectra are reconstructed as a function of the dijet invariant mass $m_{1,2}$ in five rapidity regions defined in terms of the variable $|y|_{max} = |y_{max}|$, where y_{max} corresponds to the rapidity of the outermost jet, as given by:

$$y_{max} = sign(|\max(y_1, y_2)| - |\min(y_1, y_2)|)\max(|y_1|, |y_2|) \qquad (25)$$

The double differential cross-section spectra are reconstructed from the event yield N in bins of the chosen observables, normalized to the integrated luminosity L_{int}, and accounting for the total selection efficiency ε. In the phase space under study, the ε is determined almost entirely by the trigger efficiency, which remains above 99.5%. Events are weighted according to the prescaling factor of the trigger path assigned to the corresponding phase space region. The double differential cross section is obtained as a function of y_{max} and $m_{1,2}$ according to:

$$\frac{d^2\sigma}{dy_{max}dm_{1,2}} = \frac{1}{\varepsilon L_{int}} \cdot \frac{N}{(2 \cdot \Delta|y|_{max})\Delta m_{1,2}} \quad (26)$$

here, $\Delta|y|_{max}$ and $\Delta m_{1,2}$ denote the width of the bins in the respective quantities. The measurement is performed in five rapidity bins of equal size within $0 < |y|_{max} < 2.5$ and covers an overall invariant mass range of $249 < m_{1,2} < 10050\ GeV$.

Figure 5.10 shows the detector level double differential cross-section as a function of the dijet invariant mass of the two leading jets in the event. The detector level spectra are constructed by using the turn-on points calculated with respect to the leading jet p_T. Each mass bin is filled with events that originate from a mixture of different trigger paths. Different marker styles separate the rapidity bins while each distribution is further scaled for better visualization.

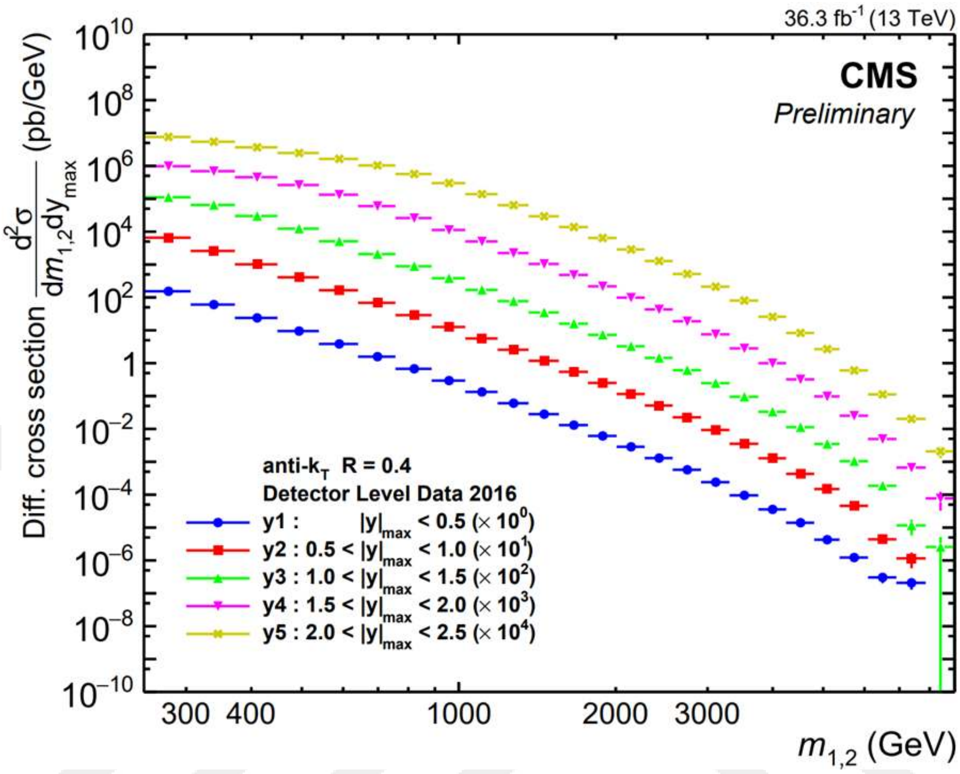

Figure 5.10. Detector level data double differential cross-section with respect to the dijet mass $m_{1,2}$ for AK4 jets. Rapidity bins are shown in different marker styles.

The contribution of each trigger to the detector level spectrum for AK4 jets is shown in Figure 5.11. Each canvas corresponds to a different rapidity bin, while the color scheme separates the contribution of each trigger for a given mass bin. The plot shows no contribution from the HLT_PFJet_40 which is justified due to the high phase space cuts (Table 5.4) imposed in the transverse momentum of the dijet pair.

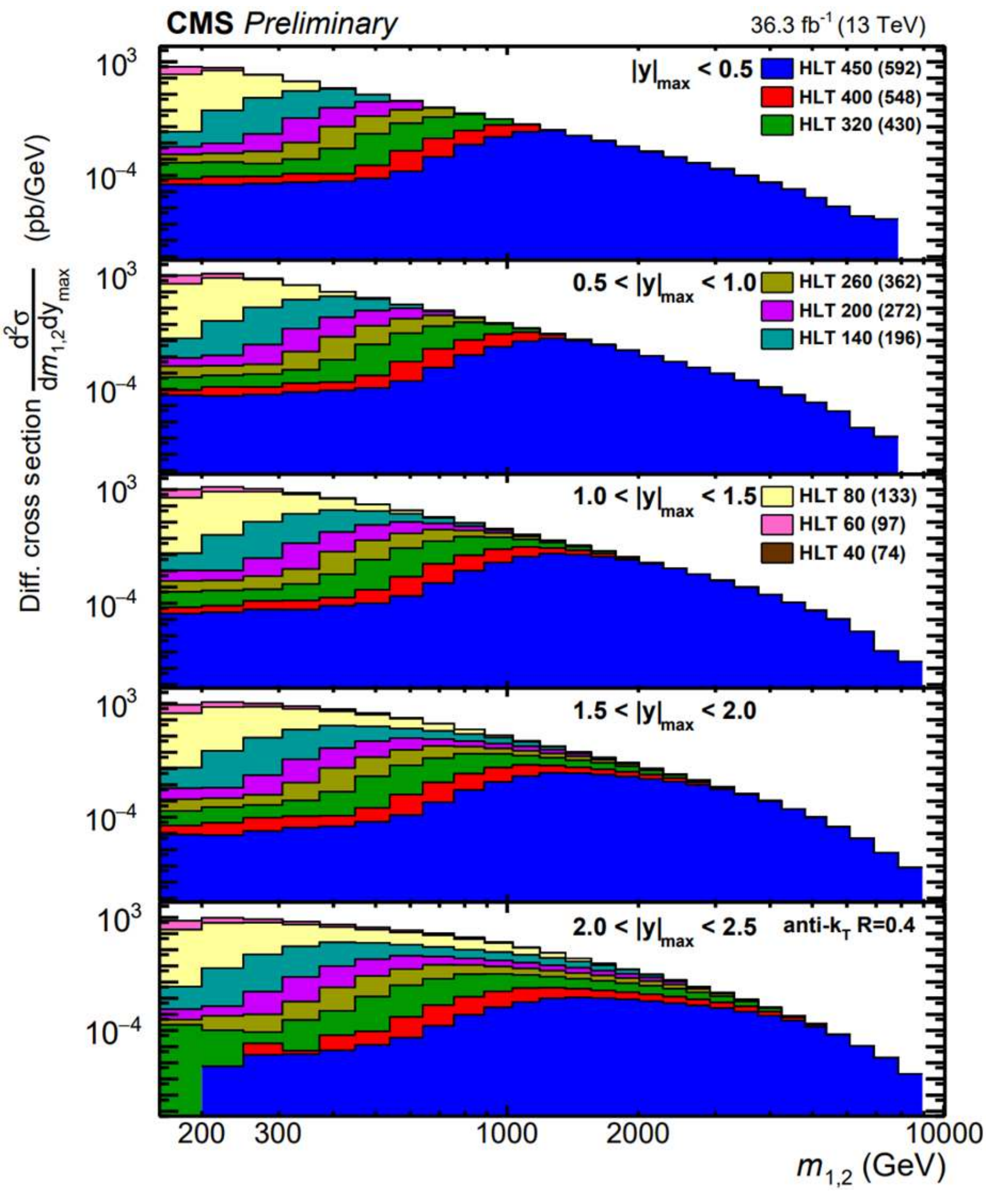

Figure 5.11. Contribution of the different trigger paths to the detector level spectrum for AK4 jets. Each canvas corresponds to a different rapidity bin, while the color scheme separates the contribution of each trigger for a given mass bin.

5.5. Unfolding

The CMS detector is composed of a variety of subsystems which are mainly based on electronics. For this reason, the machine is bound to have some limitations on the execution and precision of different procedures. Each time a measurement is performed, there is an expected deviation between the measured value and the true value of an observable. The measured value includes the impact of the detector and the observed difference compared to the true value is usually referred to as detector response. There can be many sources that lead to this effect such as the finite detector resolution and others such as the nonlinear and inhomogeneous response of the detector or the in-time and out-of-time PU. For example, a jet produced at 1 TeV after the hard process will interact with the detector and its energy will be deposited in the calorimeters, information which is later collected and fed to the reconstruction algorithm. The computed jet energy will have been shifted either to a slightly higher or lower value, an offset attributed to the detector resolution, which causes a misidentification of the jet's true energy. By comparing the measured spectrum with the true spectrum, event migrations can be observed among neighboring bins. This is a consequence of the phase space division happening in binned measurements.

Events near bin edges (mass and rapidity bins) will most likely migrate to neighboring bins because of the shift during the online measurement. Each bin of the measured spectrum contains events that migrated in from neighboring bins and miss events that have migrated out of it. The purpose of unfolding is to negate this effect of the detector from the measured cross sections (Figure 5.10) by unsmearing the detector level spectrum. After the unfolding, a comparison can be made between the resulting measurement at the particle level and the QCD theory predictions since the resulting measurement have been corrected for detector effects.

Two-dimensional unfolding is performed not only to account for migrations happening between mass bins, but also among rapidity bins. In addition, migrations in and out of the phase space are considered as miss and fake entries when going from the detector level to the particle level spectrum but are treated separately. The task at hand is addressed as a least square minimization problem according to the formula:

$$\chi^2 = \min_x[(Ax + b - y)^T V^{-1}(Ax + b - y)] \quad (27)$$

where x represents the vector containing the unfolded distribution that we want to calculate, y is the vector containing the measured distribution at the detector level, b is the background vector at the detector level, A is the probability matrix (PM), and V is the covariance matrix of the detector level distribution containing the data statistical uncertainties with correlations. The TUnfold package (S. Schmitt, 2012) is used by performing a pseudo-inversion which means using twice as more bins at the detector level than at the particle level and no Tikhonov regularization is used. Better stability is achieved compared to the plain matrix inversion by following this strategy.

5.5.1. Response Matrix (RM)

The construction of the probability matrix comes first in the unfolding process. It is a normalized version of the RM, which only includes migrations originating from inside the phase space. For the actual construction, the Pythia sliced sample is used by exploiting the fact that the distributions are provided at both generator and reconstructed levels. Only jets that are matched at both levels are included inside the RM. As mentioned in the previous section, twice the number of bins is used at the reconstructed level than at the generator level, according to the following strategy. For the first three $|y|_{max}$ regions, the amount of reconstructed level bins is doubled by dividing them in half, so that the generator level distributions have the same binning scheme as at the reconstructed level. On the other hand, the exact opposite is done for the two forward $|y|_{max}$ regions. The binning scheme at the reconstructed level is followed this time, but they are merged by two to keep half the number of bins at the generator level.

A new matrix A, known as the RM, is formed if the migration matrix is normalized, which indicates the total of all the entries in each column equals one. A_{ij} represents an estimate of the likelihood that an event generated in bin j will be reconstructed in bin i. Figure 5.12 shows the extracted probability matrices.

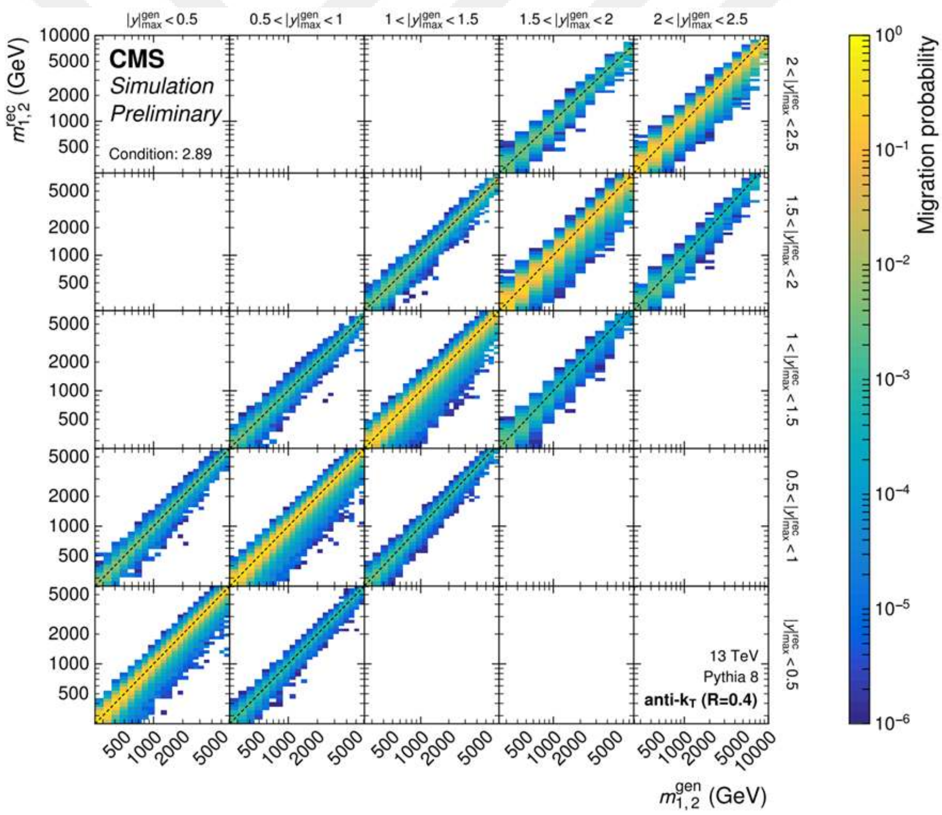

Figure 5.12. Two-dimensional probability matrix. Each cell corresponds to a different combination between $|y|_{max}^{gen}$ and $|y|_{max}^{rec}$ regions and covers the full $m_{1,2}$ phase space at both gen (X axis) and reco (Y axis) level.

