Measurement of Inclusive Jet Production at $\sqrt{s}$ = 2.76 TeV with the CMS Detector and calibration of FFTJet Algorithm

by

Terence Libeiro, B.S., M.S.

A Dissertation

In

PHYSICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Approved

Igor Volobouev
Chair of Committee

Richard Wigmans

Charley Myles

Nural Akchurin

Sung Won Lee

Mark A. Sheridan
Dean of the Graduate School

May, 2015

Acknowledgements

I would like to thank Dr. Igor Volobouev for the opportunity to work on the FFTJet algorithm. I appreciate his generous advice and patience, his enthusiasm and his continuous support. His efforts were invaluable to the completion of my dissertation and my research.
I am thankful to Dr. Sung-Won Lee for introducing me to the CMS SMP group, which allowed me to pursue my research with the help and supervision of the experts on the topic. Their timely suggestions and input made it possible to advance and conclude this research in a short timescale. I am especially grateful to Maxime Gouzevitch, Mikko Voutilainen, Marguerite Tonjes, and Samantha Dooling; they contributed significantly to this work and helped me tackle some very tedious issues. I benefited greatly from the experience of the post-docs and scientists, here at Fermilab. They were an amazing source of ideas and always gracious and helpful.
It was a great privilege and an unique experience to work in a large international collaboration, such as CMS. There are many people who were involved in the successful running of this experiment; without them this work would not be possible.
I would also like to thank my family for their steadfast support and encouragement throughout this process.

Contents

Abstract

The double-differential inclusive jet cross section is measured as a function of jet transverse momentum (p_T) and rapidity (y) with pp collision data from the CMS experiment at the LHC. The data were collected in 2013 at 2.76 TeV center-of-mass energy with an integrated luminosity of 5.43 pb^{-1}. Jets are reconstructed with anti-k$_T$ algorithm with the scale parameter R = 0.7 and measured over the p_T range from 74 to 592 GeV and $|y|$ range from 0 to 3. The reconstructed inclusive jet spectrum is unfolded to the particle-level and compared to the next-to-leading (NLO) order theoretical predictions. Next, a double ratio, $\frac{\sigma^{2.76TeV}_{Data}}{\sigma^{2.76TeV}_{Theory}} / \frac{\sigma^{8TeV}_{Data}}{\sigma^{8TeV}_{Theory}}$, is calculated using the inclusive jet cross section at 8 TeV center-of-mass energy. The results are presented with the experimental and theoretical uncertainties associated with the measurements.

The second part of this thesis describes the calibration of the FFTJet algorithm. This is a new approach to jet reconstruction which uses pattern recognition techniques to identify jet-like objects in the energy flow of the event. It allows the user to apply a jet shape model and then reconstruct the jet energies in a subsequent step. In this study, the calibration of this jet reconstruction algorithm is performed for a range of jet scale parameters (R = 0.2 – 1.0) using collision data and simulated samples. The calibration procedure is validated by reconstructing a benchmark process, semi-leptonic $t\bar{t}$ decay, and the results are presented.

List of Tables

List of Figures

List of Abbreviations

ALICE: A Large Ion Collider Experiment

ATLAS: A Toroidal LHC ApparatuS

CDF: Collider Detector at Fermilab

CERN: Conseil Européen pour la Recherche Nucléaire

CMS: Compact Muon Solenoid

CMSSW: CMS SoftWare

DAQ: Data AcQuisition system

DIS: Deep Inelastic Scattering

ECAL: Electromagnetic CALorimeter

FFTJet: Fast Fourier Transform Jet Algortithm

GeV: Giga-electron volt

HCAL: Hadronic CALorimeter

HERA: Hadron Electron Ring Accelerator

HLT: High Level Trigger

JEC: Jet Energy Correction

JES: Jet Energy Scale

JER: Jet Energy Resolution

L1: Level-1 Trigger

LO: Leading Order

LHC: Large Hadron Collider

LHCb: Large Hadron Collider Beauty experiment

MC: Monte Carlo

MET: Missing E_T

MPF: Missing energy Projection Factor

NLO: Next-to-Leading Order

NNLO: Next-to-Next-to-Leading Order

NP: Non Perturbative

PDF: Parton Distribution Function

PF: Particle Flow

PU: PileUp

pQCD: pertubative Quantum ChromoDynamics

QCD: Quantum ChromoDynamics

QED: Quantum ElectroDynamics

QFT: Quantum Field Theory

SM: Standard Model

SUSY: SUperSYmmetry

SLAC: Stanford Linear Accelerator Center

TeV: Tera-electron volt

Chapter 1

Introduction

"Enlightenment is man's release from his self-incurred tutelage."

– Kant

In the western intellectual tradition, the scientific investigation of nature began with the Greek philosophers during the fourth and fifth centuries BC. From the four fundamental elements composed by Empedocles to the indivisible and infinite atoms by Democritus and Leucippus, many ideas were proposed to explain the substance and phenomena observed in the natural world. These ideas co-existed with the homocentric ideals of the Skeptics and teleologically driven Aristotelean philosophy, which eventually prevailed through the decline of Greek civilization and into the middle ages. In the late medieval, early modern period the interest in the natural world was revived with the intellectual revolution of the Renaissance. The world view changed from one in which the universe is isolated, static and geocentric to an ever-changing, heliocentric one with planet trajectories governed by the mechanical laws of motion. With the desire to understand nature's secrets, mathematicians, philosophers and natural scientists employed the scientific method to explain the physical world.

The scientific method was established as a rational basis for validating theories by means of precise and systematic empirical tests. It is assumed that natural laws are unchanging and can be discovered with experimentation, therefore any results of repeated investigations are consistent and reproducible. This has allowed scientists to build on the

works of successive generations to produce an increasingly sophisticated and comprehensive picture of nature. Today, the scientific realm extends from the microscopic world of quantum mechanics to the astronomical phenomena observed on the cosmic scales.

The nature of matter described in terms of its most basic constituents and their interactions is studied by *particle physics*. At the subatomic scales, matter is composed of protons, neutrons and electrons. Protons and neutrons are stable composite particles made up of tightly bound quarks and gluons (collectively called partons). In order to study the properties and interactions of these elementary particles, protons (or electrons) are accelerated to very high energies and collided. The particles generated in the collisions are measured with an experimental apparatus called a detector, made up of large numbers of sub-components designed to detect and measure different types of particles.

The quarks and gluons produced from such high energy interactions are unstable. They spontaneously fragment and produce a large number of stable particles which vary in their properties and composition. A collection of these particles is called a jet. It is observed as a collimated spray of energetic particles, interacting with the detector. Due to the large number of collision events and the complexity of processing involved, the reconstruction of jets is performed with jet reconstruction algorithms in the physics analysis. The jets stand as a proxy to partons produced in the proton-proton collisions, and provide a means to understand the interactions between the elementary particles.

In this thesis, the production and the reconstruction of jets in pp collisions is studied. The motivation behind this study is twofold. First, to understand the physics of the strong interaction which dominantly contributes to the production of jets in hadron-hadron collisions. Second, to improve the software tools used to reconstruct jets. In the first running phase of the Large Hadron Collider (LHC), in the years 2010-13, proton-proton collision data was collected at four different center-of-mass energies. The data collected is used to test the well established theories, in this higher energy regime, and look for evidence of new phenomena. The high luminosity at the LHC also presents a considerable challenge to

isolate and investigate physics processes of interest from the noise and pile-up interactions. A measurement of jet cross section serves as a way to test the theory at a new, higher scale of interaction. The development of jet reconstruction software attempts to provide a safer, more resilient jet-data reconstruction in view of some unique problems faced at the LHC.

In the first part, the inclusive jet production cross section is measured at 2.76 TeV center-of-mass energy ($\sqrt{s}$). The inclusive jet production is one of the basic tests of the predictions of the theory of strong interactions, Quantum Chromodynamics (QCD). It is previously untested in the kinematic range covered by this analysis. A detailed study of the cross section data at this center-of-mass can yield independent constraints on the theory parameters and complement the measurements made at $\sqrt{s} = 7$ and 8 TeV.

In the second part, a new algorithm for reconstructing jets is described. It uses an unique pattern recognition technique which provides robust jet identification. In addition, a multi-scale jet reconstruction is tested for the first time at CMS, in which jets can be reconstructed at multiple scales. In this study, we perform a preliminary calibration of the algorithm and derive scale-dependent jet energy corrections. The upcoming run at the LHC[1] will be at higher center-of-mass energies, and will produce an ever larger number of "uninteresting" pile-up events. The precise jet identification and reconstruction, afforded by the FFTJet algorithm, will be most useful in such noisy environments. Furthermore, advanced jet reconstruction techniques such as FFTJet are promising in the searches for SUSY particles and boosted resonances where commonly used jet algorithms lack sensitivity.

The thesis is organized as follows. In the next chapter, a description of the CMS experiment and the LHC accelerator complex is provided. The theory of jet production with QCD is covered in chapter 3, which includes the basics of Standard Model physics, deep inelastic scattering of partons, and a brief overview of the simulation software used in the analysis. Chapter 4 is devoted to the inclusive jet cross section measurement at $\sqrt{s} = 2.76$ TeV and the cross section ratio with the $\sqrt{s} = 8$ TeV measurement. Chapter 5 documents the

[1] Run II at the LHC

calibration of the FFTJet algorithm. It contains a brief introduction to the main concept followed by the detailed description of the calibration procedure.

In this thesis, the natural units are employed using the convention $c = 1$ and $\hbar = 1$.

Chapter 2

The Large Hadron Collider and the Compact Muon Solenoid

The LHC at CERN (*Conseil Européen pour la Recherche Nucléaire*) is a particle collider that accelerates and collides charged particles, such as protons and heavy ions, at high energies. The LHC accelerator facility is located in a 27 km long tunnel, 50 to 175 m underground across the Swiss-French border near Geneva, Switzerland. It is designed to achieve higher collision energies and luminosities than any other hadron collider facility in the world. With the design center-of-mass energy of 14 TeV and instantaneous luminosity of $10^{-34}\text{cm}^{-2}\text{s}^{-1}$, the LHC can probe parton-parton interactions at 1 TeV or higher. This pushes the reach of experimental high energy physics further into the tera-electron volt regime and allows us to investigate physics at higher energies than ever before. The acceleration chain begins by ionizing hydrogen atoms to obtain protons or lead atoms to obtain Pb ions. The protons (or lead ions) are accelerated in incremental steps through LINAC, Booster, Proton Synchrotron (PS), Super Proton Synchrotron (SPS) and, finally, the LHC. A schematic of the accelerator complex is shown in fig. 2.1. The protons injected from SPS at 450 GeV can be accelerated to the energy of 7 TeV and circulated in 2808 separate bunches around the LHC. The protons are accelerated at the rate of 0.5 MeV per turn using radiofrequency cavities. The circular orbit of the rapidly accelerating particles is maintained by a compensating magnetic field of up to 8.3 T, provided by helium-cooled superconducting dipole magnets.

In November 2009, the LHC reached 1.18 TeV with protons and became the most powerful accelerator in the world, surpassing Tevatron's previous record of 0.98 TeV. In

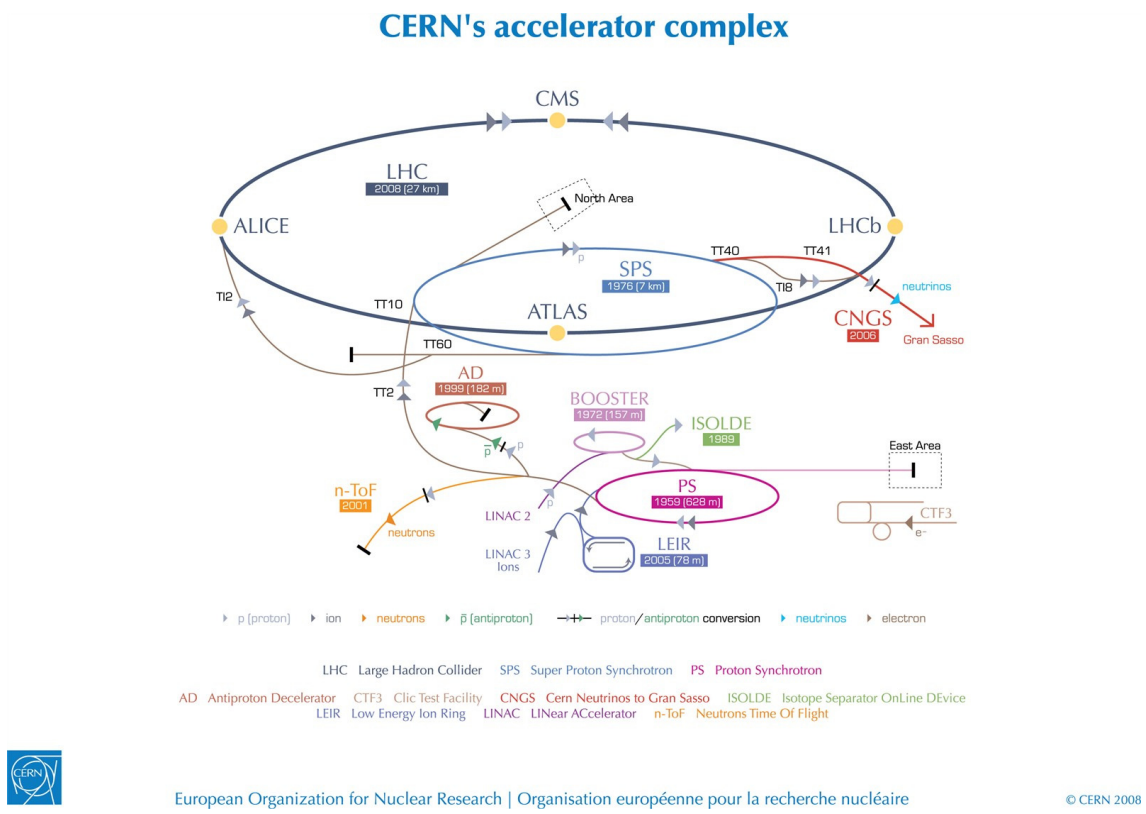


Figure 2.1: The LHC accelerator complex at CERN [2].

the subsequent run in the years 2011 and 2012, the center-of-mass energies reached 7 TeV and 8 TeV, respectively. In 2013, the LHC collided pPb at $\sqrt{s} = 5.02$ TeV and pp at $\sqrt{s} = 2.76$ TeV as a part of its heavy ion program.

Four main experiments are set up at the LHC: CMS [3], and ATLAS [8] are the two general purpose detectors, while LHCb [9] is designed for b-physics and ALICE [10] is designed for lead-ion studies. In addition, two smaller experiments, TOTEM and LHCf study forward physics[1].

The luminosity delivered by the LHC is used to calculate the cross sections of any

[1]
CMS - Compact Muon Solenoid
ATLAS - A Toroidal LHC Apparatus
LHCb - Large Hadron Collider beauty
ALICE - A Large Ion Collider Experiment
TOTEM - TOTal Elastic and diffractive cross section Measurement
LHCf - Large Hadron Collider forward

Quantity	Value for pp (pPb)	Unit
Center of mass energy	2.76 (5.02)	TeV
Instantaneous luminosity	10^{31} (10^{29})	$cm^{-2}s^{-1}$
Number of bunches per beam	510 (338)	-
Colliding bunches	504 (296)	-
Crossing angle	170.0 (145.0)	μrad
Revolution frequency	11	kHz
β^*	1100.0 (80.0)	cm
γ	1471	-
ε_n	0.14[1]	μm
Magnetic field	1.63[2]	T

Table 2.1: The LHC parameters in 2013 collisions.

physics process under investigation. It is determined using the following relation:

$$\mathscr{L} = \frac{\gamma f k_B N_p^2}{4\pi\varepsilon_n\beta^*}F \tag{2.1}$$

where,

γ is the Lorentz factor,

f is the revolution frequency,

k_B is the number of bunches,

N_p is the number of protons per bunch,

ε_n is the beam emittance,

β^* is the beta function,

F is the reduction factor due to beams crossing at a finite angle.

An overview of the LHC operating parameters in 2013 is given in table 2.1.

[1] calculated from the nominal value, 75 μm.

[2] The value is calculated for $\sqrt{s}$ =2.76 TeV from the nominal magnetic field value of the LHC.

2.1 The CMS detector

CMS is a multipurpose detector located at the LHC interaction point P5, situated 100 m underground near the village of Cessy, France [3,4]. The CMS detector is shaped as a cylinder 14.6 m in diameter and 21.6 m in length, weighing over 12,500 tons. The primary motivation behind the design of the CMS detector is to be sensitive to the manifestations of the electroweak symmetry breaking due to the "Higgs" mechanism. The design of the detector also makes it possible to probe wide range of processes, which is useful for checking the Standard Model predictions and looking for exotic physics beyond the Standard Model such as SUSY (Supersymmetry) or extra dimensions. The design specifications of the LHC would provide a seven-fold increase in energy and a hundred-fold increase in integrated luminosity compared to the previous experiments in hadron collider physics. At design energies and luminosities, on average 20 inelastic collisions (or pile-up collisions) were expected to be superimposed on the primary hard interaction of interest. The prediction of pile-up has increased since then [2]. A single bunch crossing every 50 ns generates about 1000 charged particles in the detector [3]. In order to accurately measure the energies of particles and distinguish the product(s) of the hard interaction from the pile-up, CMS employs a high resolution tracking system and high granularity calorimeters with good time resolution, all enclosed in a 4 T magnetic field. This results in millions of electronic channels which are synchronized to read out data every bunch crossing.

Some of the notable features of CMS are:

- Muon system consists of the inner tracker and the outer muon chambers. The two sub-detectors combined can identify muons over a wide range of momenta with high precision.

- Inner tracking system provides high resolution measurement of the charge and mo-

[2]The pile-up observed at $\sqrt{s}$ =8 TeV running in 2012 was about 20. At $\sqrt{s}$ =14 TeV this is predicted to increase, exceeding design expectations.

[3]collision rate will increase to 40 MHz or a bunch collision every 25 ns in 2015.

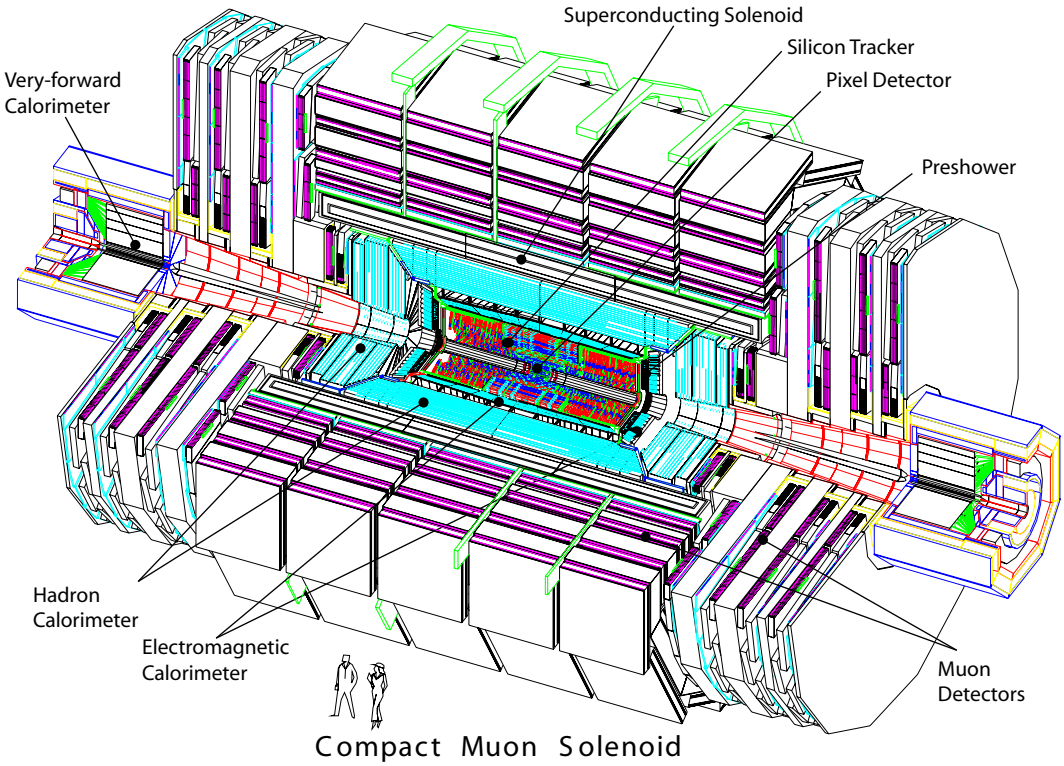


Figure 2.2: A cross section view of the CMS detector [3].

mentum of the charged particles. Interaction vertices are reconstructed using the charged tracks and used to identify displaced vertex or pile-up interactions in the event.

- Electromagnetic calorimeter provides good energy resolution for investigating electroweak processes with electrons and photons in the final state.

A detailed description of the individual components of the CMS detector is given the following sections. A cross section view of the detector is shown in fig. 2.2.

Coordinate system

The following coordinate system is used by the CMS experiment.

- Origin: Located at the nominal interaction point.
- X axis: Radially inward towards the center of the LHC.

- Y axis: Upward perpendicular to the LHC.
- Z axis: Along the beam direction (westward).
- ϕ: Azimuthal angle is measured from the x-axis in the xy plane.
- θ: Polar angle is measured from the z-axis.

The coordinate system is set up in the right handed convention, i.e., positive directions in all coordinates are linked by the right hand rule. The center-of-mass frame of parton interactions differs from the lab frame in which the measurements are made by the detector. The following invariant variables are commonly used for experimental measurements, when considering longitudinally boosted systems: Δy, $\Delta\eta$ in the massless limit, and p_T. They are defined as:

- Pseudorapidity:
$$\eta = -ln(tan\frac{\theta}{2}), \tag{2.2}$$
depends only on the polar angle of the particle (θ).

- Rapidity:
$$y = \frac{1}{2}ln\left(\frac{E+p_z}{E-p_z}\right). \tag{2.3}$$
Rapidity is commonly used in jet cross section measurements as it takes into account the energy dependent fragmentation of jets.

- E_T, p_T: Momentum in the transverse plane is conserved, while an imbalance in the longitudinal components is possible due to different longitudinal momenta of the interacting partons.

Magnet

The CMS experiment uses a superconducting solenoid magnet which creates a magnetic field of up to 4 T. It is 12.5 m in length and 6 m in diameter, and surrounds the tracking

system and the calorimeters (see fig. 2.2). The magnet has the mass of 220 tons and 2.6 GJ of stored energy at full current. The magnetic field outside the solenoid is returned through a 10,000 ton yoke made up of five barrel-shaped wheels and two three-disk end caps. A notable property of this magnet is that, due to its high energy density (11.6 KJ/kg), the mechanical structure of the detector is deformed by about 0.15% after energizing.

Tracking

The tracking system is designed to provide a high precision measurement of the momenta of charged particles produced in the proton collisions. Due to the high instantaneous luminosity delivered by the LHC, a large number of pile-up interactions are expected. Considering the high flux of particles produced in every bunch crossing, the tracking system is built to meet the following requirements:

1. Fine granularity and fast response time for prompt and accurate reconstruction of charged-particle trajectories.

2. Light material footprint to limit particle interactions.

3. Radiation hardness to operate at high radiation level for extended period.

The tracking system is 5.8 m in length and 2.5 m in diameter. It covers the barrel and end cap regions with $|\eta|$ <2.5. It is made up of inner high resolution *pixel detectors* combined with outer light and radiation hard silicon *strip trackers*. Pixel detectors make up three barrel layers, 4.4 cm to 10.2 cm away from the beam axis, and two disks in the end cap. The pixel detectors contain 1440 segmented silicon sensor modules which are read out with 66 million electronic channels. The size of a pixel cell is 100×150 μm^2 which provides the spatial resolution of 20 μm. The particle position in the $r-\phi$ direction is reconstructed with 15 μm resolution. The pixel detectors are tilted at 20° in a turbine-like geometry to induce charge sharing, which improves the overall spatial resolution.

The strip trackers are divided into 10 layers in the barrel (10 to 110 cm away from the beam) and 9 disks in the end cap. In all, the strip trackers have 15,148 modules covering an area of 198 m^2, making it the largest silicon strip tracker ever built. The spatial resolution of the detector is 230 μm in z direction and 23 to 34 μm in $r-\phi$ direction. The strip trackers have coarser resolution than the pixel detector and are subject to reduced particle flux as they are placed farther away from the interaction point. The momentum resolution of the tracking system is 1-2% in the central $|\eta| < 1.6$ region, for 10-100 GeV particles.

2.1.1 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is made up of lead tungstate ($PbWO_4$) crystals. Lead tungstate is dense (8.28 g/cm^3) with short radiation length (X_0) of 0.89 cm and Moliere radius of 2.2 cm. It covers the rapidity region $|\eta|$ <3 in the detector. The crystals are radiation resistant and emit 80% of the light signal in 25 ns, the duration of a typical bunch crossing at the LHC, making them a good choice of material for the electromagnetic calorimeter. The energy resolution of the ECAL is determined by electron test beam measurement to be $\sim$1% for the energy region of interest for Higgs searches. A schematic of ECAL is given in fig. 2.3. The ECAL consists of three parts :

1. EB (or ECAL in barrel) is made up of 61,200 crystals which cover the pseudorapidity range of $|\eta|$ <1.4. The crystals are tapered with the narrow ends facing the interaction point. The crystals are 22×22 mm^2 at the narrow end and 26×26 mm^2 at the wider end. The length of each crystal is 23 cm which corresponds to 25.8 X_0. The light produced by the interacting particles is collected by avalanche photodiodes.

2. EE (or ECAL in endcaps) is made up of 14648 crystals. Each crystal is 22 cm (24.7 X_0) in length with a vacuum phototriode at the end used for collecting the light.

3. ES (or the preshower detector) is installed before the EE in the endcap region 1.6 < $|\eta| < 2.6$. Its primary purpose is to discriminate between photon pairs produced in

π^0 decays and the photon(s) produced in all other processes. This is achieved by identifying photon pair(s) produced with an invariant mass close to the mass of π^0.

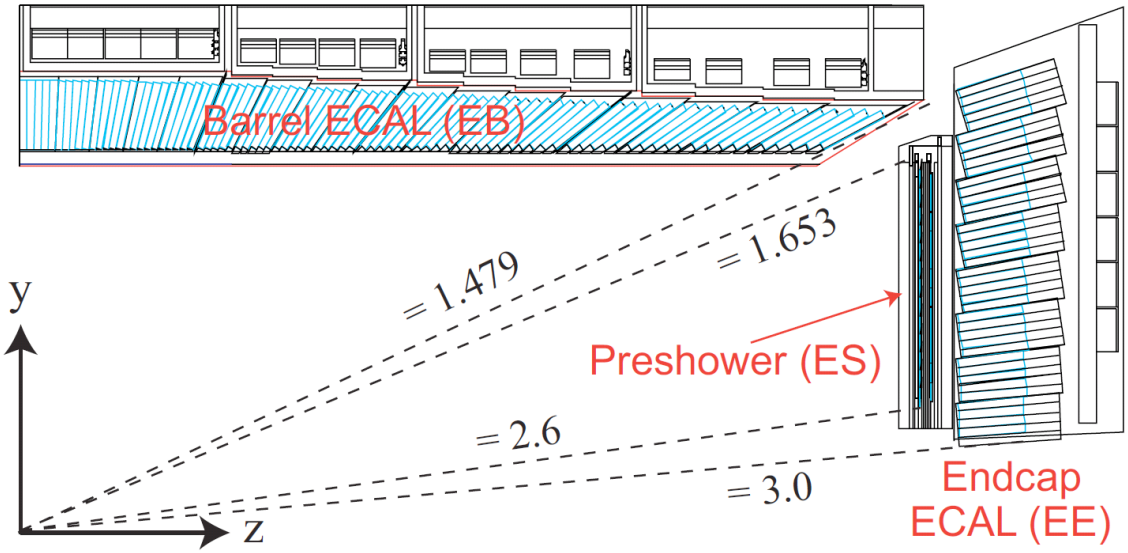


Figure 2.3: The electromagnetic calorimeter at CMS [4].

2.1.2 Hadronic calorimeter

The hadronic calorimeter (HCAL) is useful for the measurement of hadronic jets. It is a sampling calorimeter made from alternating layers of brass (absorber) and scintillating material. The brass plates are 79 mm thick with 9 mm gaps for scintillators . The scintillating material is made from plastic, doped with a wavelength shifting agent and read out by hybrid photodiode. The HB (or HCAL in barrel) extends from 1.77 to 2.95 m away from the beam axis and covers the pseudorapidity range $|\eta|$ <1.3. The HE (or HCAL in endcaps) is placed outside the EE detector and covers a range of 1.3< $|\eta|$ <3. The HO (or the outer hadron calorimeter) is placed outside the superconducting magnet to catch the tail end of highly energetic showers which traverse through the calorimeter and the solenoid. It gives a measure of shower leakage from the calorimeter and improves the energy resolution measured by the detector. The HO covers the $|\eta|$ <1.3 region and extends the total number of interaction lengths of the barrel calorimeter to 11.8. The forward calorimeter (HF) is

an extension of HCAL, located at 11.2 m from the interaction point, and covers the region $3< |\eta| <5$. A visual representation of HCAL is given in fig. 2.4.

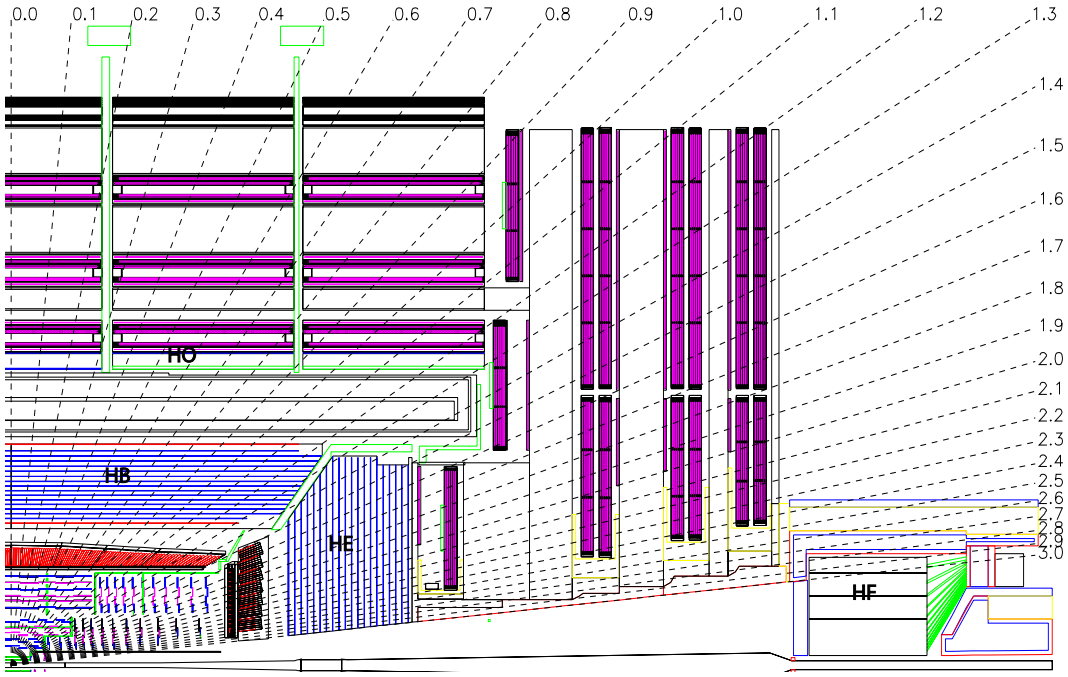


Figure 2.4: The hadronic calorimeter at CMS [3].

2.1.3 The Muon system

The muon system is designed to identify muons and to make precise measurements of their momentum. The high magnetic field inside the solenoid magnet and the flux-return yoke outside make good muon momentum resolution possible. The muon system is made up of gas detectors placed outside the HCAL, in the barrel and the end cap regions (fig. 2.5). Three different detector types are used:

1. Drift tubes (DT) surround the HO in the barrel region in four concentric cylinders. The cylinders have drift chambers made up of rectangular cells, containing positively charged wires in the center and filled with a gas mixture (Ar+CO_2). A muon traversing a chamber ionizes the gas. The negatively charged electrons produced in the process drift towards the center ionizing more gas. An avalanche develops. The

electrons created in the avalanche drift towards the anode and produce current proportional to the original number of ions. The barrel region experiences the lowest particle flux and covers the largest surface area.

2. Cathode strip chambers (CSC) are in the endcaps where the particle flux is higher. They are made up of multi-wire proportional chambers which perform better at high particle rates and non-uniform magnetic field.

3. Resistive plate chamber are gaseous, parallel plate detectors located in the endcaps. They are used for triggering, and complement the information from CSCs and DTs.

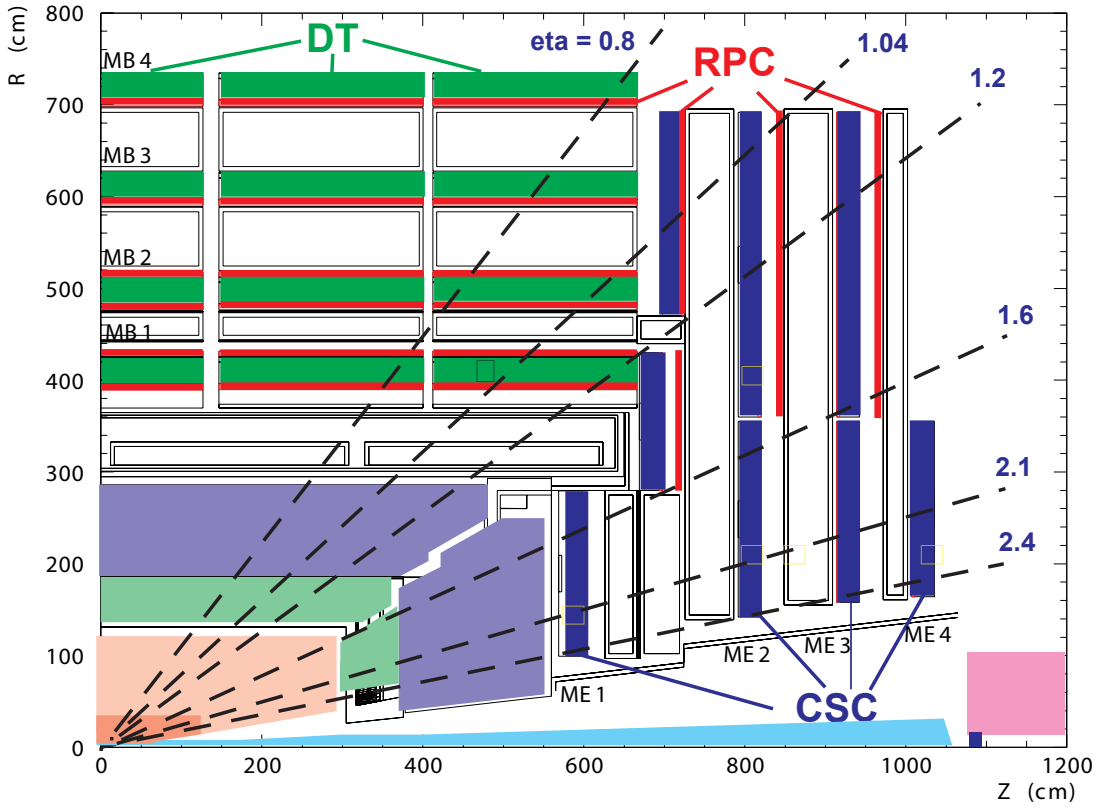


Figure 2.5: The muon system at CMS consisting of drift tubes, resistive plate chambers, and cathode strip chambers [4].

2.1.4 Data acquisition and trigger system

The LHC can provide pp collisions at the rate of 40 MHz or once per 25 ns. The high frequency of collisions generates a large amount of data which has to be analyzed and stored. In order to extract the most interesting events from the data and to reduce the event rate, triggers are applied. The trigger system is divided into two stages which are applied in sequence: Level-1 (L1) and High-Level Trigger (HLT). The maximum output rates of L1 and HLT triggers are 100 kHz and 100 Hz, respectively. The two triggers, applied in conjunction, reduce the event rate by a factor of 10^6. The L1 trigger makes a quick decision to accept or reject an event based on selective processing of the event data. If the L1 trigger decision is positive, the data from all the sub-detectors of CMS is retrieved for processing with the HLT software.