The figure contains a 5x5 plot, where each cell of the plot corresponds to a different combination between the $|y|_{max}^{gen}$ and $|y|_{max}^{rec}$ regions. Additionally, every cell has its own X and Y axis covering the full $m_{1,2}$ phase space at the generator and reconstructed level, respectively. Cells in the main diagonal represent event migrations happening among neighboring mass bins, both with the same $|y|_{max}$ value at the generator and at the reconstructed level, which corresponds to the plain one-dimensional unfolding. While off-diagonal cells represent event migrations happening between neighboring rapidity bins at different levels.

Because it expresses a linear map between the generator level and the reconstructed level distributions, the RM is an important aspect of the unfolding process. Finding an inverse of this map to recover a generator level distribution from a measurement on the reconstructed level is the core concept of unfolding.

5.5.2. Comparison to Detector Level

Figure 5.13 shows the ratio between the unfolded distributions and the detector-level spectrums shown in Figure 5.10. The plot consists of five cells corresponding to the five $|y|_{max}$ bins. By closely observing the two forward regions one can see the coarser binning scheme, in action, used in the unfolded distribution. The ratio itself is a way of illustrating the effect of unfolding and is mostly around unity, following divergence from unity across high $m_{1,2}$ phase space, where the statistical uncertainty becomes dominant. Such behavior is expected for a steeply falling spectrum since migrations caused by the detector response are corrected after the unfolding, which yields an unfolded distribution with less populated mass bins.

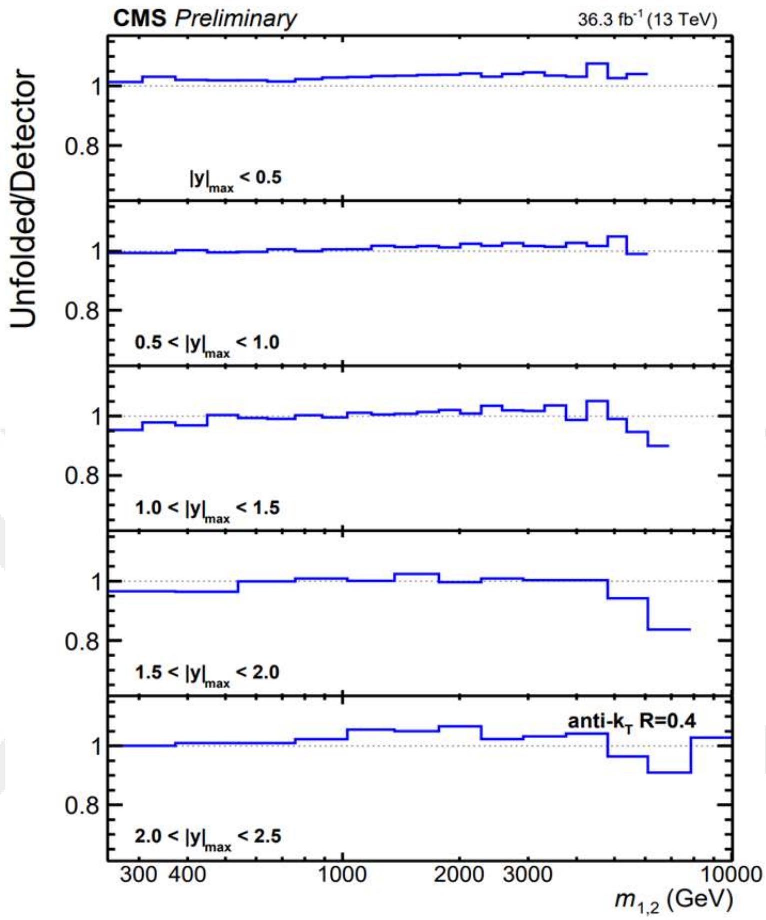

Figure 5.13. The ratio between the unfolded (particle) and detector level distributions for all rapidity bins.

5.5.3. Closure Test and Backfolding

To validate the unfolding procedure and to ensure that the result is unbiased, tests are performed on the Pythia sliced sample which is used to construct the RM. First, the reconstructed Pythia sample is unfolded using the same RM that was constructed by the same Pythia sample and then compared to the generator-level distribution of the same MC sample. This ratio is expected to be exactly at unity since both the unfolded distribution and the RM used to perform the unfolding originate from the same MC sample. This test is called the closure test and it is presented in Figure 5.14 for all rapidity regions.

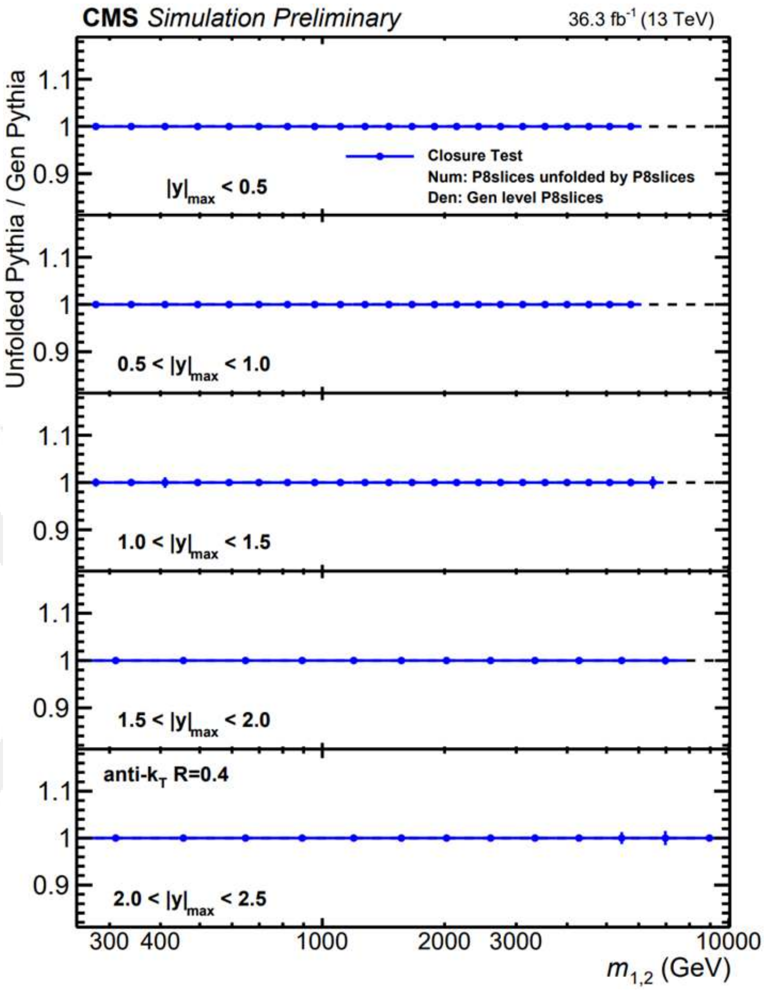

Figure 5.14. The ratio between the unfolded and generator level distributions of the Pythia sliced sample for all rapidity bins.

Following the closure test strategy, but in reverse, a comparison can be made at the generator level. This is called the backfolding test and is shown in Figure 5.15 for all rapidity bins. By applying the RM to the unfolded level distributions of the simulation, the reconstructed level distributions are obtained and compared to the original reconstructed Pythia distributions. Again, the ratio is expected to be at unity for the same reasons.

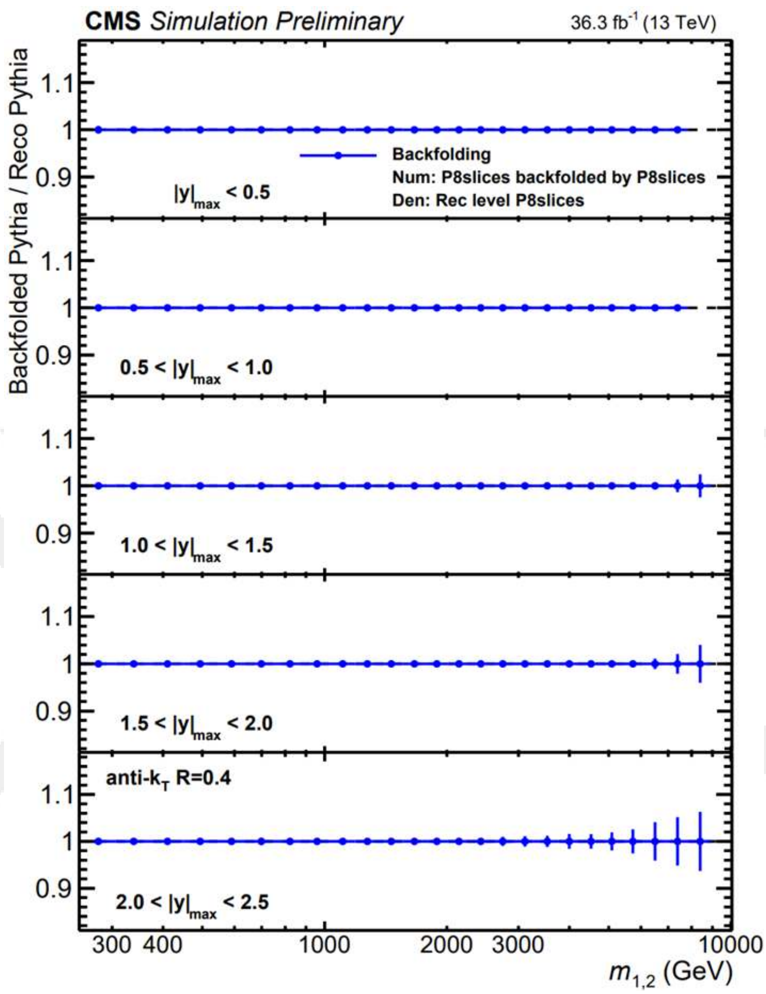

Figure 5.15. The ratio between the backfolded and reconstructed level distributions of the Pythia sliced sample for all rapidity bins.

5.5.4. Bottom Line Test

Another kind of test is the bottom-line test. The ratio between data and simulation is calculated two times, once before unfolding in the form of a detector over the reconstructed level MC and once after unfolding in the form of the unfolded data over the generator level MC. The comparison between these two ratios can be found in Figure 5.16. The important thing here is the shape agreement between the ratios before (in blue) and after (in red) unfolding. Looking at all five rapidity bins, the shapes seem to have a good agreement with the statistical uncertainties, meaning that no bias is introduced during unfolding. Again, for the two forward rapidity regions one can easily see the coarser binning applied at the particle level after unfolding, compared to the finer binning in the central rapidity regions at the same level.

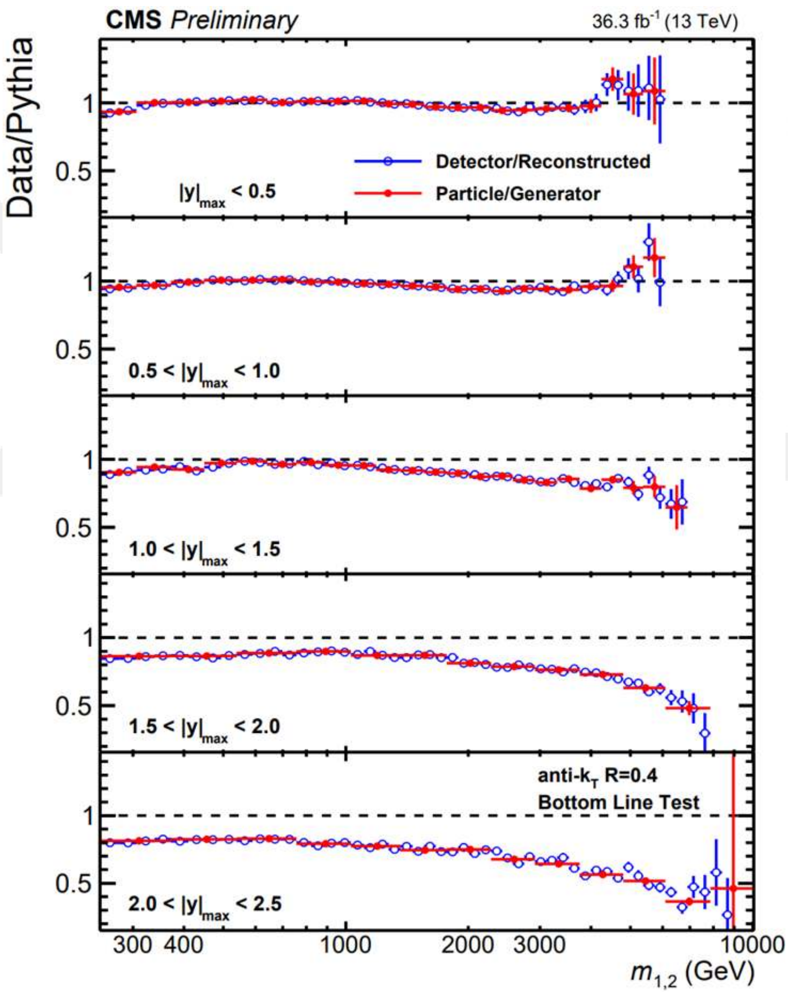

Figure 5.16. The ratio between data and simulation before and after the unfolding for all rapidity bins.

5.5.5. Model Dependence

The choice of an MC sample carries the need to account for an additional uncertainty, that of the model dependence. Each MC generator uses a different setup to simulate the event generation and reconstruction of the samples thus being prone to differences. The difference between the data unfolded by MadGraph to the data unfolded by Pythia is shown in Figure 5.17. The difference from unity is taken as an additional model uncertainty to experimental uncertainties.