The L1 trigger is implemented at an electronic level, with some components installed on the CMS detector and others located in the control room situated about 90 m away from the detectors. The data processed by the L1 trigger comes from coarsely segmented regions of the calorimeter and the muon system. The global trigger logic is determined from the individual inputs shown in fig. 2.6. At the HLT stage, the data is processed in the computing farm with more detailed event reconstruction and sophisticated selection algorithms. The data is selected with specific trigger conditions on physics objects such as photons, leptons or jets, and the passing events are written to disk.

The data acquisition system (DAQ) relies on the input from the trigger system to read out event information. The events passing both triggers are fully processed and stored for physics analysis. The size of a reconstructed event record from the CMS detector is $\approx$ 1 MB. At the L1 rate of 100 kHz, the volume of data processed by the DAQ is $\approx$ 100 GB/s. The events passing the HLT are reconstructed on the basis of complete information from the sub-detectors, and transferred to a local disk for permanent storage. The DAQ architecture is shown in fig. 2.7. The data read out from the front-end detectors is consolidated into an event by the builder network and written to disk. The triggering

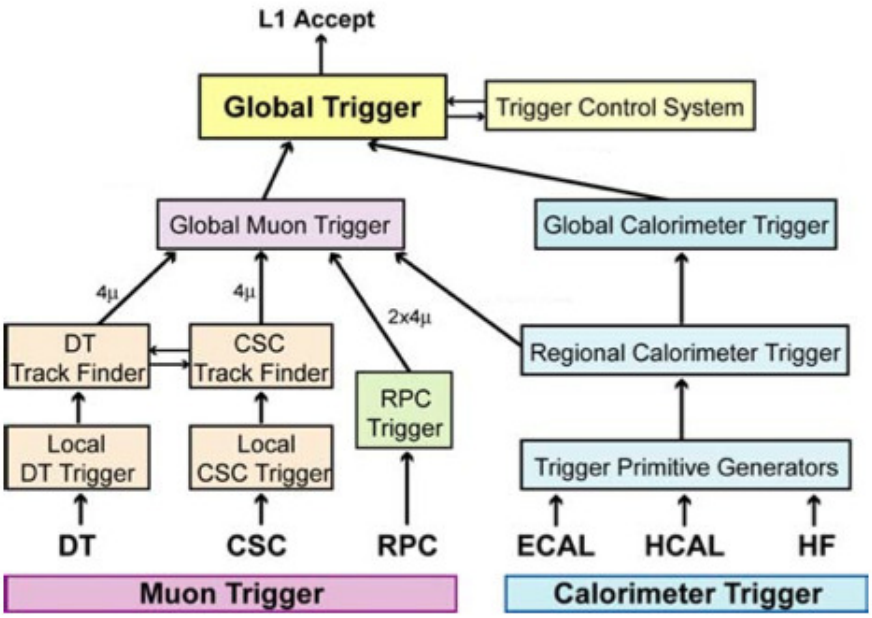


Figure 2.6: The L1 trigger architecture is shown with all the contributing sub-detectors. The final decision to accept or reject the event is made based on the trigger objects delivered by the Global Muon Trigger and the Global Calorimeter Trigger [3].

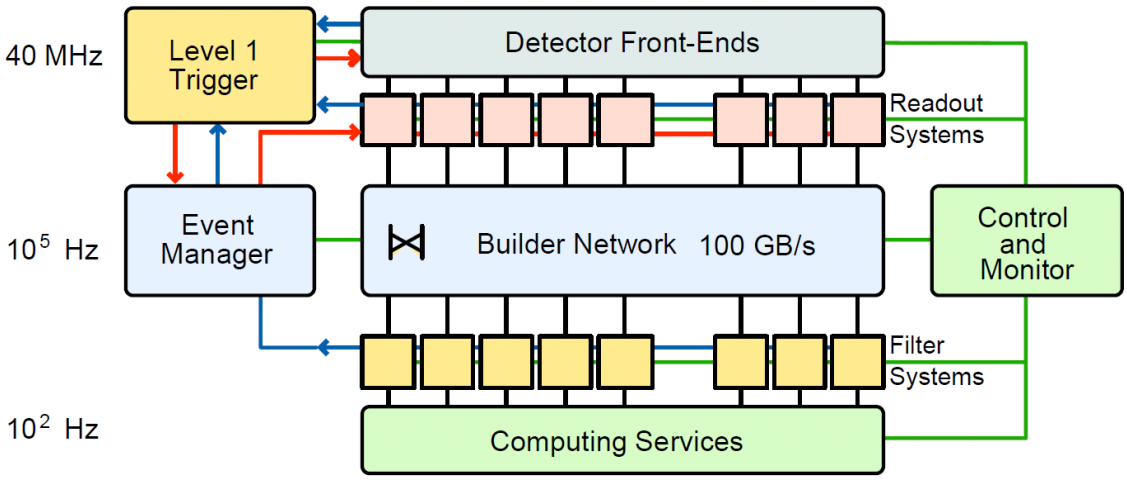


Figure 2.7: Architecture of data acquisition system at CMS [4].

systems are used to modulate the readout operation based on the processing bandwidth available.

Chapter 3

The Standard Model and Jet Physics

The production of jets at the LHC is primarily due to the strong interactions between partons inside the colliding protons. The interacting partons scatter away from the parent hadron and fragment into jets of particles which travel, on average, in the same direction as the original parton. In order to understand the physics of jets, we start with the general theory of particles and interactions called the Standard Model. In this chapter, the fundamental concepts of the Standard Model are presented. The theory of strong interactions, QCD, is discussed with a focus on the theory of inclusive jet production. Finally, a brief description of Monte Carlo (MC) simulation techniques used for making predictions, which may be used to confront the experimental data, is given.

3.1 The Standard Model

Our understanding of elementary particles and interactions has evolved tremendously in the past century. The first elementary particles to be discovered were electrons which are subject to the electromagnetic interaction via photons. Owing to many concurrent developments in technology in the 1960s, especially the development of accelerator technology and extensive use of computers for data collection and processing, detailed studies of nucleon structure became possible. This led to the discovery of quarks and a better understanding of the strong nuclear interaction. The Standard Model of particle physics was proposed as a unified framework, made up of all known elementary particles and the electromagnetic, weak and strong nuclear interactions [11–17].

3.1.1 Particles in the Standard Model

The elementary particles are described by their fundamental properties, some of the most important among them are: mass, electric charge and spin. The spin quantum number is defined in the units of $\hbar$. The particles are divided into two classes: half-integer spin particles are called fermions, and integer spin particles are called bosons. The fermions fall into two categories: leptons and quarks. The list of currently known elementary particles and their properties is given in tables 3.1 and 3.2.

3.1.2 Interactions in the Standard Model

The interactions in the Standard Model can be described by the emission and absorption of field quanta. As the exchanged field quanta carry energy and momentum, the interaction between two particles is also called *exchange force* and the exchanged particle *force carrier*. In all three classes of interactions (electromagnetic, weak and strong) spin-1 bosons are the force carriers (see table 3.2). The strength and range of the different forces vary considerably. The strong force is the strongest, the electromagnetic force is weaker with the longest range, while the weak force is the weakest and has the shortest range. A recent addition to this picture is the discovery of spin-0 Higgs boson which is postulated to explain the origin of mass [18, 19] of subatomic particles.

To date, the Standard Model is the most successful theory that ties many disparate phenomena observed in nature into a single coherent framework. It is a great feat of modern physics confirmed to high experimental precision. However, there are hints of new phenomena and particles which are as yet not included in the Standard Model. Recent discoveries in astronomy of dark matter and dark energy are not predicted or explained by the theory. And the gravitational force, described by general relativity, has not been formulated in terms of the principles of quantum mechanics.

3.2 QCD

QCD is the non-Abelian gauge theory of strong interactions [20, 21]. It describes the interactions of partons inside the nucleons with the exchange of massless gauge bosons. The partons have color charge with three degrees of freedom - red, green and blue - which are represented in the *SU(3)* symmetry group[1]. The idea of quantum theory of strong interactions has its origins in quantum electrodynamics (QED), a renormalizable field theory of point particles which successfully describes electromagnetic interactions. In the early days of development of QCD, this theory was adopted to explain the data coming from experiment; such as the general principles of scattering amplitudes. The study of the strong interactions was aided greatly by the introduction of high energy accelerators which could reach multi-GeV energy regime and probe hadrons with higher resolution than ever before. The first evidence of quarks came from electron-proton scattering experiments performed at SLAC. The data supported the existence of point-like constituents inside the proton [22]. The quark model, originally proposed by Gell-Mann [23] and Zweig [24], was used to explain the composition of hadrons. It involved fermionic quarks, possessing a fraction of the proton charge and momentum, which interacted with the exchange of the force carrying bosons called gluons. The properties of quarks and gluons are given in table 3.1. The final major contribution to this picture was made by the discovery of asymptotic freedom and confinement.

3.2.1 Lagrangian of QCD

The perturbative calculations of QCD are made with the Lagrangian density ($\mathscr{L}$) [25]

$$\mathscr{L} = \mathscr{L}_{classical} + \mathscr{L}_{gauge-fixing} + \mathscr{L}_{ghost}. \tag{3.1}$$

[1] The word "color" is used here as conceptual term, and has no relation to the color seen by the human eye.

Fermions									
	Generation			Charge	Color	Spin	Mass		
	1	2	3				1	2	3
Leptons	ν_e	ν_μ	ν_τ	0	-	1/2		-	
	e	μ	τ	-1			0.5	105.7	1777
Quarks	u	c	t	+2/3	r, g, b	1/2	2.4	1270	173,200
	d	s	b	-1/3			4.8	104	4200

Table 3.1: Fermions in the Standard Model [1]. Mass is in the units MeV.

Bosons					
Interaction	Particle	Charge	Spin	Mass	Relative Strength
Electromagnetic	γ	0	1	0	10^{-2}
Weak	W	+1,-1	1	80.4	10^{-13}
	Z	0	1	91.2	
Strong	g	0	1	0	1

Table 3.2: Bosons in the Standard Model [1]. Mass is in the units GeV.

The classical gauge invariant Lagrangian density ($\mathscr{L}_{classical}$) is given by[2]

$$\mathscr{L}_{classical} = -\frac{1}{4}F^A_{\alpha\beta}F_A^{\alpha\beta} + \sum_{flavors} \bar{q}_a(i\not{D} - m)_{ab}q_b. \tag{3.2}$$

The first term with field tensor ($F^A_{\alpha\beta}$) is derived from the gluon field $\mathscr{A}^A_\alpha$,

$$F^A_{\alpha\beta} = \partial_\alpha \mathscr{A}^A_\beta - \partial_\beta \mathscr{A}^A_\alpha - g f^{ABC}\mathscr{A}^B_\alpha \mathscr{A}^C_\beta, \tag{3.3}$$

where the indices A, B, C run over the eight color degrees of freedom of the gluon field. The first two terms in eq. 3.3 are similar to the field terms for photon in QED, while the introduction of the third non-Abelian term gives rise to the self-interacting property of the gluon. The interaction terms in eq. 3.2 ($\sum$...) describe the quark-gluon interactions with quarks of mass m and charge q. Finally, back in eq. 3.1, the gluon propagator cannot be defined in the perturbation theory without fixing the choice of gauge [25]. This introduces two more terms into the Lagrangian. The gauge-fixing term ($\mathscr{L}_{gauge-fixing}$) fixes the covariant gauges with the gauge parameter λ,

$$\mathscr{L}_{gauge-fixing} = -\frac{1}{2\lambda}(\partial^\alpha \mathscr{A}^A_\alpha)^2. \tag{3.4}$$

The ghost term ($\mathscr{L}_{ghost}$) is introduced to cancel unphysical degrees of freedom that propagate in the covariant gauges,

$$\mathscr{L}_{ghost} = \partial_\alpha \eta^{A\dagger}(D^\alpha_{AB}\eta^B). \tag{3.5}$$

where η^A is a complex scalar field which obeys Fermi statistics. A detailed discussion of the two gauge terms can be found in [25].

[2] $\not{D}$ is the covariant derivative

3.2.2 Asymptotic freedom and confinement

The strength of the strong interaction is high at low energy of interaction (Q) and decreases asymptotically approaching zero at high energy. The latter property of QCD is called asymptotic freedom. In mathematical terms, the asymptotic freedom results from the scale dependence of the strong coupling constant, $\alpha_s(Q)$.

In QCD, the calculation of an observable by a perturbation series in α_s introduces ultraviolet divergences. A renormalization scale (μ_R) is necessary to remove the divergences; this makes the observable dependent on an arbitrarily defined parameter. The requirement that the observable has to be independent of the renormalization scale produces the following renormalization group equation.

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s) \tag{3.6}$$

The β function can be written as a perturbative expansion

$$\beta(\alpha_s) = -b\alpha_s^2(1 + b'\alpha_s + b''\alpha_s^2 + \mathcal{O}(\alpha_s^3)), \tag{3.7}$$

where,

$$b = \frac{11n_c - 2n_f}{12\pi} = \frac{33 - 2n_f}{12\pi}, \tag{3.8}$$

n_f is the number of quark flavors with mass less than Q^2 and n_c is the number of color charges. The subsequent coefficients can be determined from higher order diagrams, such as the one loop approximation given in fig. 3.1. The solution in one loop approximation, neglecting higher order terms, is

$$\alpha_s(Q^2) = \frac{\alpha_s(Q_0^2)}{1 + b\alpha_s(Q_0^2)\ ln\frac{Q^2}{Q_0^2}}, \tag{3.9}$$

where Q_0 is a constant reference scale. In contrast to QED, gluons self-interact causing

the zero degree coefficient (b in eq. 3.9) to be positive in QCD. As a result, the value of α_s asymptotically decreases to zero as the interaction scale $Q^2 \to \infty$, while at small scales ($Q^2 \leq 1\ GeV^2$) it diverges to infinity. The scale dependence of α_s is determined from perturbative QCD (eq. 3.9), however its absolute value has to be calculated experimentally. A reference scale (Λ_{QCD}), below which the perturbative expansions in α_s (eq. 3.7, 3.9) are not meaningful, is found to be $\mathcal{O}(200\ MeV)$ from experiment. This scale marks the limit of perturbative QCD and the region below this limit ($\mathcal{O}(1\ GeV)$) is called the non-perturbative region. As observed by a probing particle, partons in a hadron are weakly bound in the perturbative region and strongly confined to the hadron in the non-perturbative region due to the varying strength of the strong coupling. The α_s is experimentally measured and reported as a scale dependent function, as well as a value at the reference scale M_Z (90 GeV) (fig. 3.2). The current world average of the strong coupling constant is $\alpha_s(M_Z) = 0.1185 \pm 0.0006$ [1].

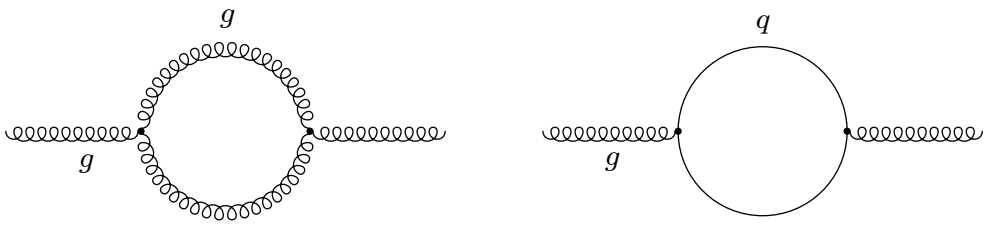


Figure 3.1: Diagrams contributing to the β function in the one loop approximation.

3.3 Deep inelastic scattering and jet production

Deep inelastic scattering (DIS) refers to the interactions which probe the constituent partons inside hadrons in high energy lepton-hadron or hadron-hadron collisions. It has been instrumental in measuring the momentum distribution of partons inside hadrons. The cross

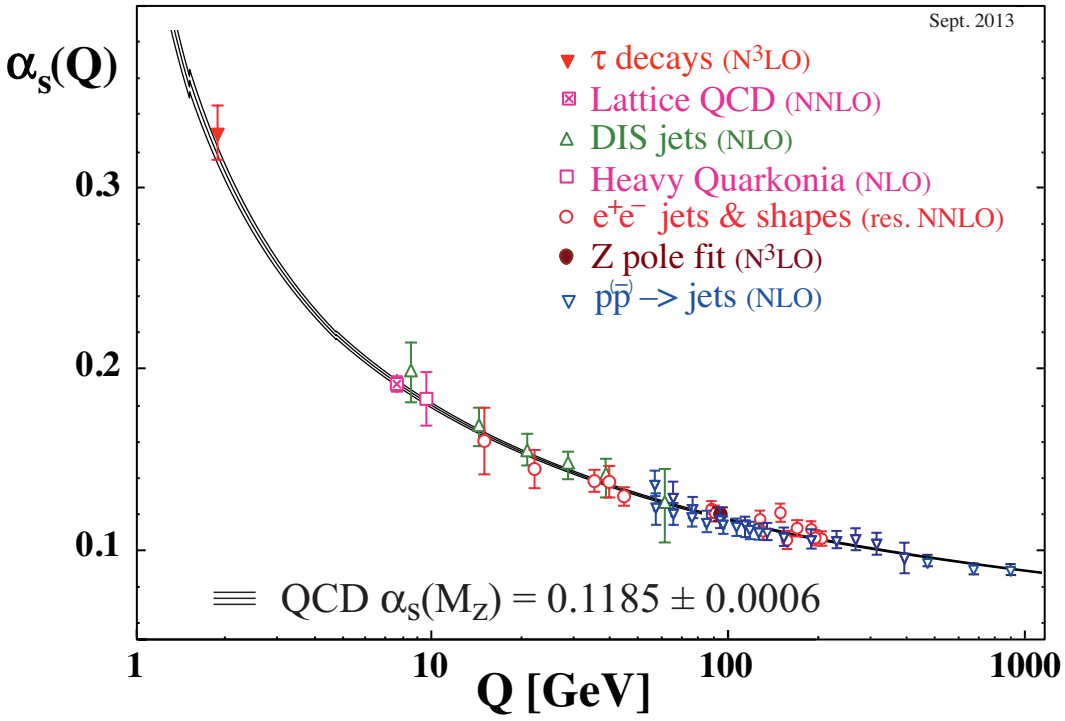


Figure 3.2: α_s plotted as a function of the interaction scale (Q) [1].

section calculations in DIS depend on the parton model that describes hadrons as bound states of gluons and quarks. An interaction between hadrons is the result of the interaction between constituent partons carrying a fraction of the momenta of their parent hadrons. The parton distribution function (PDF) describes the momentum distribution of partons inside the proton. It is the number density of a parton flavor that carries the momentum fraction x of the proton, at an interaction scale Q. The proton is made up of three valence quarks (uud) which make up its electric charge and quantum numbers, and a sea of virtual $q\bar{q}$ pairs and gluons. When probed at scale Q, the sea contains all quarks with mass $m_q < Q$. The valence quarks and sea quarks together make up for approximately 50% of the proton's momentum in the asymptotic limit ($Q^2 > 1\ GeV^2$), while the rest can be attributed to its gluon constituents. The parton distribution functions of the valence quarks, gluons and sea quarks are shown in fig. 3.3.

The cross section for an inelastic scattering process initiated by two hadrons with four-momenta P_1 and P_2 is calculated as

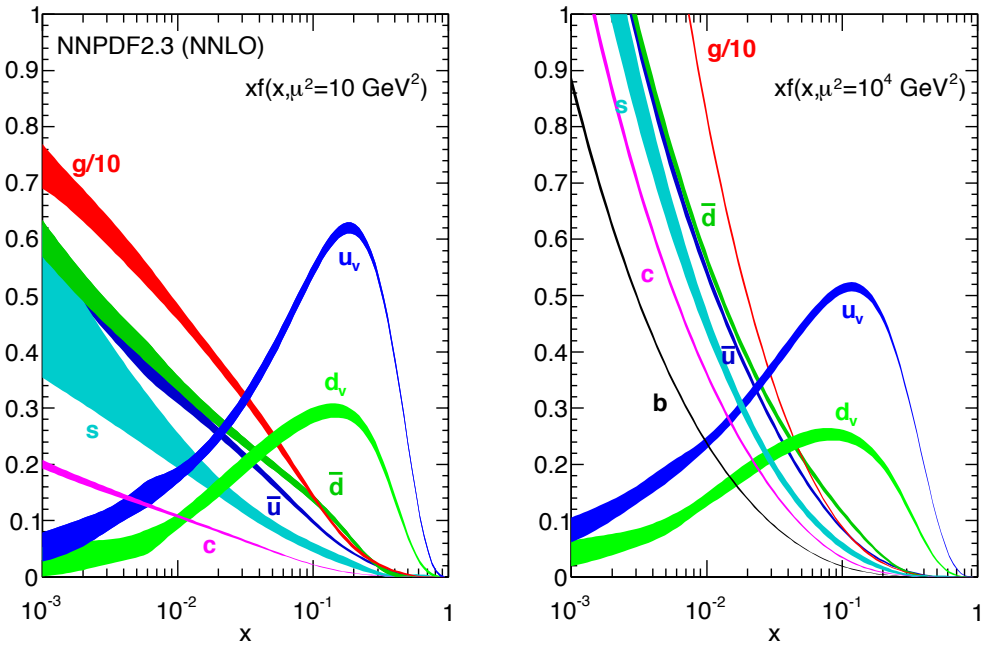


Figure 3.3: The PDFs of the proton measured by the NNPDF collaboration. The PDFs, derived from data, are evolved at the next-to-next leading order in the fitting procedure [1,5]. The PDFs are determined at the interaction scale $Q^2 = 10\ GeV^2$ (left) and $Q^2 = 10^4\ GeV^2$ (right).

$$\sigma(P_1,P_2) = \sum_{i,j} \int f_i(x_1,\mu_F^2) f_j(x_2,\mu_F^2) \hat{\sigma}_{ij}(x_1P_1, x_2P_2, \alpha_S(\mu_R^2), \mu_F^2) dx_1 dx_2, \tag{3.10}$$

where,

$f_{i,j}(x,\mu_F)$ is the parton distribution function for parton-flavor i,j, respectively,

x_1 is momentum fraction of hadron carried by parton 1 with flavor i,

x_2 is momentum fraction of hadron carried by parton 2 with flavor j,

$\hat{\sigma}_{ij}$ is the short-distance scattering cross section for partons 1 and 2 with flavors i and j, respectively,

μ_R is the renormalization scale,

μ_F is the factorization scale (defined in section 3.3.2).

Fig. 3.4 shows the corresponding hard scattering process.

Process	$\overline{\sum}\|\mathscr{M}\|^2/g^4$	Diagrams in fig. 3.5
$qq' \to qq'$	$\frac{4}{9}\frac{\hat{v}^2+\hat{u}^2}{\hat{t}^2}$	(a)
$q\bar{q}' \to q\bar{q}'$	$\frac{4}{9}\frac{\hat{v}^2+\hat{u}^2}{\hat{t}^2}$	(a)
$qq \to qq$	$\frac{4}{9}(\frac{\hat{v}^2+\hat{u}^2}{\hat{t}^2}+\frac{\hat{v}^2+\hat{t}^2}{\hat{u}^2})-\frac{8}{27}\frac{\hat{v}^2}{\hat{u}\hat{t}}$	(a)
$qq' \to q'\bar{q}'$	$\frac{4}{9}\frac{\hat{t}^2+\hat{u}^2}{\hat{v}^2}$	(a)
$q\bar{q} \to q\bar{q}$	$\frac{4}{9}(\frac{\hat{v}^2+\hat{u}^2}{\hat{t}^2}+\frac{\hat{t}^2+\hat{u}^2}{\hat{v}^2})-\frac{8}{27}\frac{\hat{u}^2}{\hat{v}\hat{t}}$	(b)
$q\bar{q} \to gg$	$\frac{32}{27}\frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}-\frac{8}{3}\frac{\hat{t}^2+\hat{u}^2}{\hat{v}^2}$	(c)
$gg \to q\bar{q}$	$\frac{1}{6}\frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}-\frac{3}{8}\frac{\hat{t}^2+\hat{u}^2}{\hat{v}^2}$	(d),(e)
$gq \to gq$	$-\frac{4}{9}\frac{\hat{v}^2+\hat{u}^2}{\hat{v}\hat{u}}+\frac{\hat{u}^2+\hat{v}^2}{\hat{t}^2}$	(f)
$gg \to gg$	$\frac{9}{2}(3-\frac{\hat{t}\hat{u}}{\hat{v}^2}-\frac{\hat{v}\hat{u}}{\hat{t}^2}-\frac{\hat{v}\hat{t}}{\hat{u}^2})$	(g)

Table 3.3: Matrix elements terms calculated for leading order subprocesses contributing to the inclusive jet production.

The short-distance cross section ($\hat{\sigma}_{ij}$) for the inelastic scattering process: $P_1 + P_2 \to P_3 + P_4$ is calculated as

$$\frac{d^3\hat{\sigma}}{d^2p_T dy} = \frac{1}{2s}\frac{1}{8\pi^2}\overline{\sum}|\mathscr{M}|^2\delta(\hat{v}+\hat{t}+\hat{u}), \tag{3.11}$$

The sum $\overline{\sum}$ denotes the average over the spins and sum over the colors. The four momenta of the initial- and final-state partons is denoted by P_1, P_2, P_3 and P_4. For convenience, we use Mandelstam variables defined as $\hat{v} = (P_1 + P_2)^2$, $\hat{t} = (P_1 - P_3)^2$ and $\hat{u} = (P_2 - P_3)^2$. The processes contributing to leading order calculations are shown in fig. 3.5 and the matrix elements ($\overline{\sum}|\mathscr{M}|^2$) are given in table 3.3.

3.3.1 Kinematics

A typical hadron scattering involves colliding two bunches of incoming partons, as illustrated in fig. 3.4. The partons have a spectrum of longitudinal momentum[3] determined by the parton distribution functions. The parton four momenta are represented as $p_1 = x_1 P_1$

[3] Along the direction of the proton

and $p_2 = x_2 P_2$, as fraction of the hadron momenta P_1 and P_2, respectively. The center of mass energy of the collision ($\sqrt{s}$) and the scale of the interaction in massless limit (Q) are defined as

$$\sqrt{s} = (P_1 + P_2), \;\; Q^2 = x_1 x_2 s,$$

The center of mass of the parton-parton scattering is usually longitudinally boosted with respect to the center of mass of the hadrons. Therefore, variables invariant with respect to longitudinal boosts such as rapidity change (Δy), transverse momentum (p_T) and azimuthal angle (ϕ) are used to measure cross sections. Commonly used experimental variables are listed and defined in section 2.1.

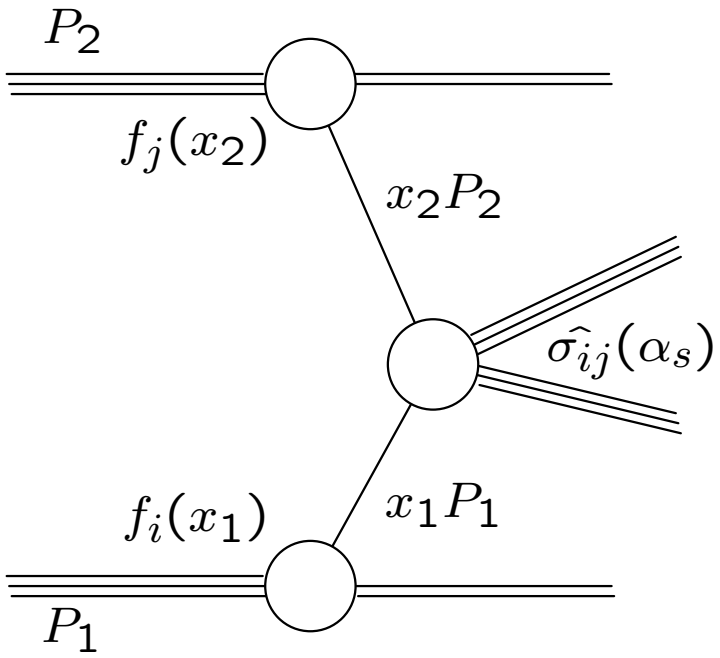


Figure 3.4: Hard scattering process involving two hadrons with momentum P_1 and P_2.

3.3.2 Factorization of the cross section

The leading order approximation to the jet cross section includes the primary hard scattering processes (diagrams shown in fig. 3.5). At higher orders of approximation, the contributions of long-distance interactions (or infrared divergences) occurring independently of the hard scattering have to be added to the leading order cross section. The factorization scale (μ_F) is an arbitrary parameter used to separate the short- and long-distance interactions. If the interacting parton is produced with energy lower than μ_F it is factored into the parton distribution functions and considered part of the hadron structure. Conversely, if the parton energy is higher than μ_F it is included in the short distance cross section calculation. The factorization scale is applied as a fraction of the interaction scale (Q/μ_F) in cross section calculations.

The long-distance interactions occur at low energy ($\mathcal{O}(\Lambda_{QCD})$) and high time scale ($t \sim \frac{1}{\Lambda_{QCD}}$) compared to the hard scattering[4], which makes the two processes factorizable to a high level of approximation. Due to the low energy of long-distance interactions, non-perturbative methods are used to calculate the long-distance contribution to the cross section. Hence, this contribution is called the non-perturbative correction. Factorization is the fundamental property of the theory and it is indispensable in any calculations made with QCD.

3.3.3 Jet fragmentation

The hard interaction produces a few energetic quarks and gluons as final-state particles. The principle of confinement requires quarks and gluons to exist only in color-neutral bound states. Hence, the quarks and gluons undergo two transient processes: fragmentation and hadronization, to generate a collection of stable particles called a jet.

Fragmentation refers to the formation of particles generated from the partons interacting

[4]The long-distance interactions such as hadronization of quarks (or gluons) occur at scales at which α_s becomes strong (Λ_{QCD}) . The distance and timescale of the hard interactions are of the order $\frac{1}{Q}$; as $Q \gg \Lambda_{QCD}$ these are short-distance small-timescale interactions described by perturbation theory.

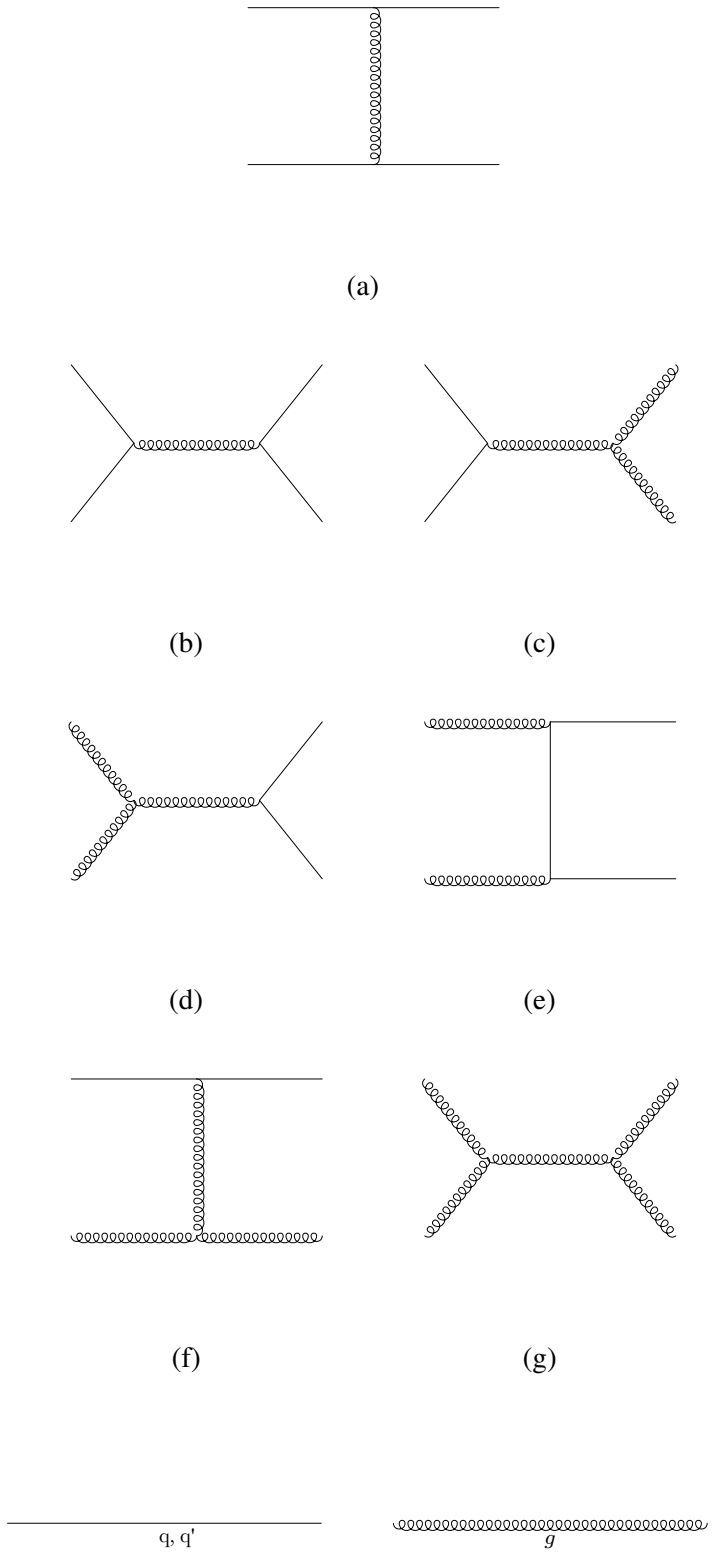


Figure 3.5: Leading order processes contributing to the inclusive jet production.

in the inelastic scattering. As the partons try to escape the proton, the energy density of the gluon field increases to the point such that $q\bar{q}$ pair production is possible. The quark-antiquark pairs are produced successively, accompanied by radiative gluon emissions until the energy density falls below the production threshold. This is followed by hadronization, a process in which the partons produced in the fragmentation process combine to form stable hadrons.

The partons emit virtual particles (quarks and gluons) which undergo successive splittings to produce more gluons and quark-antiquark pairs. At the leading order in α_s three processes contribute to parton splitting: gluon radiation ($q \rightarrow q + g$), gluon splitting ($g \rightarrow g + g$) and pair production ($g \rightarrow q + \bar{q}$). The probabilities of parton splitting, and therefore the fragmentation of parton, vary with the momentum and the flavor of the parton. The evolution of the fragmentation process is determined in the perturbative regime ($\mathcal{O}(1\ GeV)$) using DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations, which account for momentum-dependent parton splitting. The splitting processes at the leading order are given in fig. 3.6.

The momentum distribution $F^h(x)$ for hadron h in a jet is given by the fragmentation function

$$F^h(x,s) = \sum_i \int_x^1 \frac{dz}{z} C_i(z, \alpha_s(s)) D_i^h(x/z, s), \tag{3.12}$$

where,

i is the parton,

D_i^h is the probability of producing hadron h from parton i ,

$C_i(z, \alpha_s(s))$ is the coefficient function of the process,

x is the momentum fraction of the initial parton carried by the hadron ,

z is the momentum fraction of the incoming parton carried by the outgoing parton in the splitting process,

s is the scale of the hard interaction.

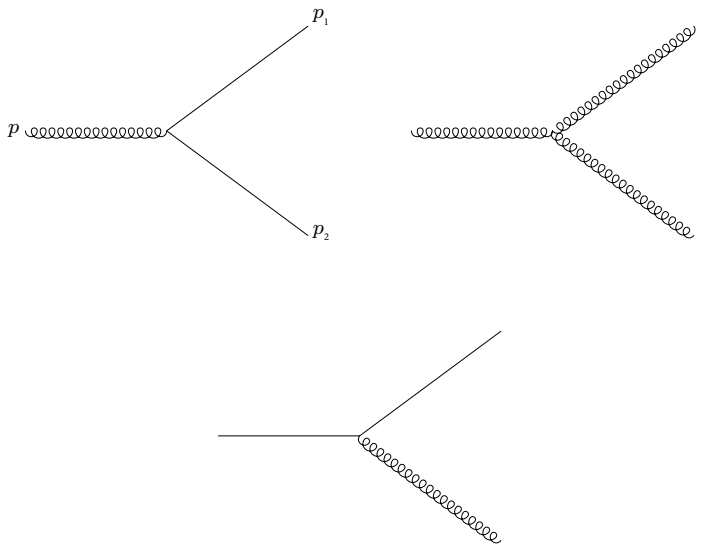


Figure 3.6: The three leading order parton splitting processes in QCD. Parton p splits into p_1 and p_2, with p_1 carrying the leading fraction, z, of the momentum of parton p.