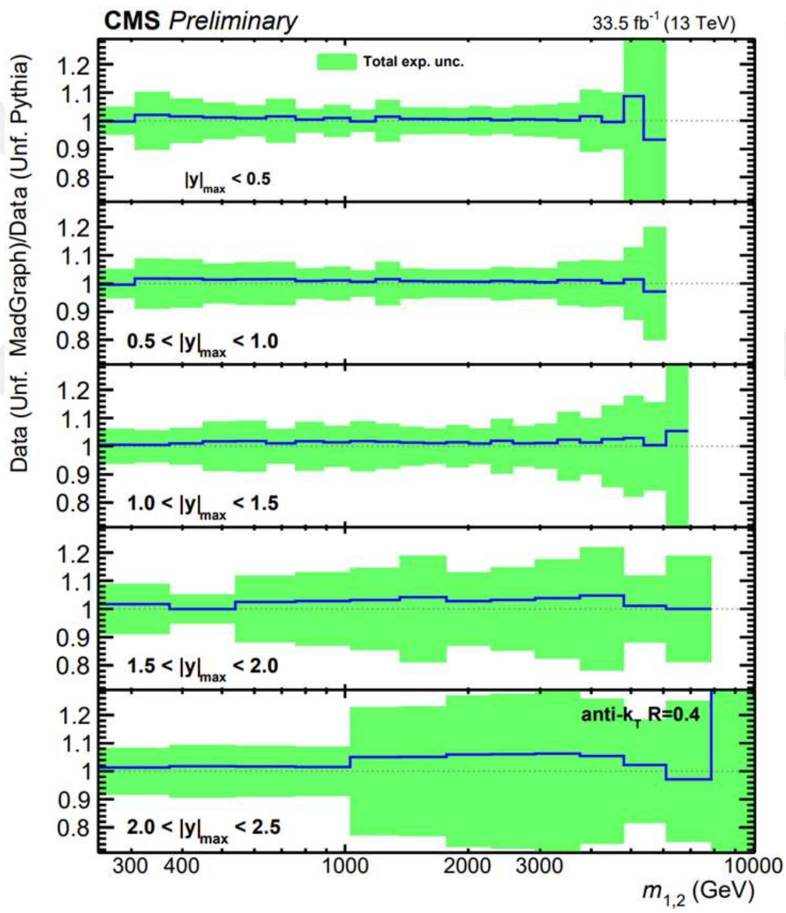

Figure 5.17. The ratio between data unfolded by MadGraph to data unfolded by Pythia is shown with the blue curve. The green band illustrates the total experimental uncertainty. Each cell corresponds to a different rapidity bin.

5.6. Experimental Uncertainties

A variety of uncertainty sources are closely examined and presented in more detail in the subsections. A breakdown of the individual contributions is presented in Figure 5.18 for all rapidity regions. The log scale is preferred for the plot's Y axis to increase the visibility of the fluctuations apparent in each contribution. Different line colors and styles are assigned to each uncertainty source. Uncertainties are symmetrized by taking the average between the up and down variations of the same source. Then the uncertainties are smoothened by performing a fit using *Chebyshev* polynomials. Among all uncertainty sources, the most dominant one is the JES across all $|y|_{max}$ bins. The grey band represents the total experimental uncertainty which is calculated as the quadratic sum of all uncertainty contributions (statistical and systematic). The total experimental uncertainty starts from $\sim 5\%$ for both cone sizes and goes up to 50%.

The impact of systemic factors on the cross-section is frequently estimated through varying experimental parameters within a 1 interval around the nominal value. The shifted detector-level cross-sections then propagate through the unfolding procedure, yielding a both upward and downward shifted estimate of the unfolded cross-section.

5.6.1. Statistical Uncertainty

Because of the limited number of events recorded, the cross-section estimates are subject to statistical uncertainty, which is often represented by a Poisson distribution. The event counts in this study are sufficiently high for this to be approximated by a Gaussian distribution instead. The standard deviation of the Gaussian approximation, which is supplied by the square root of the sum of the squares of all event weights in a bin, is used to determine the uncertainty size.

While the statistical uncertainty is naturally uncorrelated at the detector level, the situation does not remain the same after unfolding. At the hadron level, some correlation rises among neighboring mass/rapidity bins due to events that migrate back to their original bin because the migrations caused by the detector response are corrected after unfolding. These correlations are transferred through unfolding by the covariance matrix.

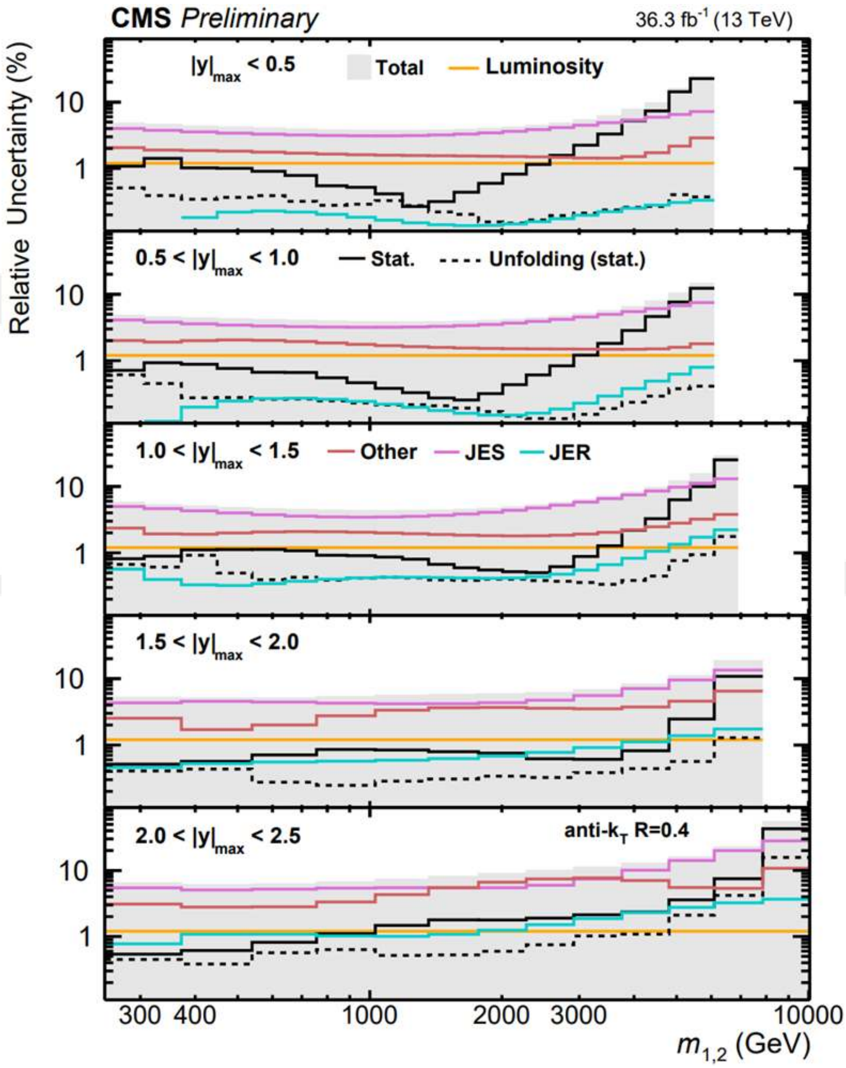

Figure 5.18. Smooth relative uncertainty (%) on the measured cross-section for all $|y|_{max}$ bins. The quadratic sum of the individual contributions is illustrated by a grey band and represents the total experimental uncertainty.

5.6.2. JES Uncertainty

Because JECs are done in phases, the transverse momentum of reconstructed jets is sensitive to many sources of systematic bias. (CMS Collaboration, 2018.). The overall uncertainty in jet energy associated with these biases is in the few percent range. The impact is increased by the steeply falling spectrum when obtaining the cross-sections, and consequently makes up the largest uncertainty in this analysis.

JES uncertainties are made up of 27 separate contributions, each of which reflects a different systematic effect. Repeating the process with systematic both upward and downward shifts applied to the momentum of each reconstructed jet yields an estimate of each uncertainty component. Along with the centrally calculated JECs, shift parameters are supplied as a function of individual jet kinematics and other event characteristics.

In descending order of their overall impact on the cross-section, the JES contributions include the change in experimental conditions over time, the calibration of the relative and absolute JES as a function of η and p_T, the change in response for jets initiated by gluons and different quark flavors, and PU. Each contribution represents a fully correlated uncertainty across all data points and is independent of the other contributions. Four broad categories of JES uncertainty components can be identified as follows:

- **Primary:** Seven JES components are found to contribute at the level of 1% or above in all rapidity regions and are thus called "primary". These are *TimePtEta* $(1-7\%)$, *AbsoluteMPFBias* $(1-5\%)$, *AbsoluteScale* $(1-3\%)$, *FlavorQCD* $(1-4\%)$, *RelativeSample* $(1-8\%)$, *RelativeBal* $(1-3\%)$, and *RelativeFSR* $(1-6\%)$, with the numbers in parentheses indicating the approximate range across all rapidity regions of the median uncertainty value.

- **Secondary:** The following eight JES components are found to be below 1% in all rapidity regions: *RelativePtBB*, *Fragmentation*, *SinglePionECAL*, *SinglePionHCAL*, *RelativeStatFSR*, *PileUpDataMC*, *PileUpPtRef*, and *PileUpPtBB*.
- **Endcap specific:** 7 JES components are only significant in the endcap regions with $2 < |\eta| < 3$, namely: *RelativeStatEC*, *RelativeJEREC1*, *RelativeJEREC2*, *RelativePtEC1*, *RelativePtEC2*, *PileupPtEC1*, and *PileupPtEC2*.
- **Negligible:** The remaining contributions are found to be negligible in the phase space measured in this analysis. These comprise four JES sources that are only relevant in the *HF* region at $|\eta| > 3$ (*RelativeJERHF*, *RelativePtHF*, *RelativeStatHF*, *PileUpPtHF*) and one contribution (*AbsoluteStat*) that evaluates to less than 0.05% in all regions.

5.6.3. JER Uncertainty

The detector response used as an input for the unfolding process is dependent on the simulation's estimated JER. While this is specifically modified to match the data resolution, applied scale factors are still susceptible to uncertainty, which is centrally provided alongside them. The unfolding technique is repeated using responses obtained through a systematic variation of the JER to evaluate the impact of JER modeling on the unfolded cross-section. This is accomplished by varying the simulation to the data scale factor used in JER smearing. Upward and downward variations are readily available. The overall uncertainty because of JER, like JES, is derived from an asymmetric confidence range around the nominal cross-section, as supplied by the shifted estimations.

5.6.4. Luminosity Uncertainty

The integrated luminosity enters into the calculation of the effective number of events, and thus affects the cross-section measurement. In line with the recommendation of the LUMI POG, the uncertainty in the final cross-section due to the luminosity measurement is estimated to be 1.2% (CMS Collaboration, 2021b).

5.6.5. Unfolding Uncertainty

The finite size of the simulated sample used to determine the detector response adds an additional factor to uncertainty through the unfolding process. As a result, the generated RM is subject to statistical uncertainty, which can produce variations in the unfolded cross sections even when none are present in the input detector level distribution. This impact is mitigated by adding a new unfolding uncertainty. TUnfold calculates this uncertainty by mathematically propagating the statistical uncertainties of the RM to the unfolded cross-sections. The unfolding uncertainty computed in this manner, like the statistical uncertainties of the input distributions, display correlations between neighboring bins and are consequently supplied as a complete covariance matrix.

Overall, the unfolding uncertainties obtained are of the order of 1%, remaining mostly below this value in well-populated areas of the phase space. In bins at the upper edge of the phase space, where the event numbers are lower, the unfolding uncertainty increases to %4.

5.6.6. Other Uncertainties

A mixture of different uncertainty sources is included in the other contribution. A 5% normalization uncertainty is inserted before unfolding with the call of the TUnfold objects due to the background subtraction (fake entries). A 5% normalization uncertainty is also estimated by hand due to miss entries by rescaling the inefficiency. Additionally, a prefiring uncertainty is included to account for the inefficiencies caused by triggers that were firing prematurely. A PU profile uncertainty is added to account for the relevant correction. Lastly, model uncertainty is introduced as discussed due to the MC choice for the construction of the RM used to unfold the dijet mass distributions. Overall, the evaluation of the other contribution is of the order of 1% or below in the whole $m_{1,2}$ phase space, except for the forward region, where it ranges between 2% and 5%. The higher values in the outer $|y|_{max}$ region ($|y|_{max} > 2.0$) are justified by the effect of the prefiring correction, which is dominant there, in contrast with the rest of the phase space, which is mostly negligible.

5.7. Comparison to Fixed Order Theory Predictions

Comparison between data and fixed order predictions is achieved both at NLO and at NNLO in pQCD. The theory is obtained with the NNLOJET program (T. Gehrmann et al., 2018) using the APPLFAST (D. Britzger et al., 2019) interface to the FASTNLO package (T. Kluge, K. Rabbertz, and M. Wobisch, 2006). The FASTNLO package provides the necessary tables and allows the use of alternative PDF sets. The statistical precision of the prediction ranges between 0.5% and 2% depending on the phase space region.

Theory production happens in two stages. First, the ME needs to be calculated according to pQCD which is then combined with a PDF set to account for effects happening at larger distances. Still, the final result is given at a parton level meaning that NP effects like MPI and HAD are not included. Before making a comparison with the unfolded data, the theory must be corrected for the NP processes. By using MC event generators, the particle level information both with and without the MPI and HAD processes can be accessed since they can be easily turned on or off. The ratio between the two cross-sections is a way to parametrize the lack of the NP information into a factor that can be then applied directly to the prediction. Mathematical representation of the NP correction factor c^{NP} is:

$$c^{NP} = \frac{\sigma^{PS+\ MPI+HAD}}{\sigma^{PS}} \tag{28}$$

where the PS is a perturbative component.

The model dependence of the NP corrections is evaluated by comparing results from several MC event generators. LO particle level predictions are obtained from PYTHIA 8 using the CUETM1 (CMS Collaboration, 2016a) and CUETM2 (CMS Collaboration, 2016b) tunes, and HERWIG++ (M. Bahr et al., 2008) using the EE5C (M. H. Seymour and A. Siodmok, 2013) tune. These generators are interfaced with POWHEG (P. Nason, 2004) to provide NLO predictions. An additional set of predictions is obtained from HERWIG 7.2 (J. Bellm et al., 2016) with the CH3 tune (CMS Collaboration, 2021a) at both LO and NLO. To reduce the effect of the statistical fluctuations, the values of c^{NP} are fitted as a function of $m_{1,2}$ using a three-parameter function shown in Equation 29. The final correction factor is obtained as the average between the largest and smallest value of c^{NP} obtained across all MC configurations, and half the difference between these values is assigned as an uncertainty.

$$f\left(m_{1,2}\right) = A + B * m_{1,2}^{C} \tag{29}$$

Figure 5.19 shows the fit performed for PYTHIA 8 using the CUETM1 tune to obtain the NP correction factors c^{NP} at LO.