The hadronization process occurs at long time scale ($t \sim \frac{1}{\Lambda_{QCD}}$) and low momentum transfers ($\mathcal{O}(\Lambda_{QCD})$). Since this is below the perturbative limit ($\mathcal{O}(1\ GeV)$), non-perturbative methods such as lattice QCD are used for modeling the process.

There are two commonly used approaches to explain the fragmentation and hadronization processes, the Cluster model [26] and the String model [27]. In the Cluster model, gluons are split into $q\bar{q}$ pairs followed by the recombination of the quarks with the adjacent anti-quarks to produce color-neutral hadrons. The hadrons are clustered together and allowed to decay isotropically into lighter particles as allowed by the available phase space. The String model assumes a relativistic string stretched between the initial interacting partons. The string represents a color flux tube with potential energy proportional to its length. As the distance between the quarks increases beyond the linear confinement limit, the string potential increases up to the level at which a quark-antiquark pair can be produced. The string breaks due to the intense color field and the initial $q_1\bar{q}_1$ pair is split into two, $q_1\bar{q'}_1$ and $\bar{q}_1 q'_1$. The process is repeated until the hadrons are not energetic enough to continue hadronization. In a multi-parton final-state energetic gluons produce kinks in

the string, while soft low energetic gluons are absorbed without any modifications. Due to the property of factorization, the hadronization process is not specific to the hard scattering that precedes it. It is therefore possible to use the same model in any scenario that involves partons in the final state.

3.4 Simulation

The MC generators are used to simulate hadron-hadron interactions and make theoretical predictions. A wide gamut of processes can be tested in detailed simulations of observables such as cross sections and angular distributions. Therefore, unsurprisingly, MC generators are most essential and widely used tools in experimental and theoretical studies.

The MC generation at CMS proceeds in two steps.

1. The physics process is simulated by probabilistically modeling the parton scattering along with the associated fragmentation and hadronization processes. The output at this stage simulates the final-state stable particles in a high energy collision event.

2. The collection of stable particles is passed through a simulation of the CMS detector. This includes detailed information on the material components and the geometric layout of the detector. The process models the interaction of particles with the detector material and provides a realistic template for comparison with the experimental data.

The two most widely used packages for MC generation at CMS are PYTHIA6 and HERWIG++. They use different combination of approaches and models for event generation[5].

3.4.1 Pythia

PYTHIA6 is an event generator which simulates the particle interactions in which many

[5]These include different PDFs and underlying event models. The role of underlying event in jet cross section calculation is discussed in section 4.4.2.

final-state particles are produced starting with two initial particles [28, 29]. It can handle large number of Standard Model and Beyond Standard Model processes. The leading order pQCD calculations are performed, by default, with the proton PDF CTEQ6L1[6] [30]. The parton fragmentation is simulated at the leading log approximation using the Lund string fragmentation model. The simulated events can be generated with any choice of pile-up and underlying event settings. The PYTHIA6 package was originally written in Fortran and recently rewritten in C++ in the PYTHIA8 release.

3.4.2 Herwig

HERWIG++ is a general purpose event generator which simulates large number of high-energy interactions in lepton-lepton, hadron-lepton and hadron-hadron collisions [31]. In addition, novel interactions can be added to the generator in the Les Houches Accord convention [32]. The QCD radiation is simulated by including color coherence effects: soft gluon interference and radiation suppression from massive particles. The underlying event is simulated with the eikonal model for multiple parton scattering [33]. The parton hadronization is performed with cluster hadronization model and hadron decays are simulated taking into account spin correlations and off-shell effects. The pQCD calculations are performed at leading order with standard PDF and underlying event settings [34].

3.4.3 NLOJet++

In this study, the next-to-leading order theory calculations are performed with NLOJET++ [35–39]. It is a general-purpose MC program which can generate QCD predictions of jet cross sections in arbitrary scattering processes. In the cross section calculations, the contributions to the cross section from the infrared emissions are included. The contributions from hadronization and underlying event processes are not calculated and have to be added using non-perturbative corrections[7]. The predictions are generated using the FASTNLO

[6] A different choice of PDF may be used if necessary.

framework which takes the PDF and the factorization and renormalization scales as input. In this study, the cross section is generated at the level of final-state particles without passing them through the CMS detector simulation (step 2 in MC generation). In order to generate the steeply falling jet spectrum with good statistical confidence, large numbers of events are simulated. However, due to the large sample size the detector simulation is not possible.

3.4.4 Geant

The detector simulation is performed with GEANT4 package [40]. The final-state hadrons from the collisions interact with the detector via many different processes, e.g., ionization, bremsstrahlung, radiation, nuclear excitation and elastic scattering. All interactions occurring in the detector and the resulting readout from the DAQ system are simulated.

[7]The derivation of non-perturbative corrections is discussed in the section 4.4.2.

Chapter 4

Inclusive jet cross section at $\sqrt{s}$ =2.76 TeV

4.1 Introduction

In proton-proton collisions at the LHC, constituent partons from the colliding protons interact to produce jets in the detector. The production of jets results from the hard interactions of the partons as well as from the softer radiative processes and spectator parton interactions. The interactions depend on the composition of the proton described by the PDFs and on the strength of the strong interaction given by α_s. Despite being the two most important contributors in QCD calculations, the value of α_s and the PDF parameterization are not predicted by the theory and have to be determined experimentally.

Jet cross section measurements depend on these two parameters and are therefore very useful in setting stringent constraints on them. A significant contribution in measuring PDFs comes from the deep inelastic scattering experiments performed with an $e^{\pm}p$ collider called HERA. The valence quark PDFs are determined with good precision unlike the gluon and virtual quark PDFs which are not. The large volume of data produced by the LHC serves to improve the precision with which the PDFs are measured and therefore improve our understanding of the structure of the proton [41–43].

In this analysis, the production of inclusive jets is studied with pp data at $\sqrt{s} = 2.76$ TeV. The cross section is measured as a function of jet transverse momentum and rapidity. This measurement is complementary to similar measurements at 7 and 8 TeV and tests QCD at another center-of-mass energy. The inclusive jet cross section measurements at

$\sqrt{s} = 2.76$ TeV, at the LHC have been performed with smaller datasets (0.2 pb^{-1}) over the momentum range up to 430 GeV [42, 44]. The data used in this analysis provides much better statistical precision (5.3 pb^{-1}) over a higher kinematic range (74-592 GeV) than any previous measurements. In addition to the cross section measurement, the double ratio $(\frac{\sigma^{2.76TeV}_{Data}}{\sigma^{2.76TeV}_{Theory}} / \frac{\sigma^{8TeV}_{Data}}{\sigma^{8TeV}_{Theory}})$ is measured. The 2.76 TeV data is collected with low pile-up conditions. Therefore, in relation to the 7 and 8 TeV data, the 2.76 TeV data is analyzed with a smaller systematic uncertainty, which is useful for precision measurements such as jet cross section ratios at different center-of-mass energies or at different jet reconstruction scales.

The workflow of this analysis is organized as follows: the data is filtered by applying the event and jet quality selection procedures. Next, the "unsmearing" procedure is applied to data[1], and the theory predictions are derived with MC generators. The systematic uncertainties on the experimental measurement and the theory are calculated. Finally, the data is compared to theoretical predictions of the cross section and double ratio measurements.

4.2 Data Samples and Event Selection

4.2.1 Datasets

The data was collected by the CMS detector in 2013 with low instantaneous luminosity. The pile-up in this sample is low, with the mean interaction rate of 0.29 every bunch crossing. The dataset includes twelve runs with the integrated luminosity of 5.43 pb^{-1}. The MC samples used in this study are produced with leading order PYTHIA6 (tune UEZ2) generator and GEANT4 detector simulation [28, 29, 40]. A complete list of samples used in data and MC is given in Appendix B.

[1]The term 'data' specifically refers to the experimental data, while simulated data are referred to as MC, or simulation.

4.2.2 Jet reconstruction

The jets are reconstructed using the anti-k$_T$ clustering algorithm with jet scale R = 0.7 [45]. The jet scale (R) determines the jet size. It is defined as the largest distance, in $\eta - \phi$ space, from the jet axis within which the constituents are included to reconstruct the jet[2]. The latest jet energy corrections are applied to correct for the non-linear detector response [46]. The corrections were originally derived at $\sqrt{s} = 8$ TeV and co-opted for the smaller $\sqrt{s} = 2.76$ TeV dataset. A description of the jet reconstruction process is not given here, it is covered in detail in section 5.2.

4.2.3 Trigger Studies

Three single-jet triggers are used to collect data. The triggers are applied by the HLT software and return a positive decision if a jet with the p_T higher than the set threshold is found. The p_T thresholds are set at 40, 60 and 80 GeV for the three triggers. The trigger efficiency for a given trigger is calculated using another trigger with a lower threshold. The ratio of number of jets passing the high threshold trigger to the low threshold trigger is the trigger efficiency (eq. 4.1). For example, trigger efficiency for trigger HLT_JetY is defined as

$$\text{HLT Efficiency} = \frac{\text{number of jets when HLT\_JetY is true}}{\text{number of jets when HLT\_JetX is true}}, \quad \text{where, Y>X.} \tag{4.1}$$

The trigger turn-on curves are built to determine the jet p_T at which the trigger becomes fully efficient[3] (fig. 4.1). Only jets collected from fully efficient triggers are used for the cross section calculation; the minimum jet p_T used is 74 GeV (table. 4.1).

[2]This is strictly true only in the case of cone algorithms. The scale parameter for sequential recombination algorithms is defined as the minimum distance between any two jets. With this definition, it is possible to have an energy deposit at a distance larger than R from the jet axis, included in the jet [45].

[3]99% efficient or higher. The jet definition used for the triggering differs from the jet definition used for the analysis. Therefore the triggers do not become fully efficient at the predefined p_T threshold.

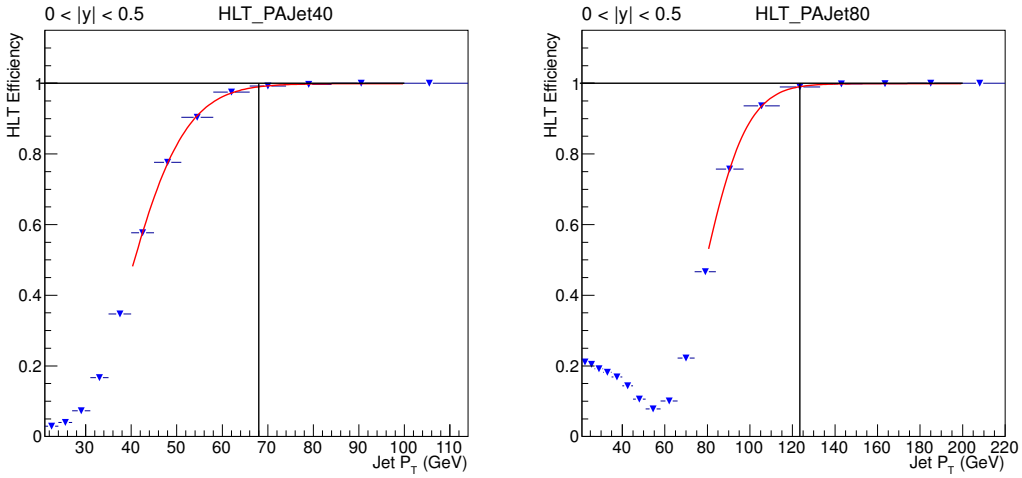


Figure 4.1: Turn on curves for triggers HLT_Jet40 and HLT_Jet80.

Trigger Name	Turn-on p_T (GeV)	p_T range (GeV)
HLT_PAJet40	68	74-97
HLT_PAJet60	94	97-133
HLT_PAJet80	123	133-2500

Table 4.1: Turn-on p_T and jet p_T range for the individual triggers.

4.2.4 Event selection

The event selection is applied to ensure that only data collected in good data-taking conditions are used for the analysis. Poor beam quality or inoperative detector components can compromise the data collected by the experiment. Hence, the CMS experiment maintains a list of certified run periods which can be safely used for data analysis. This list of runs is included in the data processing chain, and few additional selection requirements are added to ensure that data reconstruction from the detector is good.

- At least one good reconstructed vertex must be present in the event, and
- $\frac{E_T^{miss}}{\Sigma E_T} < 0.3$, where E_T^{miss} is the missing energy in the transverse plane, and ΣE_T is the scalar sum of the transverse energies of all reconstructed particles.

4.2.5 Jet selection

The jets are selected with identification criteria designed to reject objects, such as energetic leptons or photons, incorrectly identified as jets. The standard jet identification requirements for particle-flow jets at CMS are as follows [47] :

- For jets throughout the detector:
  - The jet should contain more than one particle.
  - The fraction of jet energy carried by photons or neutral hadrons must be less than 99%.
  - The fraction of jet energy carried by electrons and muons must be less than 90%.
- For jets within $|\eta| \leq 2.4$, additional cuts are applied based on the tracking information available in this region:
  - The jets should contain at least one charged particle.
  - The jet must have some (non-zero) energy contribution from the charged hadrons.
  - The fraction of jet energy carried by photons or neutral hadrons must be less than 90%.

The jet energy components in MC are compared to data after applying the jet selection (fig. 4.2). All variables used for jet identification are plotted in the appendix (figs. B.2–B.6). A typical jet event in the CMS detector is shown in fig. 4.3.

4.3 Experimental Measurement

The double-differential cross section is calculated by combining data collected from different triggers. The jet yield N is scaled by the luminosity and the corresponding p_T and

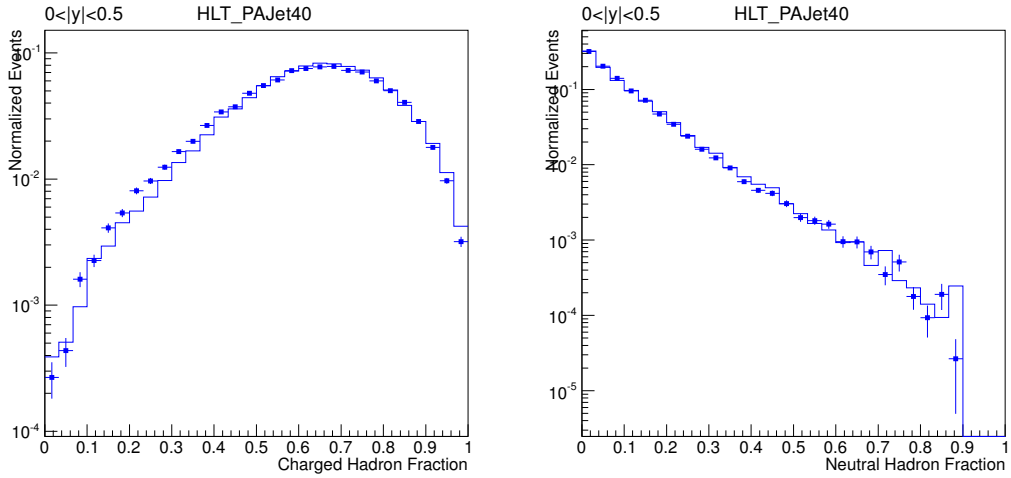


Figure 4.2: The contribution to jet energy from charged hadrons (left) and neutral hadrons (right) is plotted for jets in the central rapidity region ($0 < |y| < 0.5$). The data is shown with points while MC is shown as a solid line histogram.

y bin-widths to get the cross section (eq. 4.2). Only jets in the p_T range in which the trigger is fully efficient are considered (table 4.1). The prescale factors are applied to reduce the large flux of data, collected by a trigger, to fit the limited data processing bandwidth available to DAQ. The prescaling procedure involves randomly sampling a limited fraction of this data. The fraction is determined by the prescale factor. The triggers with lower p_T threshold, 40 and 60 GeV, are subject to prescales. The double-differential cross section is calculated as,

$$\frac{d^2\sigma}{dp_T dy} = \frac{1}{2}\frac{1}{\varepsilon \mathscr{L}_{\text{int,eff}}}\frac{N}{\Delta p_T \Delta|y|} \tag{4.2}$$

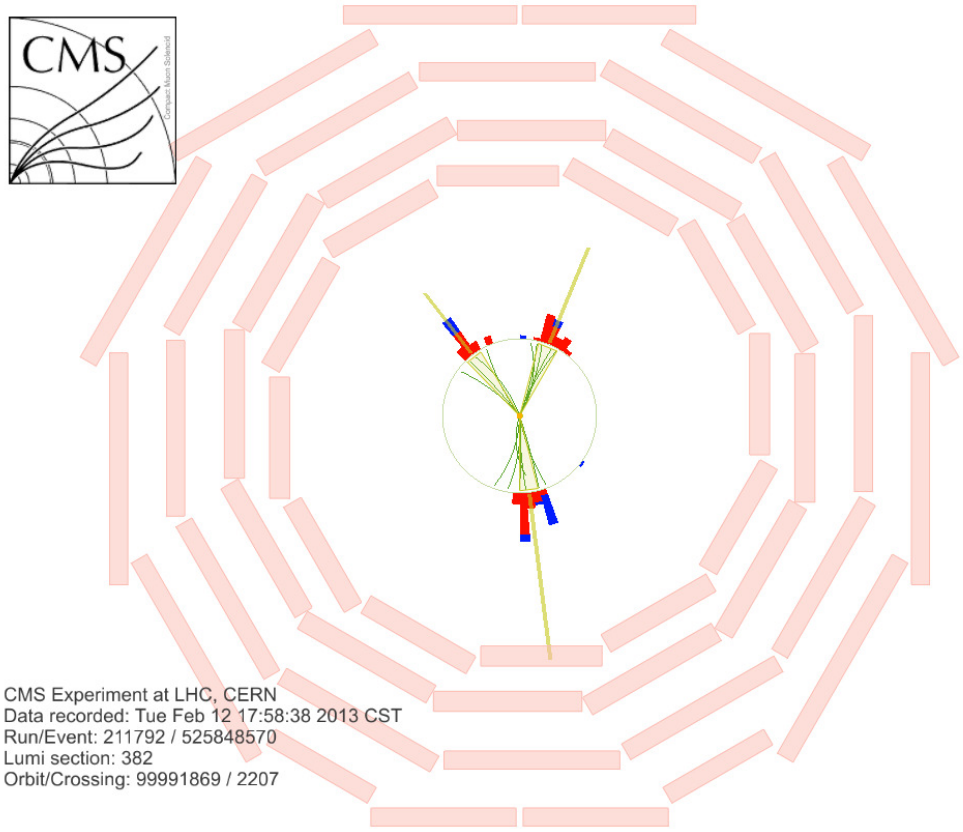
where,

N is number of jets in the p_T and $|y|$ bin,

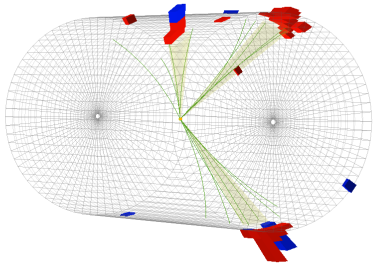
$\mathscr{L}_{\text{int,eff}}$ is effective luminosity corrected for prescales,

Δp_T is jet p_T bin-width,

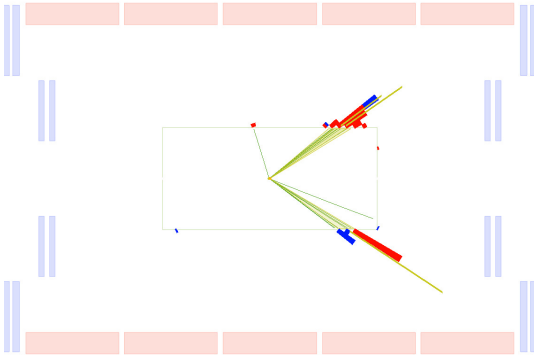
$\Delta|y|$ is rapidity bin-width,



(a) Projection in r–ϕ plane.



(b) 3D view.



(c) Projection in r–z plane.

Figure 4.3: A typical jet event in the CMS detector from different points-of-view [6]. HCAL deposits are shown in blue, ECAL deposits are shown in red and tracks are shown in green. The shaded area and solid yellow line indicates reconstructed jets.

ε is reconstruction efficiency which combines trigger and jet id efficiencies.

The observed cross section is plotted in fig. 4.4. The jets used for building this spec-

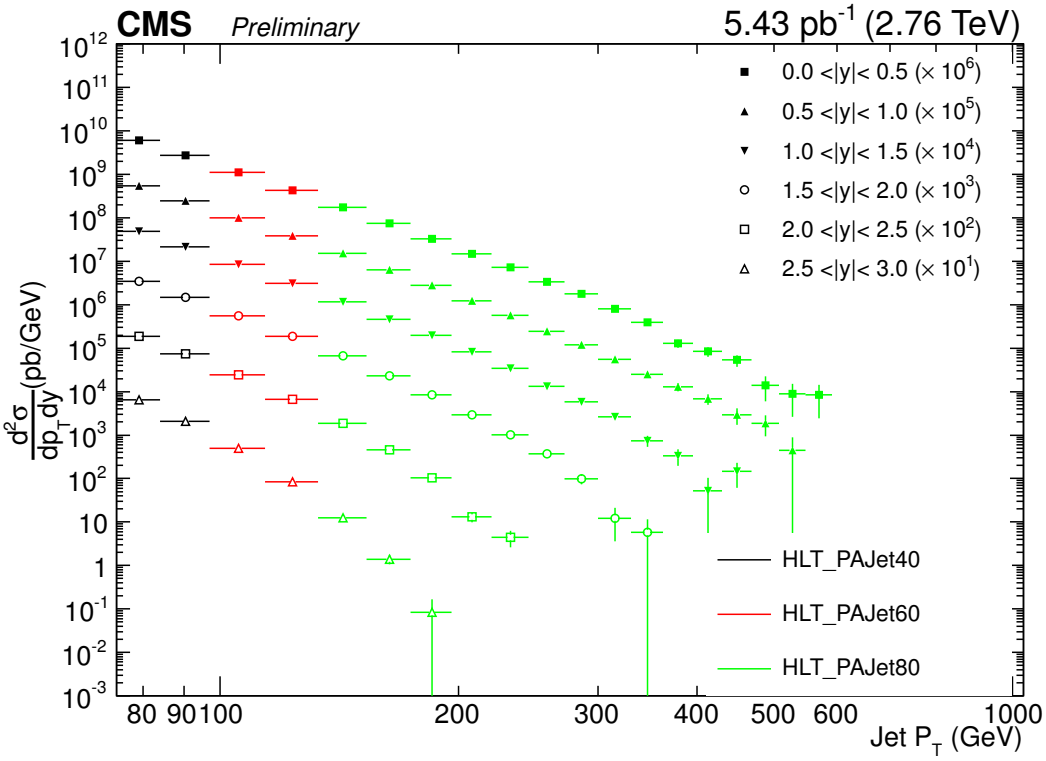


Figure 4.4: The double-differential inclusive jet cross section built from reconstructed jets.

trum are "detector-level" jets affected by the detector resolution. To get the "unsmeared" particle-level spectrum, the unfolding procedure is performed.

4.3.1 Unfolding

The jets produced in the hard interaction are composed of final-state particles which interact with the detector and deposit energy. The energy resolution of the detector alters the particle-level p_T spectrum of the jets. The original shape of the spectrum is steeply falling, hence the bin migration of jets between the adjacent p_T bins is unequal. For instance, a low p_T bin has larger number of jets migrating into the adjacent high p_T bin than the jets migrated from the high p_T bin into the low p_T bin. The jet yield is enhanced by 20–30%

in the detector-level spectrum, as seen in fig. 4.5. The unfolding procedure is employed to correct for this effect and recover the original particle-level spectrum produced in the pp collisions. This procedure estimates the migration of jets associated with each bin, using a simulated sample of jets and the measured energy resolution of the detector.

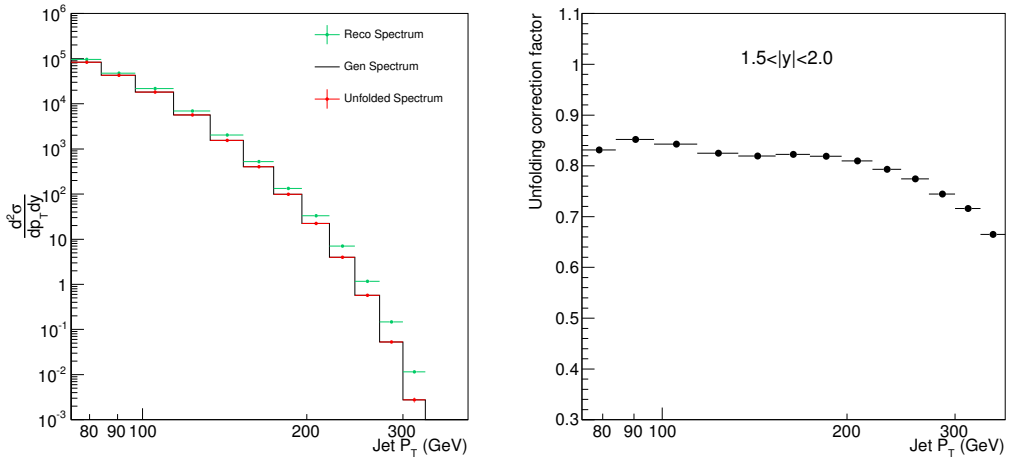


Figure 4.5: Comparison of particle-level (*gen*), detector-level (*reco*) and unfolded spectrum (left) and the correction factors applied by the unfolding procedure (right). The spectrum is generated from NLO theory and the response matrix is constructed with toy MC.

The unfolding is performed using the ROOUNFOLD package in two steps [48]:

1. The response matrix is constructed using NLO inclusive jet theory spectrum.

2. The detector-level spectrum from data is unfolded using the response matrix.

The response matrix is generated to map the bin migration between the particle-level and the detector-level spectrum, using toy MC simulation. With this method, large number of particle-level events are generated following a theoretical jet spectrum. The particle jet p_T is then smeared with the jet energy resolution (JER) to get the detector jet p_T. The particle and detector jet p_T values are used to build the response matrix for the unfolding procedure (fig. 4.6).

The expected number of events in the unfolded spectrum ($n(C)$) is determined as follows. The probability of observing the jet yield E_j in bin j, given the particle-jet yield

C_i in bin i is denoted as $P(E_j|C_i)$, and determined from the response matrix. Then, the probability that the particle-jet yield causes a certain observed-jet yield ($P(C_i|E_j)$) can be calculated as below, assuming the initial probabilities $P_0(C_i)$ [49, 50]:

$$P(C_i|E_j) = \frac{P(E_j|C_i) \cdot P_0(C_i)}{\sum_l P(E_j|C_l) \cdot P_0(C_l)}, \tag{4.3}$$

where, l goes over all particle bins. Therefore, the particle-level yield is calculated from the observed yield using the relation,

$$n(C_i) = \sum_j n(E_j) \cdot P(C_i|E_j), \tag{4.4}$$

where, j goes over all detector bins. This process is iterated using the output from a given iteration ($n(C_i)$) considered as the input in the successive one. A maximum of four iterations are used.

Jet Energy Resolution (JER)

As an input to the unfolding procedure, the JER of the detector is calculated in MC and data. First, the JER is determined as a function of jet p_T and rapidity using the MC samples. The detector jets are matched to particle jets within the distance (dR) of 0.3 and $p_T^{detector}/p_T^{particle}$ distribution is built[4]. The resolution is determined by calculating 1σ width of this distribution [5]. The p_T dependent jet resolution is determined with the function,

$$\sigma_{MC}(p_T) = \sqrt{\frac{N \cdot (|N|)}{p_T^2} + \frac{S^2}{p_T} + C^2}, \tag{4.5}$$

where, N, S and C refer to the noise, stochastic and constant terms. The noise term models

[4] $dR = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, i.e., distance in η-ϕ space.

[5] The 15.83 and 84.17 quantile values are used to determine the 1σ width.

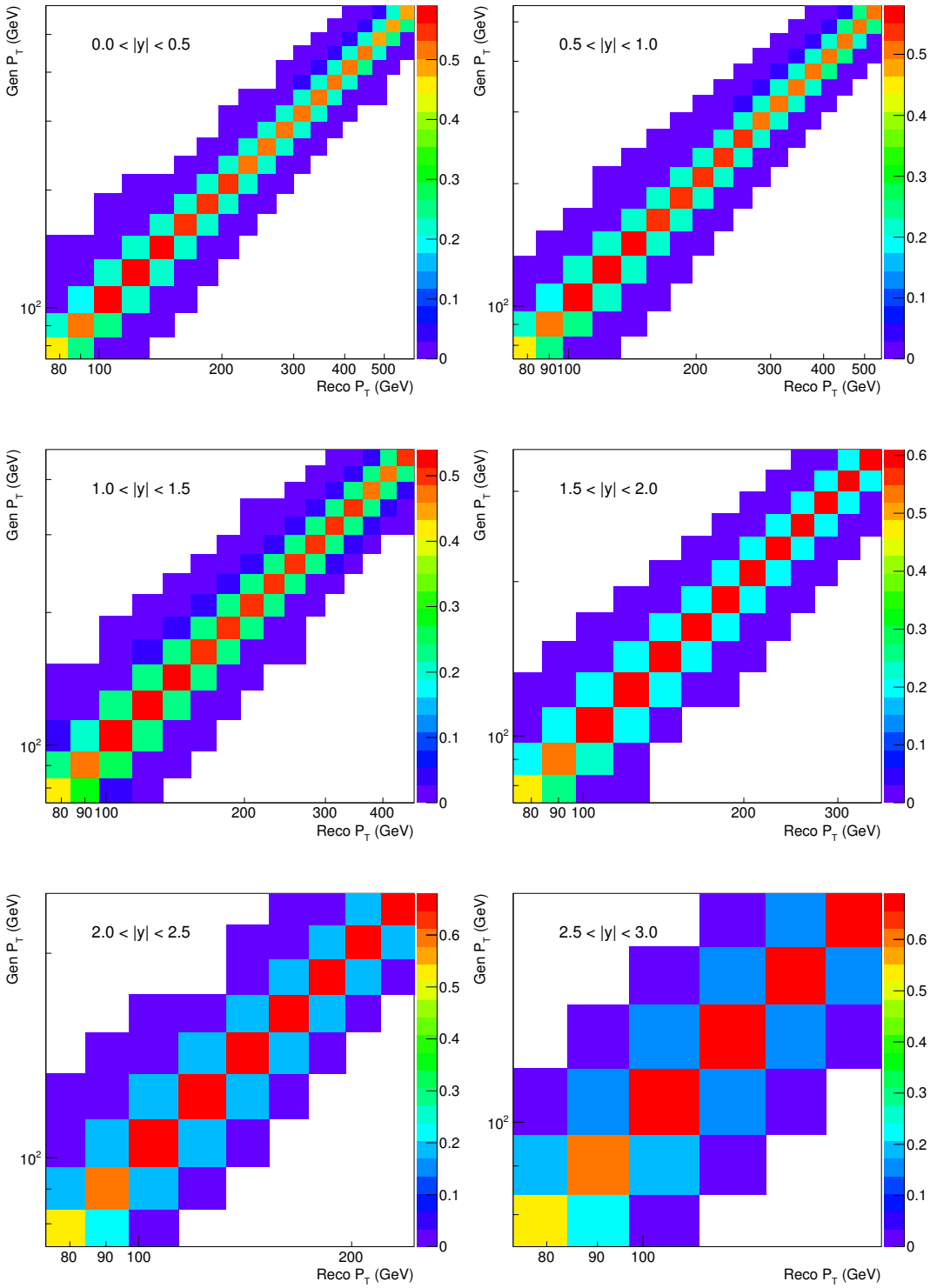


Figure 4.6: The response matrix constructed with toy MC for all rapidities.

fluctuations that dominate primarily at low p_T, the stochastic term models the broad intermediate p_T region and the constant term is significant at setting the asymptotic limit at high p_T. To estimate JER in data, the resolution values from MC are smeared with $c^{data/MC}$ factor to account for the discrepancy in resolutions that has been observed between data and MC (eq. 4.6). The jet resolution is modelled for each rapidity bin. The parameter values used of N, C, S and $c^{data/MC}$ are given in the table 4.2. The $c^{data/MC}$ factor is derived in jet resolution studies at CMS documented in [51].

$$\sigma_{Data} = \sigma_{MC} \times c^{data/MC} \tag{4.6}$$

y	N	S	C	$c^{data/MC}$
0.0-0.5	-3.550(37)	0.892(3)	0.0373(2)	1.079(26)
0.5-1.0	-3.342(50)	0.902(3)	0.0380(2)	1.099(28)
1.0-1.5	-2.803(110)	0.986(6)	0.0396(5)	1.121(29)
1.5-2.0	+2.233(196)	0.776(9)	0.0275(12)	1.208(46)
2.0-2.5	+1.700(276)	0.621(22)	0.0254(26)	1.254(62)
2.5-3.0	-0.679(353)	0.686(13)	0.0000(294)	1.395(63)

Table 4.2: The values of parameters with uncertainties used to model the JER function.

4.4 Theory Prediction

The inclusive jet cross section is derived from theory at next-to-leading order accuracy for comparison with data. The data is compared to five different theory cross sections, each derived from a specific approximation of the proton PDF. The derivation of the theory cross section includes the hard scattering cross section and the non-perturbative effects. In the following section, the two components of the derivation are covered in detail.

4.4.1 Inclusive cross section at NLO

The theory parameters include the choice of renormalization and factorization scales (μ_R, μ_F) and the PDFs.

The calculations of inclusive jet cross section are susceptible to infrared and ultraviolet divergences (see chapter 3 for details). The renormalization scale parameter is added to regularize the logarithmically divergent calculations in the ultraviolet regime. The infrared contributions to the cross section are removed by introducing a scale cutoff (on collinear radiation) called the factorization scale. The values of μ_F and μ_R are set in relation to the scale of the interaction called Q. For theory calculations, the following default values are used: $\mu_F = \mu_R = Q$.

The PDFs are crucial in predicting the parton-parton interactions in the high energy collisions. They are studied by probing physics processes which are sensitive to the momentum distribution of partons inside hadrons. A number of processes, observed in hadron-hadron and lepton-hadron collision data, are used for PDF estimation and yield tighter constraints in combination[6]. The five sets of PDFs used for deriving theory cross sections are evolved with next-to-leading order accuracy. Each set of PDFs uses a different combination of datasets, physics processes, basic models and assumptions about the active quark composition in the proton.

The physics processes used for the derivation of the PDFs are given below, with the number of active quarks flavors used in the fitting procedure.

1. CT (Coordinated Theoretical-Experimental Project on QCD) – The PDFs are derived from the deep inelastic scattering (DIS), vector boson production and inclusive jet production from HERA and Tevatron experiments. A general-mass framework is adopted which assumes up to five active quarks flavors [30].

[6]Note: The experimentally derived PDF and the cross section theory derived using such PDF are named with a PDF specific labels. For example, HERA PDF refers to the PDF derived from the H1 and ZEUS experiments on the HERA collider.

2. NNPDF (Neural Network Parton Distribution Function) – NNPDF methodology uses neural nets to determine PDFs and includes up to six active quarks in the proton. The datasets are adopted from NNPDF2.0 which include HERA (DIS), Drell-Yan, and electroweak and inclusive jet production from Tevatron. An updated version of the charm structure function is used for fitting this PDF [52].

3. HERA (H1 and ZEUS Collaborations) – The DIS data from $e^{\pm}p$ scattering is used to determine PDFs. The data from the two running periods, 1992-2000 and 2000-2007, is combined with variable-flavor numbering scheme which assumes up to five active quarks [53].