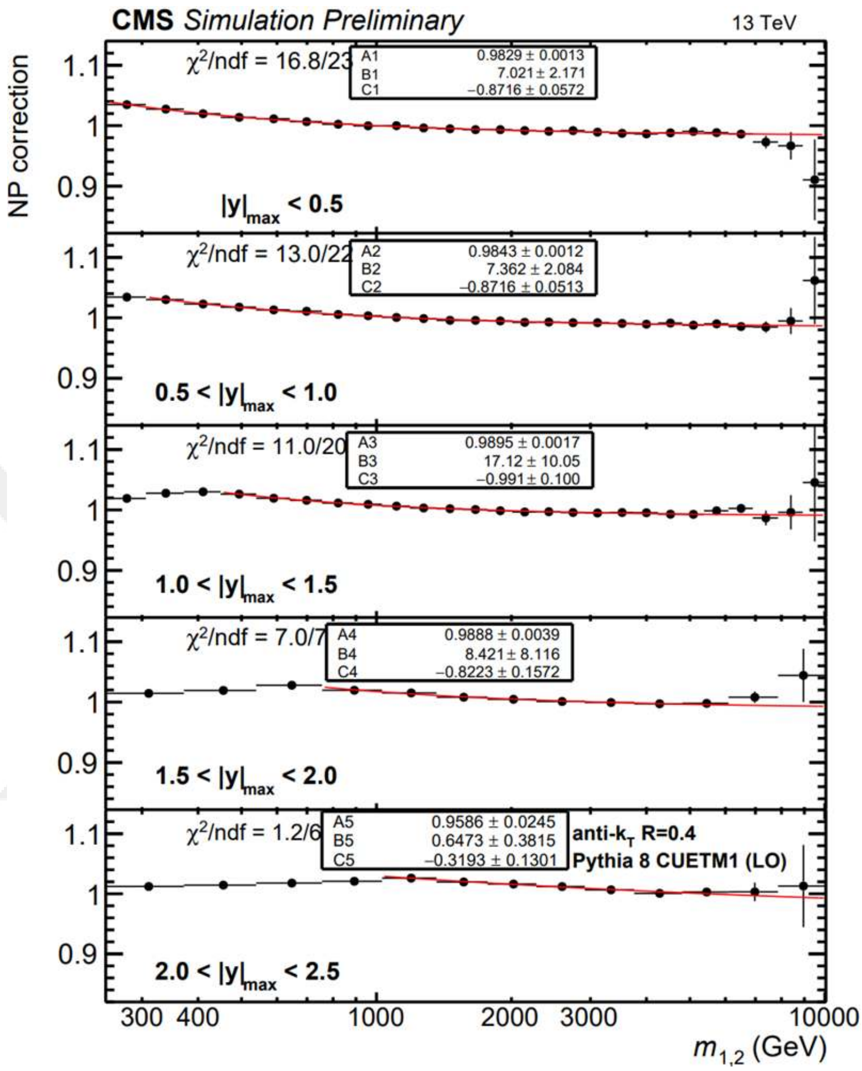

Figure 5.19. NP correction factors derived from Pythia 8 CUETM1 at LO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.20 shows the fit performed for PYTHIA 8 using the CUETM2 tune to obtain the NP correction factors c^{NP} at LO.

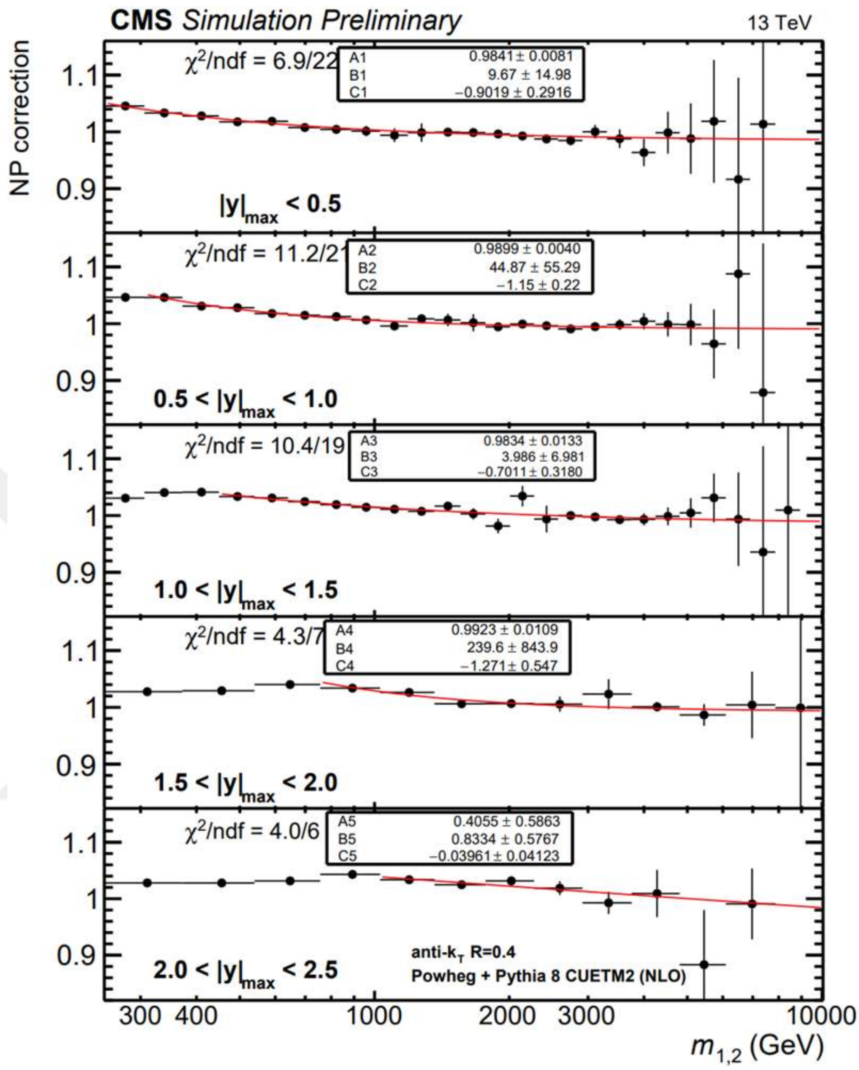

Figure 5.20. NP correction factors derived from Pythia 8 CUETM2 at LO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.21 shows the fit performed for HERWIG++ using the EEC5 tune to obtain the NP correction factors c^{NP} at LO.

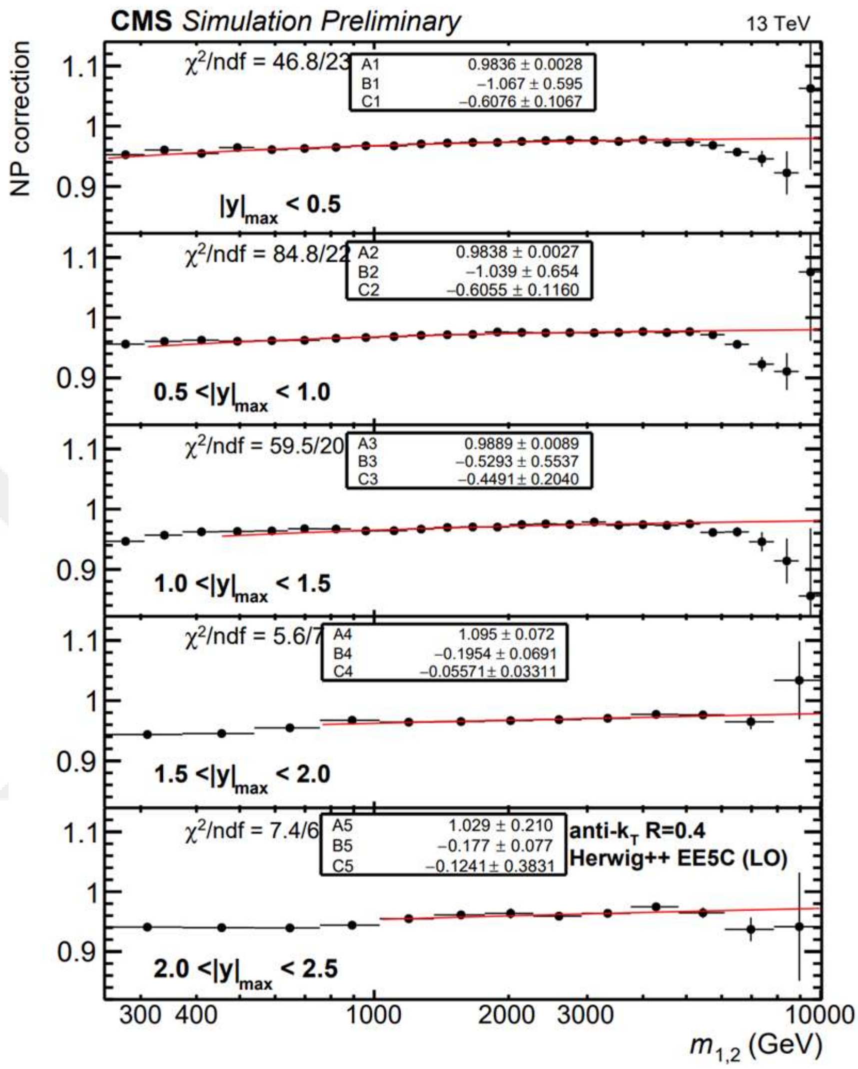

Figure 5.21. NP correction factors derived from HERWIG++ EE5C at LO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.22 shows the fit performed for HERWIG 7.2 using the CH3 tune to obtain the NP correction factors c^{NP} at LO.

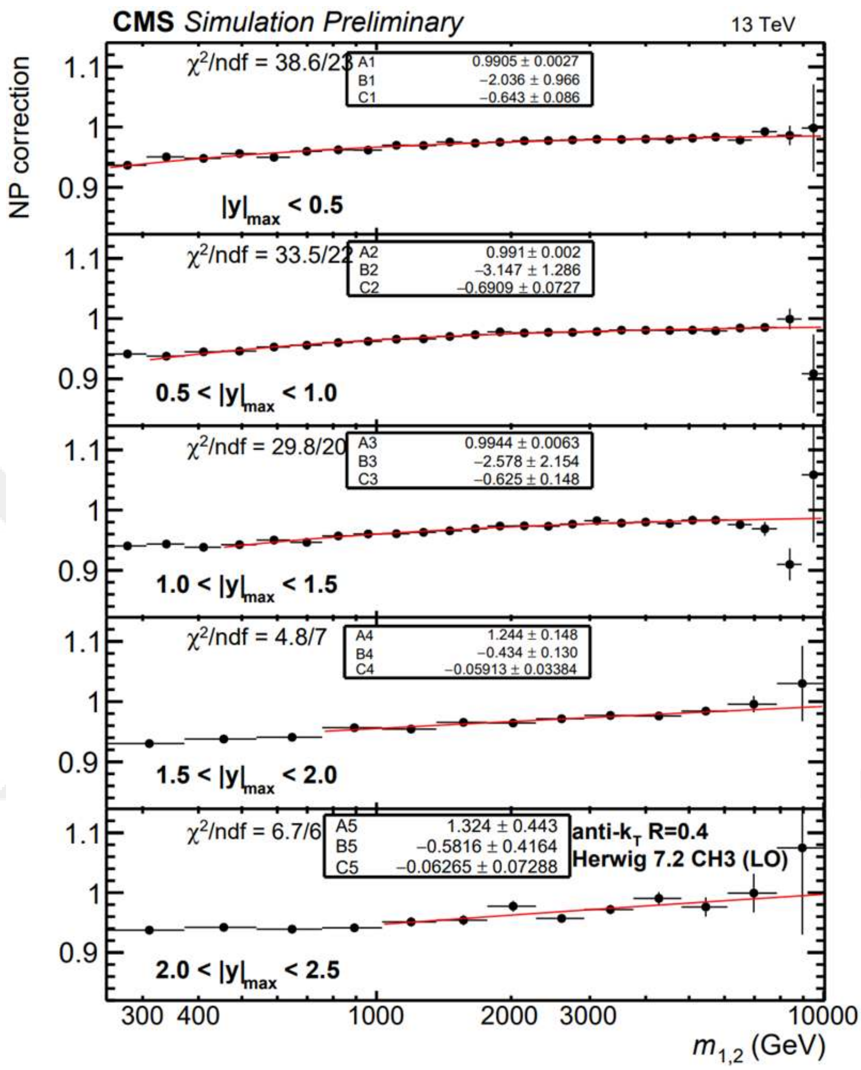

Figure 5.22. NP correction factors derived from HERWIG 7.2 CH3 at LO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.23 shows the fit performed for Powheg + Pythia 8 using the CUETM1 tune to obtain the NP correction factors c^{NP} at NLO.