4. MMHT (Martin, Motylinski, Harland and Thorne) – Three main processes are considered for PDF fits: fixed-target experiments, HERA (DIS), inclusive jet production, and W asymmetry and Z rapidity distributions measured at Tevatron and LHC. The PDFs are fitted with the aim of making accurate W,Z cross section predictions at the LHC [54].

5. ABM (Alekhin, Blümlein, Moch) – The PDFs are derived from DIS in several experiments at CERN and Fermilab, and Drell-Yan process in two Fermilab fixed-target experiments. The PDF fits are made with the assumption of fixed number of quark flavors ($n_f = 3,4,5$). The PDF used for comparison in the analysis is derived with $n_f = 5$ [55].

4.4.2 Non Perturbative Corrections

The inclusive jet cross section is derived with QCD calculations in the perturbative regime which includes initial and final-state emissions. The fixed order pQCD provides a good description of hadron interactions at the parton level. However, the jet spectrum is modified by hadronization of partons as well as by the multiple interactions between spectator partons in the colliding hadron. The contributions of hadronization and multiple parton

interactions to the jet spectrum are calculated using non-perturbative (NP) theory.

The NP corrections to the NLO cross section are derived with PYTHIA6 and HERWIG++ generators [28, 29, 31, 34][7]. The average of the correction factors derived from these two sources is considered as the NP correction factor. The cross sections (σ) are calculated with and without multiple parton and hadronization effects to derive the correction factor C_{NP}. The generator settings are modified in order to simulate QCD interactions for the two cases. The NP correction is calculated as:

$$C_{NP} = \frac{\sigma(MPI + HAD + PS + LO)}{\sigma(PS + LO)}. \tag{4.7}$$

The NP correction factor is defined as the ratio of jet cross section calculated with and without the hadronization (HAD) and multiple parton interactions (MPI) (eq. 4.7). The cross section is determined with the leading-order (LO) pQCD calculation and the contribution from the parton shower (PS). The theoretical cross section derived at NLO is corrected for the non-perturbative contributions:

$$\sigma_{Theory} = \sigma_{NLO} \times C_{NP} \tag{4.8}$$

The NP correction factors decrease with increasing interaction scale (Q) since the contributing processes are softer than the primary interaction and become less significant at higher scales. The correction factors range from 2-10% across the rapidity range.

[7]The multiple parton interaction is also called 'underlying event' and modeled differently by each generator. HERWIG++ v.2.5.0, underlying event model UE_EE_3C. PYTHIA6, underlying event model UEZ2.

4.5 Systematic Uncertainties

The systematic uncertainties of the experimental measurement and the theoretical calculations are evaluated. A description of sources contributing to the uncertainties is given in the following sections.

4.5.1 Experimental uncertainties

The experimental uncertainties refer to all uncertainties that contribute to the experimental measurement. These include jet energy correction scale (JEC) uncertainty, JER uncertainty and luminosity uncertainty. The efficiencies of the jet and trigger selection criteria are high (99% or higher). Hence, a conservative estimate of 1% uncorrelated uncertainty is added across all jet p_T and rapidity bins, to account for the small contributions from the trigger and jet selection.

Jet energy scale uncertainty

The jet spectrum falls steeply as a function of jet p_T. As a consequence, any variation in jet energy correction scale can cause a large variation in the spectrum[8]. The uncertainty on the jet corrections depends on the physics processes used for its derivation. The main sources include noise and pile-up, extrapolation beyond the reach of available dataset, time-dependent correction factors which vary over the data-taking periods, and flavor composition of jets. The uncertainty on the cross section due to these individual sources is calculated. The uncertainties are added in quadrature as they are uncorrelated, to obtain the total JEC uncertainty.

The JEC uncertainty is calculated by varying jet energy corrections by one standard deviation and estimating the change in the cross section. Due to the steep shape of the cross section at outer rapidities and high JEC uncertainty values ($> 12\%$), the calculated uncertainty is very large in this region. The total uncertainty on the inclusive cross section

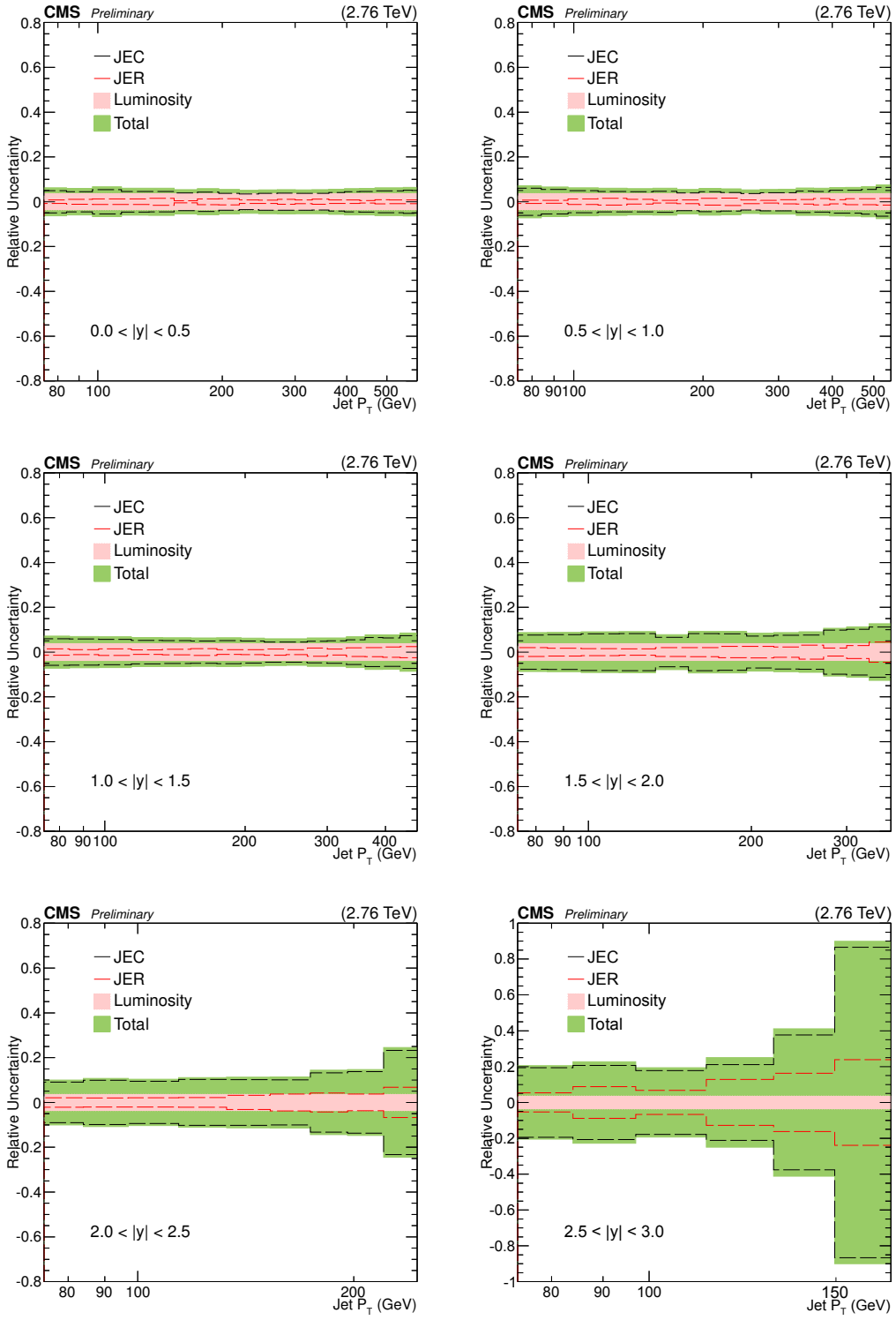


Figure 4.7: Total relative uncertainty on the experimental inclusive jet cross section.

varies between 5% and 85% due to our imprecise knowledge of the jet energy scale.

JER uncertainty

The JER uncertainty is calculated by varying the fitted parameters (in table 4.2) by one standard deviation. The uncertainty values on the parameters are obtained from the fitting procedure. The modified JER is used to calculate the cross section, and the largest deviation from the central value is considered as uncertainty. The JER uncertainty varies from 2-27%.

Luminosity Uncertainty

The uncertainty on the luminosity measurement is estimated to be 3.7% [56].

The total uncertainty is calculated by adding the uncertainties from the three uncorrelated sources in quadrature. The total experimental uncertainty and the contributing sources are shown in fig. 4.7. The JEC dominates the total experimental uncertainty while JER and luminosity make comparable contributions. The total uncertainty is representative of one standard deviation fluctuation in the experimentally measured data. The statistical uncertainties on (unfolded) data are correlated, as are the systematic uncertainties due to JEC and JER. Hence, any changes in measurement due to these effects apply to all bins simultaneously.

4.5.2 Theoretical uncertainties

The uncertainty on the theoretical cross section is introduced due to the uncertainty in PDF determination, the choice factorization and renormalization scales, and imprecise modelling of non-perturbative effects.

1. The PDFs are determined using experimental data from collider and fixed target experiments. A variety of model parameters are tested to find the best fit to data. The

[8] The cross section varies, approximately, as $\sigma \propto \frac{1}{p_T^5}$, or $\frac{d\sigma}{\sigma} \propto 5\frac{dp_T}{p_T}$. Therefore, an uncertainty of 10% in the jet p_T gives 50% uncertainty on the cross section.

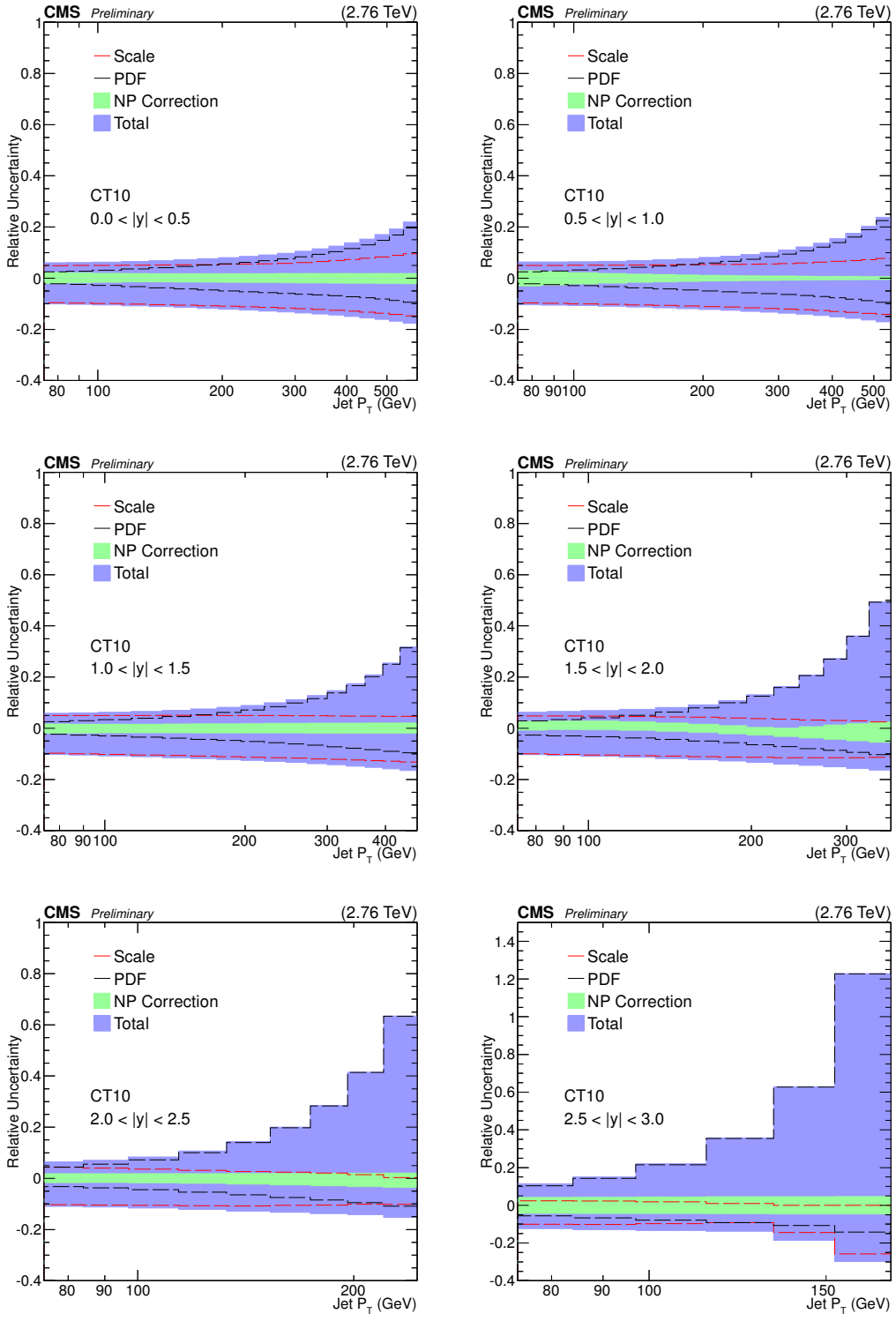


Figure 4.8: Total relative uncertainty on the theoretical inclusive jet cross section calculated with the CT10 PDF.

uncertainty on the PDF include uncertainties associated with data (systematic and statistical), and model and fit parameterization. The PDFs are provided with the set of central values and the associated uncertainties corresponding to one standard deviation. The PDF uncertainty is calculated by varying the PDF using the uncertainty parameters.

2. Scale uncertainty is determined by varying the factorization and renormalization scales in the following combinations: $(\mu_F/\mu, \mu_R/\mu)$ where $\mu = Q$ (0.5,0.5), (1,0.5), (0.5,1), (1,2), (2,1), (2,2). The largest deviation from the central value of the inclusive cross section is considered as the scale uncertainty.

3. NP corrections are calculated using PYTHIA6 and HERWIG++ generators independently. The largest deviation from the average value is considered as uncertainty.

The three sources are uncorrelated and their corresponding uncertainties are added in quadrature to calculate the total uncertainty. Overall, the total uncertainty varies from 10 to 120% with a dominant contribution from the PDF (fig. 4.8).

4.6 Inclusive cross section ratio at $\sqrt{s} = 2.76$ TeV/$\sqrt{s} = 8$ TeV

The inclusive cross section measurement obtained from this analysis is compared to a similar measurement at $\sqrt{s} = 8$ TeV. This measurement was performed with data collected by the CMS detector in 2012 [57]. The binning of jet p_T and rapidity is consistent between the two analyses, which makes a direct comparison between the two cross sections possible. The experimental systematics, which dominate both measurements, are significantly reduced in the ratio when measured in the same bins.

The double ratio of inclusive jet cross sections calculated as

$$\rho = \frac{\sigma_{Data}^{2.76TeV} / \sigma_{Theory}^{2.76TeV}}{\sigma_{Data}^{8TeV} / \sigma_{Theory}^{8TeV}} \tag{4.9}$$

4.6.1 Systematic uncertainty on the cross section ratio

The systematic uncertainty between the two measurements is correlated. The individual sources contributing to the theoretical and the experimental uncertainty are examined. The calculation of the systematic uncertainty is summarized below:

1. The jet energy correction scale used for the two measurements is the same. Hence, the JEC uncertainties are fully correlated. The 8 TeV measurement has an added uncertainty due to the pile-up corrections which is considered uncorrelated.

2. The unfolding, scale, PDF and NP correction uncertainties are considered correlated, since they are derived with the same procedure in the two studies.

3. The luminosity uncertainty is conservatively considered to be uncorrelated.

4. The uncertainty on the double ratio due to correlated source "s" is calculated as

$$\delta_s^{ratio} = \frac{1 \pm \delta_s^{2.76TeV}}{1 \pm r\delta_s^{8TeV}} - 1 \tag{4.10}$$

   where,

   δ_s^{ratio} relative uncertainty on ratio ρ due to source "s"

   $\delta_s^{2.76TeV}$ relative uncertainty due to "s" on 2.76 TeV cross section

   δ_s^{8TeV} relative uncertainty due to "s" on 8.0 TeV cross section

   r is the correlation coefficient (+1 or -1) for the two uncertainties, $\delta_s^{2.76TeV}$ and δ_s^{8TeV}.