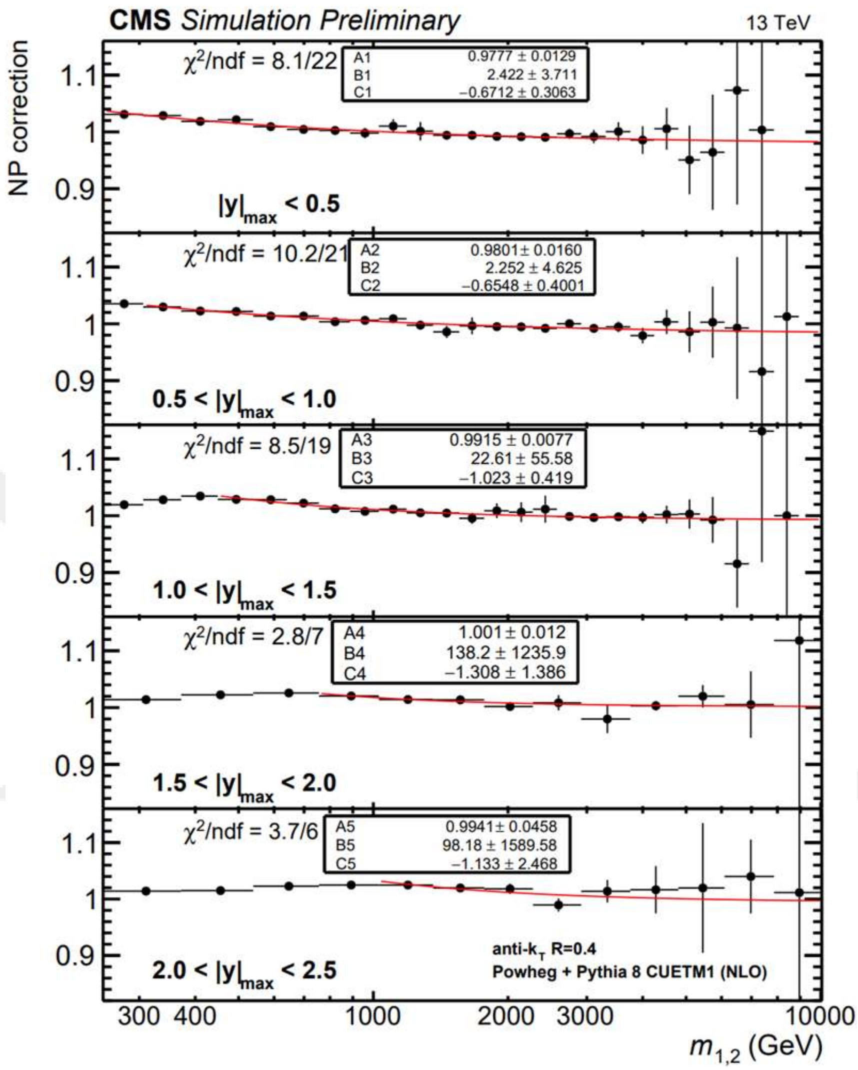

Figure 5.23. NP correction factors derived from Powheg + Pythia 8 CUETM1 at NLO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.24 shows the fit performed for Powheg + Pythia 8 using the CUETM2 tune to obtain the NP correction factors c^{NP} at NLO.

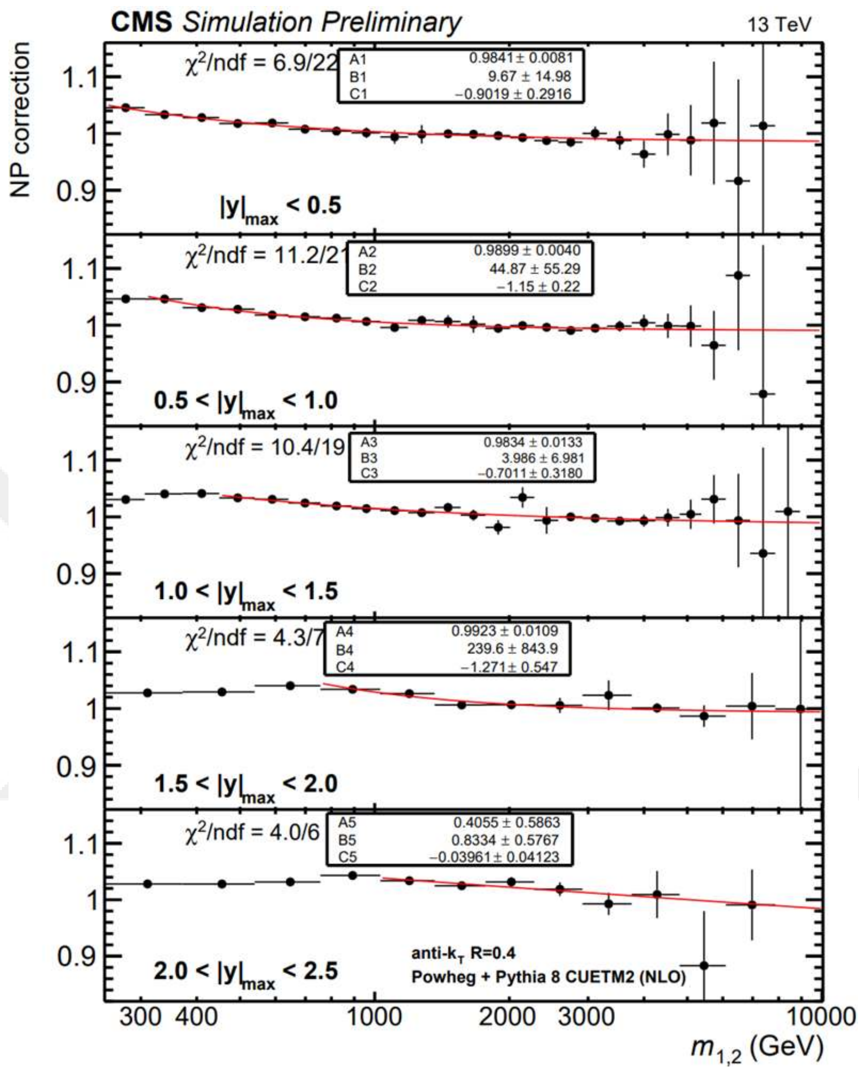

Figure 5.24. NP correction factors derived from Powheg + Pythia 8 CUETM2 at NLO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.25 shows the fit performed for Powheg + HERWIG++ EE5C using the CUETM2 tune to obtain the NP correction factors c^{NP} at NLO.

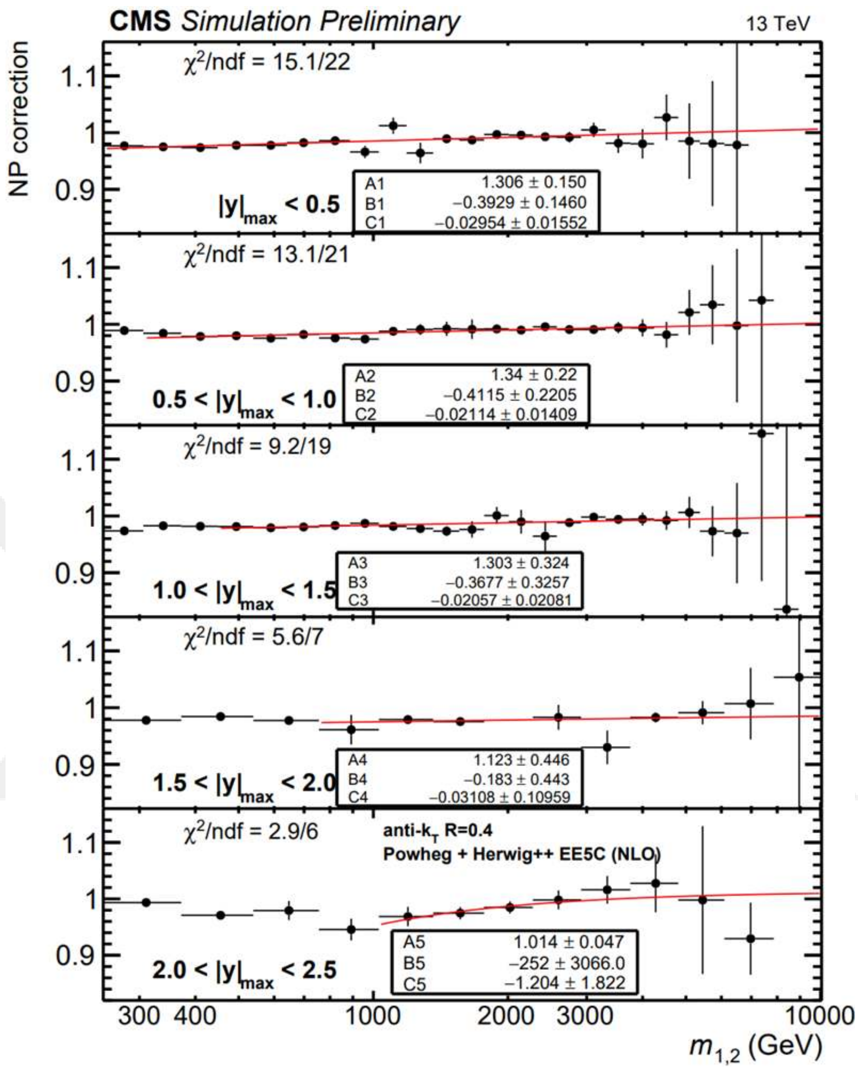

Figure 5.25. NP correction factors derived from Powheg + HERWIG++ EE5C at NLO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Figure 5.26 shows the fit performed for Powheg + HERWIG 7.2 using the CH3 tune to obtain the NP correction factors c^{NP} at NLO.

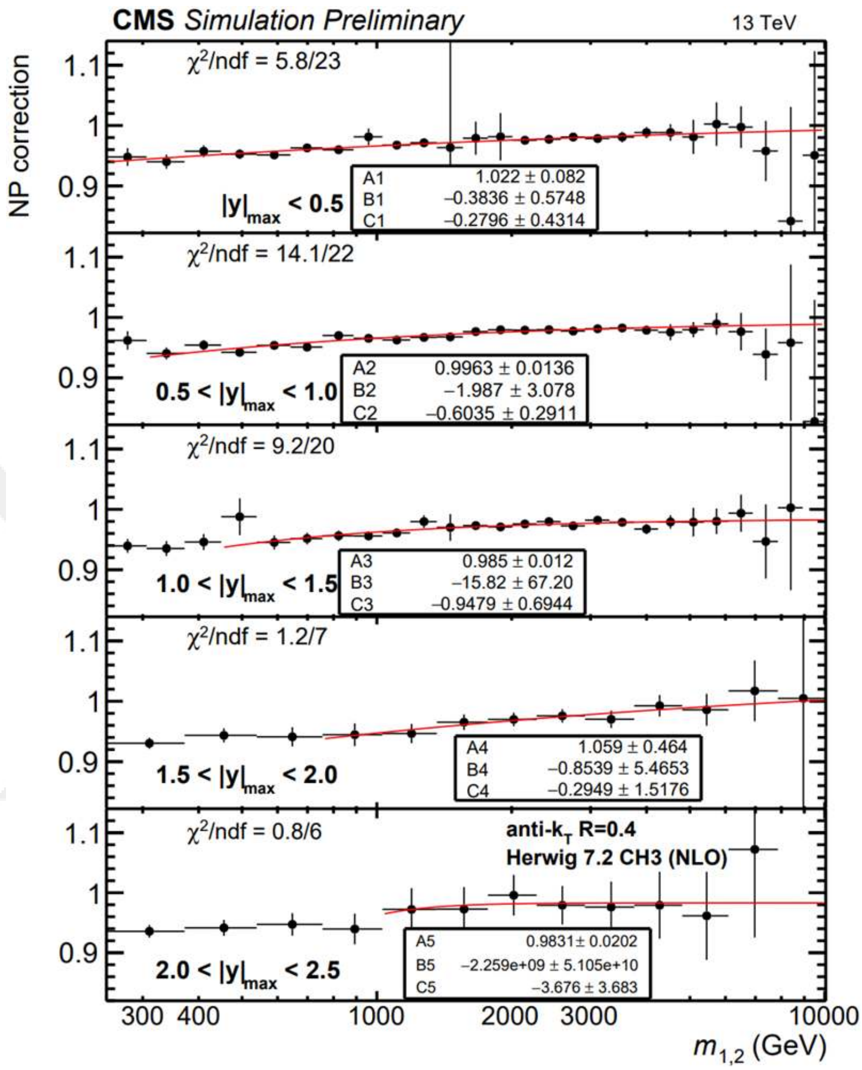

Figure 5.26. NP correction factors derived from Powheg + HERWIG 7.2 CH3 at NLO for all five rapidity regions. The red line represents a FIT parameterization inside a subset of the $m_{1,2}$ phase space in each $|y|_{max}$ bin.

Considering the contributions from all MC generators, an envelope of NPs at both LO and NLO is formed by taking minimum and maximum values. The comparison between these two envelopes is shown in Figure 5.27. The correction factors c^{NP} are taken as the middle point of the envelope indicated by the solid/dashed lines while the NP uncertainty due to the model dependence of the generators is taken as the half spread of the envelope, resulting in a symmetric uncertainty. The vertical dashed black lines indicate the point above which the NP correction is given by the fit parametrization shown in Figures 5.19 – 5.26. In most cases, the values obtained at NLO are seen to be comparable to those obtained at LO from the same generator. The NP correction is dominant mostly in the lower mass region, while for higher energies c^{NP} approaches unity. Theory predictions are corrected according to the NP correction factors given by the envelope.

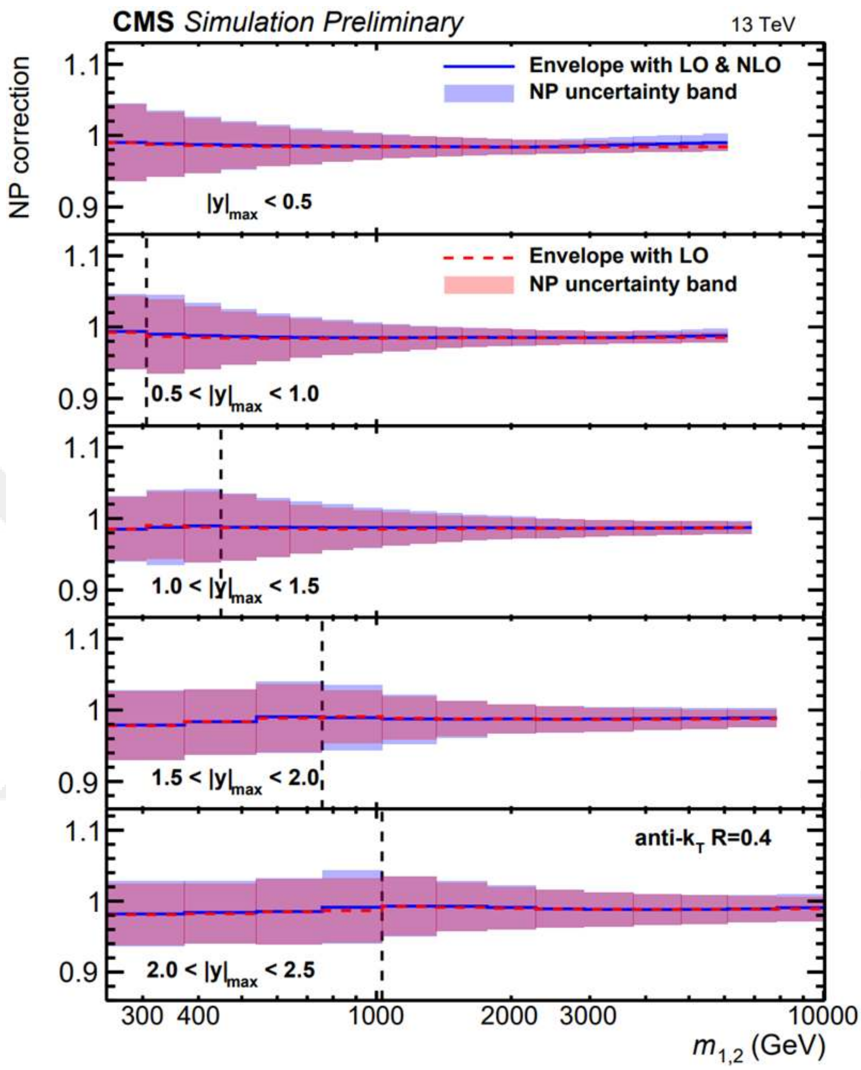

Figure 5.27. Comparison between two versions of the NP correction factors envelope. Each cell in the plot corresponds to a different $|y|_{max}$ The blue envelope is obtained by combining all eight MC predictions both at LO and NLO. The red envelope is obtained by combining only the LO MC predictions. The solid blue line and the dashed red line indicate the central value of the two envelopes, while the blue and red bands indicate the corresponding NP uncertainty. The vertical dashed black lines indicate the point above which the NP correction factors are taken by the fit, while correction factors are taken by value from the bin content below the lines.

For jet transverse momenta in the TeV range, EW contributions to the differential dijet cross section become important and must be considered in addition to the pQCD calculation (S. Dittmaier, A. Huss, and C. Speckner, 2012). These effects, which arise due to the virtual exchange of soft or collinear W or Z bosons, are accounted for by applying a multiplicative correction factor to the pQCD prediction. As shown in Figure 5.28, these factors exhibit a strong dependence on the rapidity and transverse momentum of the jets, reaching 15% at the highest probed values of $m_{1,2}$. The uncertainty in these correction factors is considered to be negligible.

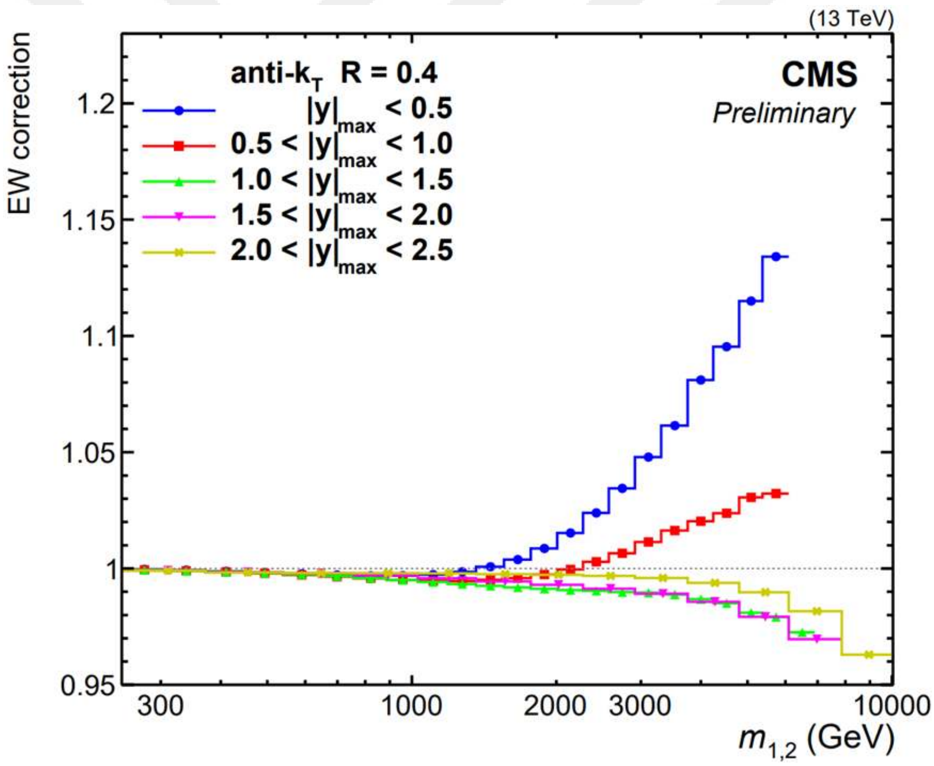

Figure 5.28. EW correction factors obtained as a function of $m_{1,2}$ in the five different rapidity regions. The corrections depend strongly on the kinematic properties of the jets and are observed to be largest at central rapidities for $m_{1,2} > 1\ TeV$.

A comparison of the present measurement to predictions from the fixed-order calculations at NNLO complemented by the NP and EW corrections can be seen in Figure 5.29. The reference predictions are obtained with the CT18 NNLO PDF set (T.-J. Hou et al., 2021) and compared to the data.

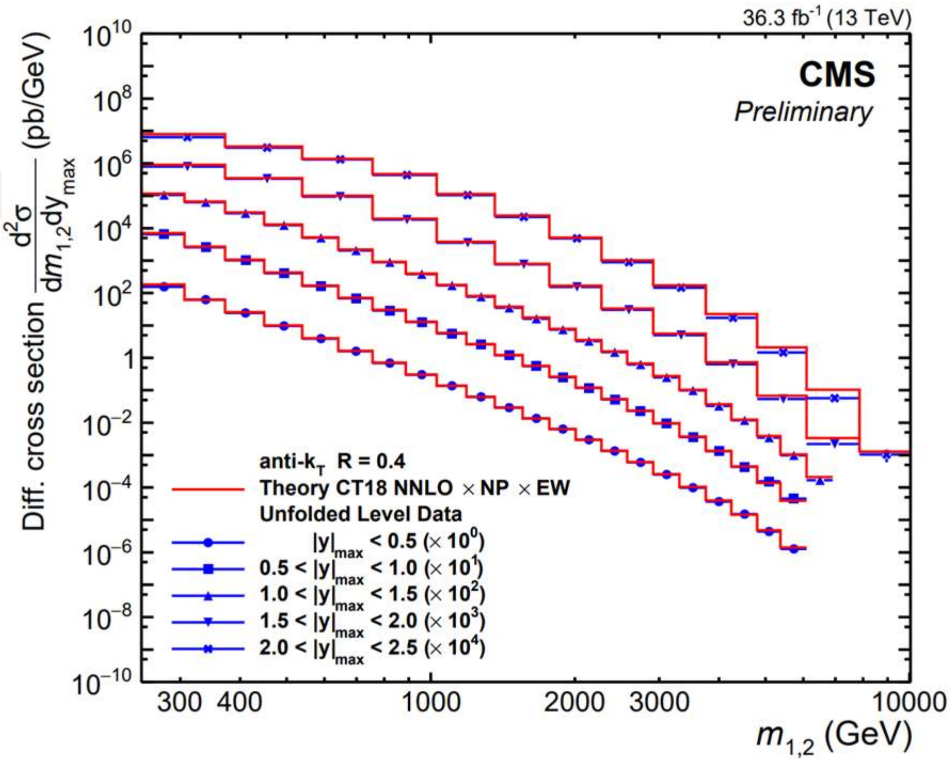

Figure 5.29. Overview of the dijet cross sections illustrated as a function of $m_{1,2}$. The markers and lines indicate the measured unfolded cross sections and the corresponding NNLO predictions, respectively. For better visibility, the values are scaled by a factor depending on the rapidity region as indicated in the legend.

For a more detailed comparison, ratios of the measured cross sections to the theory predictions are shown in Figure 5.30.

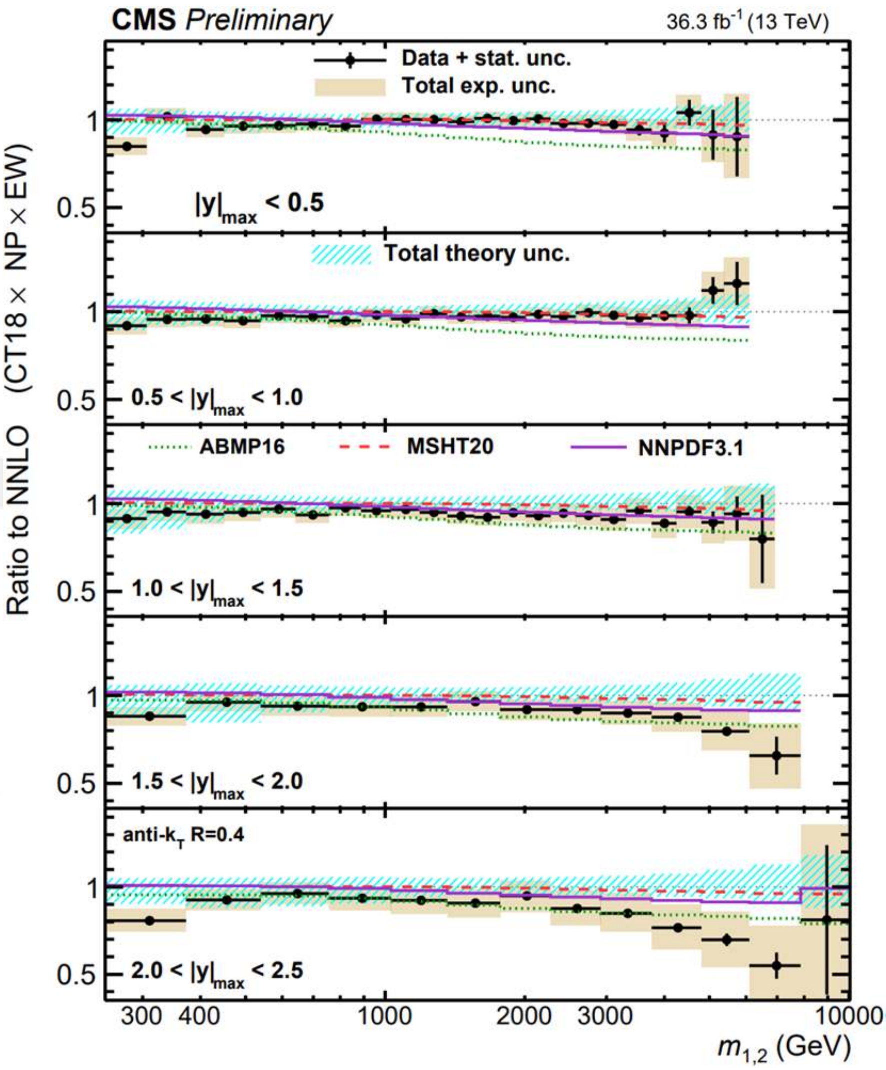

Figure 5.30. Comparison of the dijet cross section as a function of $m_{1,2}$ to fixed-order theory calculations at NNLO. The ratios of the measured cross sections (markers) to predictions obtained using the CT18 NNLO PDF set are shown. The error bars and shaded yellow regions indicate the statistical and the total experimental uncertainties of the data, respectively, and the hatched teal band indicates the total theory uncertainty. Alternative theory predictions obtained using other global PDF sets are shown as colored lines.

The black error bars around the markers represent the relative statistical uncertainty of the data, while the yellow band represents the relative total experimental uncertainty of the data. Around unity with a teal dashed band is indicated the relative total theory uncertainty, which is calculated as the square root of the quadratic sum of the individual PDF, scale, and NP correction uncertainties. Also, in the same figure comparisons with different PDF sets are presented. These are ABMP16 (S. Alekhin, J. Blumlein, S. Moch, and R. Placakyte, 2017), MSHT20 (S. Bailey et al., 2021), and NNPDF3.1 (NNPDF Collaboration, 2017) in the form of theory ratios between the PDF sets and the CT18 NNLO used as reference. The level of agreement is observed to be good in most phase space regions, with some deviations at the lower ends of the spectra and in the outer rapidity regions.

6. QCD ANALYSIS

To evaluate the impact of the dijet measurements on the determination of the proton PDFs and the strong coupling constant $\alpha_s(M_z)$, a QCD analysis is performed following the approach taken by the latest HERAPDF analyses (H1 and ZEUS Collaboration, 2010, 2015, 2022). The data entering this analysis consist of measurements of DIS obtained in $e^{\pm}p$ collisions at the HERA collider experiments H1 and ZEUS as a function of the momentum transfer Q^2, supplemented by the present measurements of the dijet cross-section. The HERA data are restricted to Q^2 values above a threshold $Q^2_{min} = 10$ GeV to ensure a good description of the measurements by the theory predictions and are fitted together with the double differential dijet cross sections. It is a well-known fact that fixed-order predictions work most reliably for inclusive observables, where the phase space for QCD radiation is not restricted. Otherwise, for very small jet distance parameters, the cancellation of singularities between real emissions and virtual corrections can lead to large logarithms, which spoil the perturbative convergence. Thus, to profit most from the available predictions at NNLO, only data for the jets with the larger jet distance parameter of R = 0.8 are used in the fits.

Theory predictions for the dijet cross sections are obtained from NNLOJET and FASTNLO at NNLO accuracy, considering the full dependence of the NNLO cross sections on the PDFs, the strong coupling constant $\alpha_s(M_z)$, and the renormalization and factorization scales. The dijet invariant mass $m_{1,2}$ is chosen as the central value for both the renormalization and factorization scales. Fits of the PDFs are performed using the xFITTER program (S. Alekhin et al., 2015 and V. Bertone et al., 2018), interfaced to FASTNLO for access to the theory predictions for the CMS dijet cross sections. The evolution of PDFs following the DGLAP equations (V. N. Gribov and L. N. Lipatov, 1972 and Y. L. Dokshitzer, 1977) is performed using the QCDNUM package (M. Botje, 2011). Contributions from heavy quarks are treated in the Thorne–Roberts optimal variable flavor number scheme (RTOPT) (R. S. Thorne and R. G. Roberts, 1998 and R. S. Thorne, 2016), with the masses of the charm and bottom quarks set to $m_c = 1.43$ GeV and $m_b = 4.5$ GeV.

In the HERAPDF approach, the proton structure is expressed in terms of the gluon distribution function $g(x)$, the up and down valence quark distribution functions $u_v(x)$ and $d_v(x)$, and the up and down type sea antiquark distribution functions $\overline{U}(x)$ and $\overline{D}(x)$. For each of these distributions f, the dependence on the momentum fraction x is parametrized at a starting scale $\mu_{F.0}^2 = 1.9$ GeV^2 as

$$xf(x, \mu_{F.0}^2) = A_f x^{B_f}(1-x)^{C_f}(1 + D_f x + E_f x^2) \tag{30}$$

The overall normalization of the PDFs is given by the A parameters, with A_{u_v}, A_{d_v}, and A_g being constrained by the quark number and momentum sum rules. The B and C parameters control the shape of the distribution as x approaches the edges of its domain at 0 and 1, respectively. The D and E parameters represent additional degrees of freedom related to the functional forms.

There are total of ten A, B, and C parameters. Values of these parameters are determined during the fit procedure. Additional D and E parameters are included if they lead to improved fit quality in terms of the total χ^2 value. Starting from the initial parametrization with no additional parameters, the change in χ^2 resulting from the inclusion of any of the remaining D and E parameters in the fit is evaluated. The parameter resulting in the largest decrease in χ^2 is added to the parametrization and the procedure is repeated until no further significant improvement is observed. The parameters added in the final parametrizations are E_{u_v}, D_{u_v} and $D_{\overline{U}}$. The final χ^2 values obtained by combining HERA DIS and CMS data for the double differential fits can be found in Table 6.1.

Table 6.1. Partial χ^2 values for fits to the HERA DIS + CMS double differential dijet measurements.

Dataset	**Partial χ^2/n_{data}**	
	HERA DIS only	**HERA DIS + CMS 2D**
CMS dijet 2D, $\|y\|_{max} < 0.5$		$24/22 = 1.1$
CMS dijet 2D, $0.5 < \|y\|_{max} < 1.0$		$14/22 = 0.63$
CMS dijet 2D, $1.0 < \|y\|_{max} < 1.5$		$22/23 = 0.97$
CMS dijet 2D, $1.5 < \|y\|_{max} < 2.0$		$15/12 = 1.2$
CMS dijet 2D, $2.0 < \|y\|_{max} < 2.5$		$30/12 = 2.5$
HERA1+2 CCem	$52/42 = 1.2$	$52/42 = 1.2$
HERA1+2 CCep	$37/39 = 0.94$	$37/39 = 0.94$
HERA1+2 NCem	$221/159 = 1.4$	$221/159 = 1.4$
HERA1+2 NCep 460	$199/177 = 1.1$	$197/177 = 1.1$
HERA1+2 NCep 575	$186/221 = 0.84$	$186/221 = 0.84$
HERA1+2 NCep 820	$56/61 = 0.91$	$55/61 = 0.9$
HERA1+2 NCep 920	$363/317 = 1.1$	$368/317 = 1.2$

The uncertainties in the fitted PDFs are investigated using a procedure that defines separate contributions from fit, model, and parametrization uncertainties. Fit uncertainty is estimated using the Hessian method (J. Pumplin et al., 2001) and represents a propagation of the uncertainties in the measurements, theory predictions, and theory correction factors to the PDFs. The model uncertainty arises from the choices made for the values of certain non-PDF parameters such as the minimum Q^2 value used for restricting the HERA data, the strangeness fraction f_s, the charm and bottom quark masses m_c and m_b, and the value of the starting scale $\mu_{F,0}$. It is estimated by varying the values of these parameters up and down from their nominal values as indicated in Table 6.2 and adding the differences to the nominal fit result in quadrature separately for each variation direction. A further uncertainty arises due to the choice of PDF parametrization. It is estimated by performing alternative fits that include one additional D or E parameter compared to the nominal parametrization. The maximum deviation between the nominal PDF and those obtained from the alternative parametrizations is taken as an additional parametrization uncertainty.

Table 6.2. Nominal values and variations of parameters used to determine the PDF model uncertainty.

Parameter	Nominal Value	Variations up	Variations down
$Q^2_{min}(\mathrm{GeV}^2)$	10	7.5	12.5
f_s	0.4	0.32	0.48
$m_c(\mathrm{GeV})$	1.41	1.37	1.45
$m_c(\mathrm{GeV})$	4.5	4.25	4.75
$\mu^2_{F,0}(\mathrm{GeV}^2)$	1.9	1.6	2.2

Finally, scale uncertainty is estimated to account for missing higher perturbative orders on the PDFs by varying the renormalization and factorization scales.

The PDFs resulting from the fits to the double differential dijet measurements are shown in Figure 6.1 along with the different uncertainty contributions. The total uncertainty in the PDFs is obtained by adding together the parametrization uncertainty and the sum in quadrature of the fit, model, and scale uncertainty contributions.

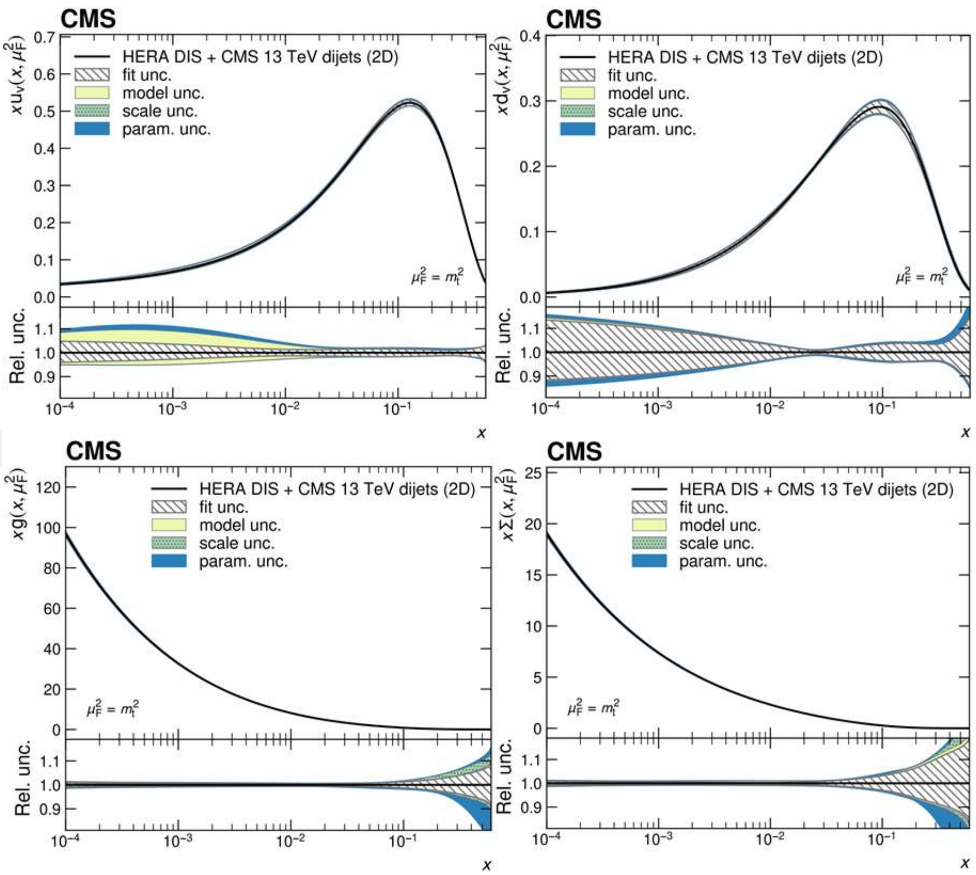

Figure 6.1. PDFs obtained in a fit to HERA DIS together with CMS 2D dijet data. The PDFs of the up and down valence quarks (top row), of the gluon (bottom left), and of the total sea quarks (bottom right) are shown as a function of the fractional parton momentum x at a factorization scale equal to the top quark mass. The colored bands correspond to the individual contributions to the total PDF uncertainty.

To evaluate the impact of the present measurements on the PDF determination, fits are performed using only the HERA DIS data. Figure 6.2 shows the resulting distributions along with the associated Hessian fit uncertainty. For a consistent comparison, the DIS data are fitted using the same PDF parametrizations as the fits including the dijet measurements.

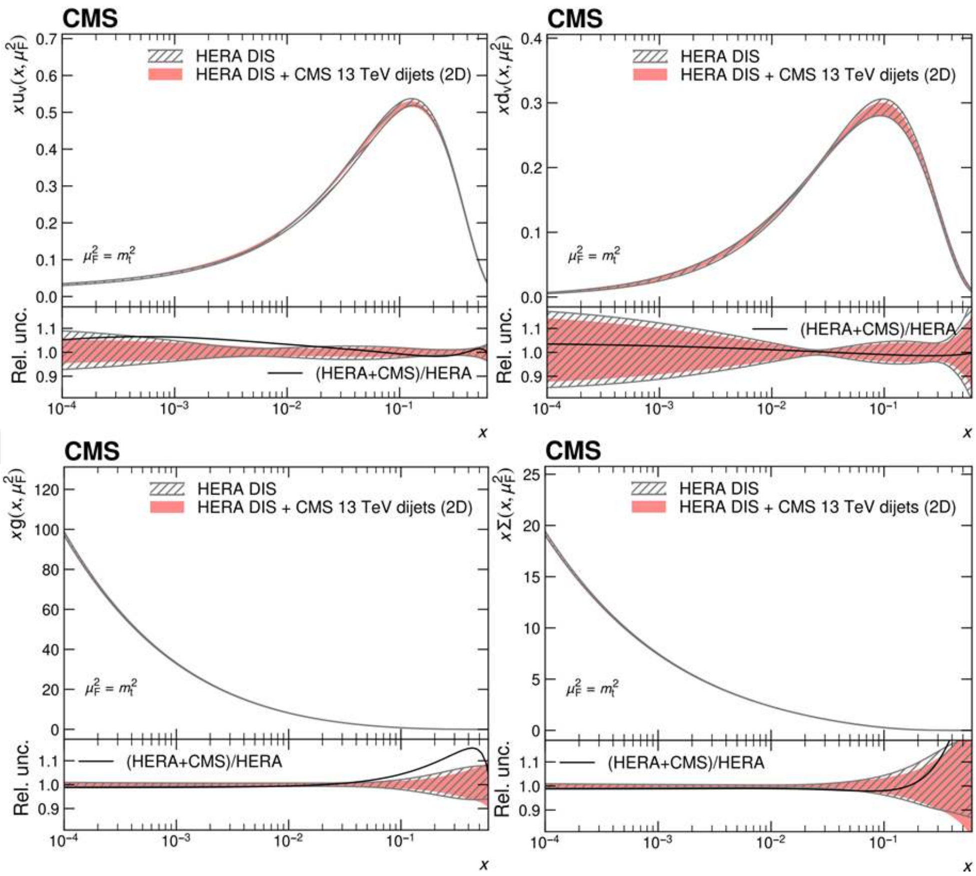

Figure 6.2. PDFs obtained in a fit to HERA DIS together with CMS 2D dijet data compared to a fit to HERA DIS data alone. The PDFs of the up and down valence quarks (top row), of the gluon (bottom left), and of the total sea quarks (bottom right) are shown as a function of the fractional parton momentum x at a factorization scale equal to the top quark mass. The hatched bands indicate the Hessian fit uncertainty and are shown in the lower panels as a relative uncertainty with respect to the corresponding central values. The solid black line shows the ratio between the fitted central values.

The inclusion of the dijet measurements results in an overall reduction of the PDF fit uncertainty. The distributions obtained with and without the CMS data appear largely compatible even within the fit uncertainty alone, with the notable exception of the gluon PDF at parton momentum fractions $x > 0.1$, where the inclusion of the double differential dijet cross sections results in an increased gluon contribution.

For the PDF determinations presented above, the value of the strong coupling constant at the scale of the Z boson mass was fixed to $\alpha_s(m_Z) = 0.118$. To assess the sensitivity of the dijet cross-section prediction to this parameter, additional fits are performed where $\alpha_s(m_Z)$ is determined simultaneously with the PDFs, using the same PDF parametrizations. Since in particular the gluon distribution is known to be correlated with $\alpha_s(m_Z)$, this enables a consistent treatment of PDFs and the strong coupling constant. The value of $\alpha_s(m_Z)$ obtained simultaneously in a fit to the double differential dijet cross sections is:

$$\alpha_s(m_Z) = 0.1201 \pm 0.0012\ (fit) \pm 0.0008\ (scale) \pm 0.0008\ (model) \\ \pm 0.0005\ (param.) = 0.1201\ \pm 0.0021\ (total)$$

The central value obtained from the dijet measurement is about one standard deviation away from the world average value of $\alpha_s(m_Z) = 0.1179 \pm 0.0009$ and larger by about 1.6 standard deviations than those obtained from measurements of the inclusive jet cross section at 13 TeV (CMS Collaboration, 2022).

7. CONCLUSION

A double differential measurement of the dijet cross section using p-p collision data from the CMS experiment at $\sqrt{s} = 13$ TeV in 2016, equal to integrated luminosity of $36.3\ \text{fb}^{-1}$, has been presented. Dijet cross sections are measured as a function of dijet invariant mass $m_{1,2}$ in five regions of the two leading jets' maximum absolute rapidity $|y|_{max}$.

The dijet spectrum is reconstructed by combining data from multiple triggers across a wide kinematic range, starting at a value of 200 GeV in $m_{1,2}$. The measurement is performed in five rapidity bins ($|y|_{max}$) of the whole phase space, going up to 2.5. Detector effects are corrected through a two-dimensional unfolding procedure, which corrects migrations both in mass and rapidity bins.

The primary causes of uncertainty for the cross-section are estimated. The JES is shown to be the main source of uncertainty, ranging from 5% to more than 25% at high rapidities.

The unfolded cross-sections and fixed-order predictions are compared in order to assess the effect of the double differential dijet cross-section measurements on PDF determination. Correction factors are determined and applied to the fixed-order theory to consider HAD, MPI, and EW contributions.

After using fixed-order theory calculations as a reference point for comparison, it is observed that theory predictions are mostly in agreement with the observed data in most regions of phase space. However, some noteworthy differences from the fixed-order theory are observed, especially at high rapidities where the uncertainty induced by PDFs is greatest.

Focusing on the larger jet radius and the dijet invariant mass $m_{1,2}$ as an observable, the PDFs of the proton are determined in fits to the dijet measurements together with deep-inelastic scattering data from the HERA experiments. The inclusion of the presented measurements leads to an improved determination of the

PDFs compared to fits to HERA data alone. In particular, the uncertainty in the gluon PDF at medium to large fractional momenta x of the proton is reduced. The results obtained from the double differential measurements are shown to be compatible within the estimated uncertainties.

REFERENCES

Alekhin S. et al., 2015. "HERAFitter: open source QCD fit project", Eur. Phys. J. C 75 (2015) 304, doi:10.1140/epjc/s10052-015-3480-z, arXiv:1410.4412.

Alekhin S., Blumlein J., Moch S., and Placakyte R., 2017. "Parton distribution functions, ¨ αs, and heavy-quark masses for LHC Run II", Phys. Rev. D 96 (2017) 014011, doi:10.1103/PhysRevD.96.014011, arXiv:1701.05838.

Alioli S. et al., 2011. "Jet pair production in POWHEG", JHEP 04 (2011) 081, doi:10.1007/JHEP04(2011)081, arXiv:1012.3380.

Alioli S., Nason P., Oleari C., and Re E., 2010. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", JHEP 06 (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

Altarelli G. and Parisi G., 1977. "Asymptotic freedom in parton language", Nucl. Phys. B 126 (1977) 298, doi:10.1016/0550-3213(77)90384-4.

Alwall J. et al., 2014. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", JHEP 07 (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

ATLAS Collaboration, 2011a. "Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass Energy with the ATLAS detector". In: Eur. Phys. J. C (2011) 71: 1512 DOI: 10.1140/epjc/s10052-010-1512-2.

ATLAS Collaboration, 2014. "Measurement of dijet cross sections in pp collisions at 7 TeV centre-of mass energy using the ATLAS detector", JHEP 05 (2014) 059, doi:10.1007/JHEP05(2014)059, arXiv:1312.3524.

ATLAS Collaboration, 2017. "Measurement of the inclusive jet cross-sections in proton–proton collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector", JHEP 09 (2017) 020, doi:10.1007/JHEP09(2017)020, arXiv:1706.03192.

ATLAS Collaboration, 2018. “Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at √s = 13 TeV with the ATLAS detector”, JHEP 05 (2018) 195, doi:10.1007/JHEP05(2018)195, arXiv:1711.02692.

Bahr M. et al., 2008. “HERWIG++ physics and manual”, ¨ Eur. Phys. J. C 58 (2008) 639, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

Bahr M. et al., 2008. “HERWIG++ Physics and Manual”, Eur.Phys.J.C58:639-707,2008, doi: 10.1140/epjc/s10052-008-0798-9, arXiv: 0803.0883.

Bailey S. et al., 2021. “Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs”, Eur. Phys. J. C 81 (2021) 341, doi:10.1140/epjc/s10052-021-09057-0, arXiv:2012.04684.

Banfi A., Salam G. P., and Zanderighi G., 2010. “Phenomenology of event shapes at hadron colliders”, JHEP 06 (2010) 038, doi:10.1007/JHEP06(2010)038, arXiv:1001.4082.

Bellm J. et al., 2016. “Herwig 7.0/HERWIG++ 3.0 release note”, Eur. Phys. J. C 76 (2016) 196, doi:10.1140/epjc/s10052-016-4018-8, arXiv:1512.01178.

Bertone V. et al., 1977. “xFitter 2.0.0: An open source QCD fit framework”, PoS DIS2017 (2018) 203, doi:10.22323/1.297.0203, arXiv:1709.01151.Y. L. Dokshitzer, “Calculation of the structure functions for deep inelastic scattering and e+e− annihilation by perturbation theory in quantum chromodynamics”, Sov. Phys. JETP 46 (1977) 641.

Botje M., 2011. “QCDNUM: Fast QCD evolution and convolution”, Comput. Phys. Commun. 182 (2011) 490, doi: 10.1016/j.cpc.2010.10.020, arXiv:1005.1481.

Britzger D. et al., 2019. “Calculations for deep inelastic scattering using fast interpolation grid techniques at NNLO in QCD and the extraction of αs from HERA data”, Eur. Phys. J. C (2019) 845, doi:10.1140/epjc/s10052-019-7351-x, arXiv:1906.05303.

Britzger D. et al., 2022. “NNLO interpolation grids for jet production at the LHC”, Eur. Phys. J. C 82 (2022) 930, doi:10.1140/epjc/s10052-022-10880-2, arXiv:2207.13735.

Britzger D., Rabbertz K., Stober F., and Wobisch M., 2012. “New features in version 2 of the fastNLO project”, in 20th International Workshop on Deep-Inelastic Scattering and Related Subjects, p. 217. 2012. arXiv:1208.3641. doi:10.3204/DESY-PROC-2012-02/165.

Buckley A. et al., 2015. “LHAPDF6: parton density access in the LHC precision era”, Eur. Phys. J. C 75 (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.

Cacciari M. et al., 2004. “The tt cross-section at 1.8 TeV and 1.96 TeV: a study of the systematics due to parton densities and scale dependence”, JHEP 04 (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.

Cacciari M., Salam G. P., and Soyez G., 2008. “The anti-kT jet clustering algorithm”, JHEP 04 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

Cacciari M., Salam G. P., and Soyez G., 2012. “FastJet user manual”, Eur. Phys. J. C 72 (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

Catani S., de Florian D., Grazzini M., and Nason P., 2003. “Soft gluon resummation for Higgs boson production at hadron colliders”, JHEP 07 (2003) 028, doi:10.1088/1126-6708/2003/07/028, arXiv:hep-ph/0306211.

Chen X. et al., 2022. “NNLO QCD corrections in full colour for jet production observables at the LHC”, JHEP 09 (2022) 025, doi:10.1007/JHEP09(2022)025,arXiv:2204.10173.

CMS Collaboration, 2008. “The CMS experiment at the CERN LHC”, JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

CMS Collaboration, 2011. “Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s}$ = 7 TeV”, Phys. Lett. B 700 (2011) 187, doi:10.1016/j.physletb.2011.05.027, arXiv:1104.1693.

CMS Collaboration, 2013a. “Measurements of differential jet cross sections in proton-proton collisions at √s=7 TeV with the CMS detector”. In: Phys. Rev. D 87, 112002. DOI: 10.1103/PhysRevD.87.112002.

CMS Collaboration, 2014. “Measurement of the ratio of inclusive jet cross sections using the anti-kT algorithm with radius parameters R = 0.5 and 0.7 in pp collisions at √s = 7 TeV”, Phys. Rev. D 90 (2014) 072006, doi:10.1103/PhysRevD.90.072006, arXiv:1406.0324.

CMS Collaboration, 2015. “Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid”, CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015.

CMS Collaboration, 2016a. “Event generator tunes obtained from underlying event and multiparton scattering measurements”, Eur. Phys. J. C 76 (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

CMS Collaboration, 2016b. “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of tt at √s = 8 and 13 TeV”, CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016.

CMS Collaboration, 2016c. “Measurement of the double-differential inclusive jet cross section in proton-proton collisions at √s = 13 TeV”, Eur. Phys. J. C 76 (2016) 451, doi:10.1140/epjc/s10052-016-4286-3, arXiv:1605.04436.

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

CMS Collaboration, 2017b. “Measurement of the triple-differential dijet cross section in proton-proton collisions at √s = 8 TeV and constraints on parton distribution functions”, Eur. Phys. J. C 77 (2017) 746, doi:10.1140/epjc/s10052-017-5286-7, arXiv:1705.02628.

CMS Collaboration, 2017c. “Particle flow reconstruction and global event description with the CMS detector”, JINST 12 (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

CMS Collaboration, 2017d. “The CMS trigger system”, JINST 12 (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

CMS Collaboration, 2017e. “Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at √s =8 TeV and cross section ratios to 2.76 and 7 TeV”, Journal of High Energy Physics volume 2017, Article number: 156 (2017), doi: 10.1007/ JHEP03(2017)156, arXiv: 1609.05331

CMS Collaboration, 2018. “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018”, CMS Detector Performance Note CMS-DP-2020-019, CERN, 2020.

CMS Collaboration, 2019. “Performance of missing transverse momentum reconstruction in proton-proton collisions at √s = 13 TeV using the CMS detector”, JINST 14 (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

CMS Collaboration, 2020a. “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, Eur. Phys. J. C 80 (2020) 4, doi: 10.1140/epjc/s10052-019-7499-4, arXiv: 1903.12179

CMS Collaboration, 2020b. “Performance of the CMS Level-1 trigger in proton-proton collisions at √s = 13 TeV”, JINST 15 (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

CMS Collaboration, 2021a. “Development and validation of HERWIG 7 tunes from CMS underlying-event measurements”, Eur. Phys. J. C 81 (2021) 312, doi:10.1140/epjc/s10052-021-08949-5, arXiv:2011.03422.

CMS Collaboration, 2021b. “Precision luminosity measurement in proton-proton collisions at √ s = 13 TeV in 2015 and 2016 at CMS”, Eur. Phys. J. C 81 (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

CMS Collaboration, 2022. “Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at √s = 13 TeV”, JHEP 02 (2022) 142, doi:10.1007/JHEP02(2022)142, arXiv:2111.10431.

Currie J. et al., 2017. “Precise predictions for dijet production at the LHC”, Phys. Rev. Lett. 119 (2017) 152001, doi:10.1103/PhysRevLett.119.152001, arXiv:1705.10271.

Currie J. et al., 2018. “Jet cross sections at the LHC with NNLOJET”, PoS LL2018 (2018) 001, doi:10.22323/1.303.0001, arXiv:1807.06057.

Dittmaier S., Huss A., and Speckner C., 2012. “Weak radiative corrections to dijet production at hadron colliders”, JHEP 11 (2012) 095, doi:10.1007/JHEP11(2012)095,arXiv:1210.0438.

Fruhwirth R., Waltenberger W. and Vanlaer P., 2007. “Adaptive vertex fitting”, Journal of Physics G: Nuclear and Particle Physics, Volume 34, Number 12 (2007).

GEANT4 Collaboration, 2003. “GEANT4—a simulation toolkit”, Nucl. Instrum. Meth. A 506 (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

Gehrmann T. et al., 2018. “Jet cross sections and transverse momentum distributions with NNLOJET”, PoS RADCOR2017 (2018) 074, doi:10.22323/1.290.0074, arXiv:1801.06415.

Gehrmann-De Ridder A. et al., 2019. “Triple differential dijet cross section at the LHC”, Phys. Rev. Lett. 123 102001, doi:10.1103/PhysRevLett.123.102001, arXiv:1905.09047.

Gribov V. N. and Lipatov L. N., 1972. “Deep inelastic ep scattering in perturbation theory”, Sov. J. Nucl. Phys. 15 (1972) 438.

H1 and ZEUS Collaborations, 2010. “Combined measurement and QCD analysis of the inclusive e ± p scattering cross sections at HERA”, JHEP 01 (2010) 109, doi:10.1007/JHEP01(2010)109, arXiv:0911.0884.

H1 and ZEUS Collaborations, 2015. “Combination of measurements of inclusive deep inelastic e±p scattering cross sections and QCD analysis of HERA data”, Eur. Phys. J. C 75 (2015) 580, doi:10.1140/epjc/s10052-015-3710-4, arXiv:1506.06042.

H1 and ZEUS Collaborations, 2022. “Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions”, Eur. Phys. J. C 82 (2022) 243, doi:10.1140/epjc/s10052-022-10083-9, arXiv:2112.01120.

Hou T. J. et al., 2021. “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC”, Phys. Rev. D 103 (2021) 014013, doi:10.1103/PhysRevD.103.014013, arXiv:1912.10053.

Kiminsu, U., 2022. “Measurement Of Double Dıfferantial Dijet Mass Cross Section With The CMS Detector at √s = 13 TeV”, Doktora Tezi, 2022, Çukurova Üniversitesi FBE

Kluge T., Rabbertz K., and Wobisch M., 2006. “fastNLO: Fast pQCD calculations for PDF fits”, in Proceedings, 14th International Workshop on Deep-Inelastic Scattering (DIS 2006), p. 483. Tsukuba, Japan, 20-24 April, 2006. arXiv:hep-ph/0609285. doi:10.1142/9789812706706_0110.

Nason P., 2004. “A new method for combining NLO QCD with shower Monte Carlo algorithms”, JHEP 11 (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

NNPDF Collaboration, 2017. “Parton distributions from high-precision collider data”, Eur. Phys. J. C 77 (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

Pei Y., Biswas S., Fussell D. S., Pingali K, 2019. “An Elementary Introduction to Kalman Filtering”, doi:10.48550/arXiv.1710.04055, arXiv:1710.04055.

Pumplin J. et al., 2001. “Uncertainties of predictions from parton distribution functions. ii. The Hessian method”, Phys. Rev. D 65 (2001) 014013, doi:10.1103/PhysRevD.65.014013, arXiv:hep-ph/0101032.

S. Frixione, P. Nason, and C. Oleari, 2007. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, JHEP 11 (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

Schmitt S., 2012. “TUnfold: an algorithm for correcting migration effects in high energy physics”, JINST 7 (2012) T10003, doi:10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.

Seymour M. H. and Siodmok A., 2013. “Constraining MPI models using σeff and recent Tevatron and LHC underlying event data”, JHEP 10 (2013) 113, doi:10.1007/JHEP10(2013)113, arXiv:1307.5015.

Sjostrand T., Mrenna S., and Skands P. Z., 2008. “A brief introduction to PYTHIA 8.1”, ¨ Comput. Phys. Commun. 178 (2008) 852, doi: 10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

Thorne R. S. and Roberts R. G., 1998. “An ordered analysis of heavy flavor production in deep inelastic scattering”, Phys. Rev. D 57 (1998) 6871, doi:10.1103/PhysRevD.57.6871, arXiv:hep-ph/9709442.

Thorne R. S., 2006. “A variable-flavor number scheme for NNLO”, Phys. Rev. D 73 (2006) 054019, doi:10.1103/PhysRevD.73.054019, arXiv:hep-ph/0601245.

Thorne R. S., 2012. “Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections”, Phys. Rev. D 86 (2012) 074017, doi:10.1103/PhysRevD.86.074017, arXiv:1201.6180.

Workman R. L. et al., 2022. “Review of particle physics”, Prog. Theor. Exp. Phys. 2022 (2022) 083C01, doi:10.1093/ptep/ptac097.

CURRICULUM VITAE

İbrahim Soner ZORBAKIR, studied Physics at Çukurova University and graduated in 2012. For his MSc, he contributed several detector performance studies for CMS experiments. He focused jet physics on his MSc degree. After finishing his MSc in 2015, he started his Ph.D. with the hadronic jet physics group in 2017. During this, he also performed a QCD analysis by using CMS collision data which was taken in 2016.