5. The uncertainty on the double ratio due to uncorrelated errors (δ_s^{ratio}) is :

$$\delta_s^{ratio} = \sqrt{(\delta_s^{2.76TeV})^2 + (\delta_s^{8TeV})^2} \tag{4.11}$$

The total experimental uncertainty on the double ratio is shown in the fig. 4.9. The correlated errors are smaller and make non-leading contribution to the total experimental

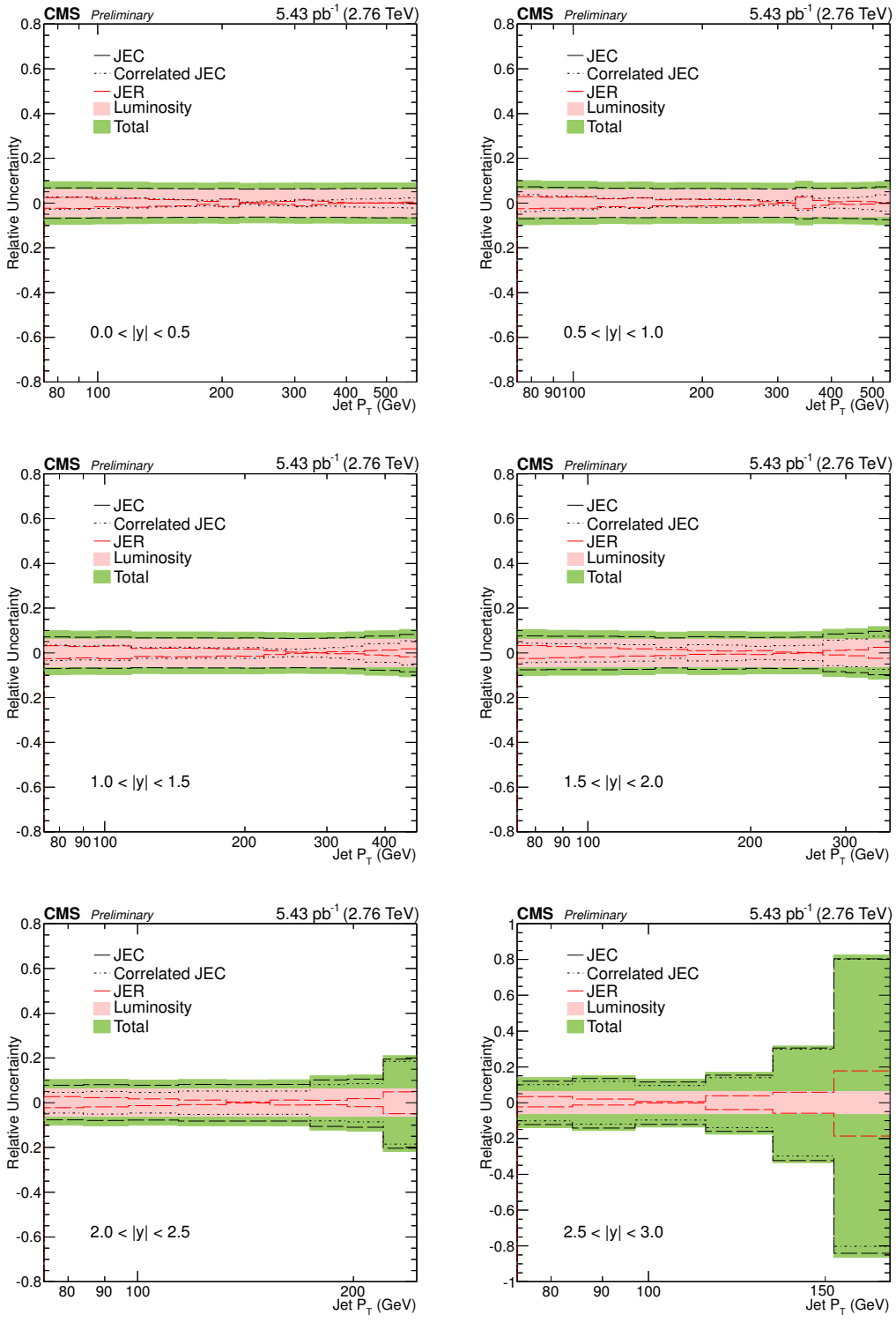


Figure 4.9: Total relative uncertainty on the experimental inclusive jet cross section double ratio.

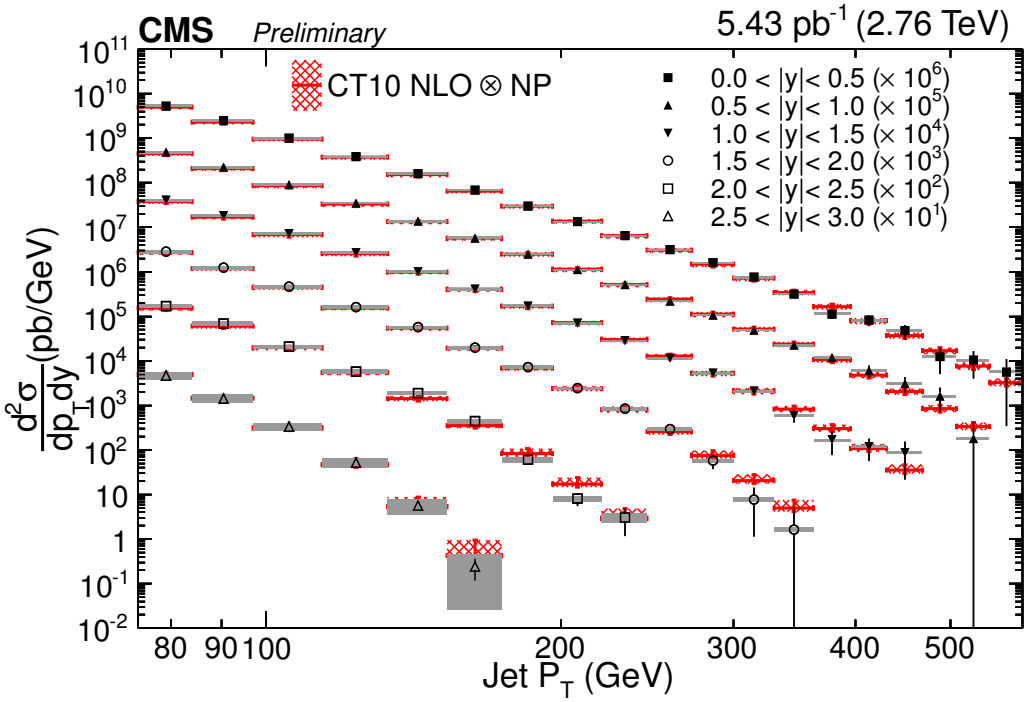


Figure 4.10: The double-differential jet cross section plotted in comparison to the CT10 theory.

uncertainty in most cases. The theoretical uncertainties on the double ratio are not shown since they are comparable to the theoretical uncertainties on the cross section and do not change significantly.

4.7 Summary and results

The double-differential cross section is shown in fig. 4.10. A closer look at data is shown in figs. 4.11 and 4.12. The ratio of data to the theoretical prediction derived with the CT10 PDF fluctuates within 10% of unity, and is covered by the systematic uncertainty. In fig. 4.12, the ratio of data to theoretical prediction derived with CT10 is compared in relation to the predictions from other PDFs. The colored dashed lines indicate the ratio of the other PDFs to CT10. The data shows good agreement with all PDFs with the exception of ABM11 which deviates by as much as 30% from unity. In fig. 4.13 the double ratio of cross sections at 2.76 TeV and 8 TeV is presented. The CT10 PDF shows the double

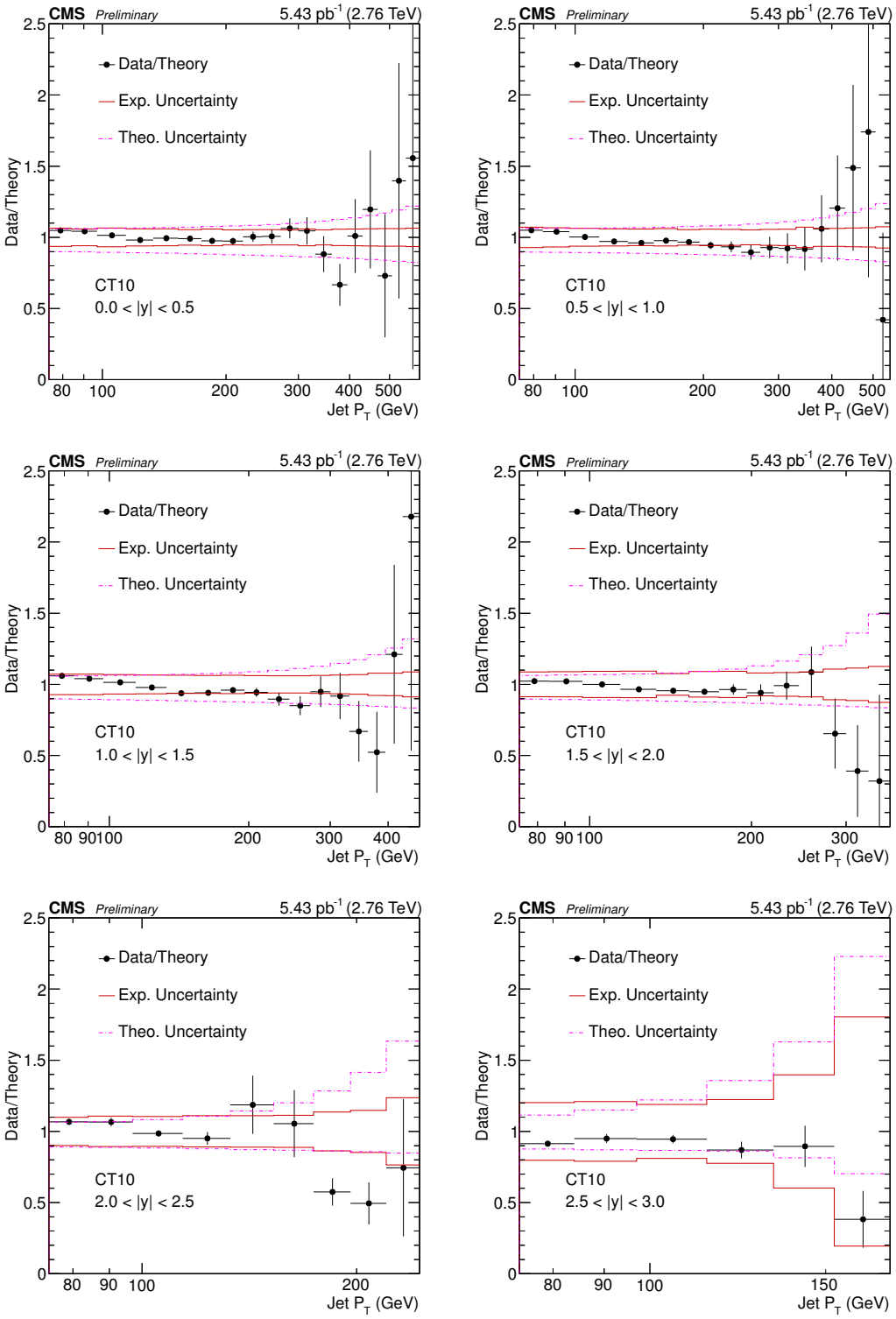


Figure 4.11: Data to theory comparison with CT10 theory (comparison with other theories in Appendix B).

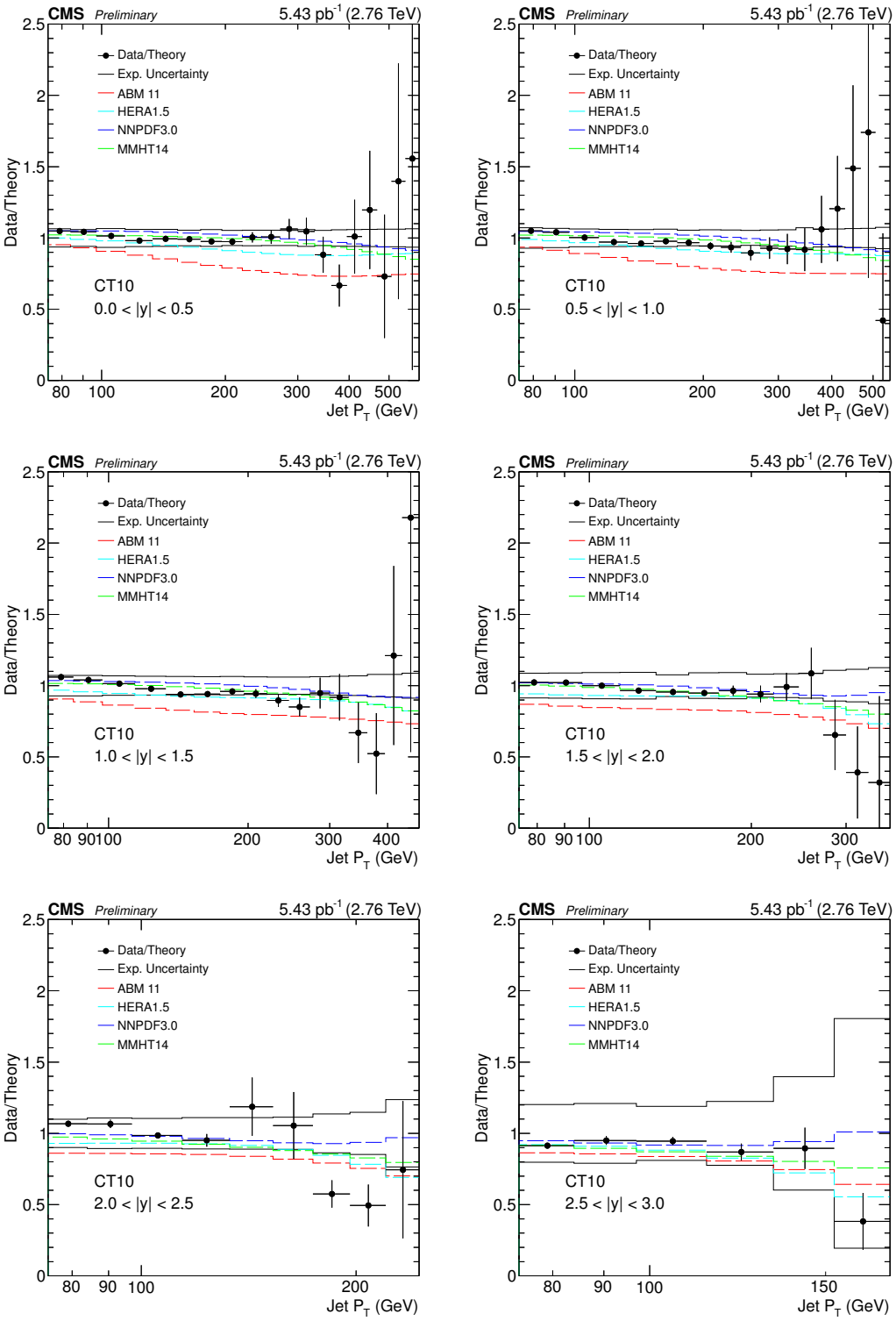


Figure 4.12: Data and other theories compared to CT10 theory.

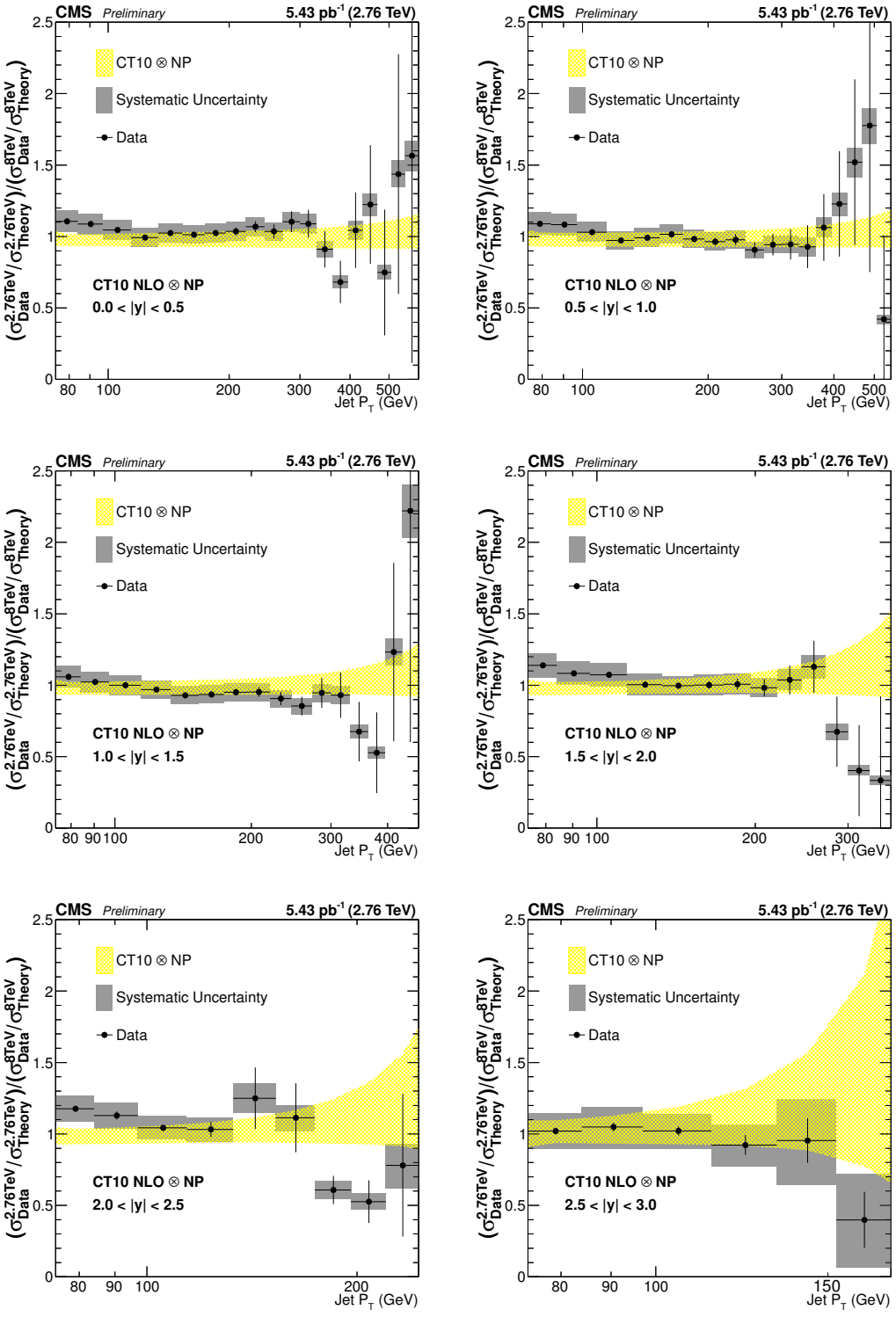


Figure 4.13: Inclusive jet cross section ratio at 2.76 TeV and 8 TeV compared to CT10 theory (comparison with other theories in Appendix B).

ratio consistent with unity within uncertainties. Double ratios calculated with other PDFs exhibit a small positive deviation up to 20%, and 30% in the case of the ABM PDF. Overall, theory describes the data very well over the observed kinematic region. The cross section predictions are tested with the systematic uncertainty of 5-10% over most rapidities, except the outermost rapidity where the experimental uncertainty deteriorates. The double ratio tests the agreement of data with theory at two different center-of-mass energies with the systematic uncertainty of 10%. Since the two datasets were collected during the same data-taking period and the analyses are performed with similar techniques, the systematic uncertainties on the measurements are highly correlated. Hence, the theory is tested with an improved precision.

The results presented in this analysis represent the first inclusive cross section measurement from CMS at $\sqrt{s} = 2.76$ TeV and extend the probed kinematic region compared to the previous measurements. The inclusive jet cross section is measured over the p_T range 74 to 592 GeV and absolute rapidity range 0 to 3.0. The results from the analysis can be used for more detailed studies, to improve our understanding of the fundamental parameters of QCD. The 2.76 TeV measurement can be combined with other measurements to improve global fits on PDF and α_s, or these fits can be derived independently for comparison. The double ratio at 2.76 and 8 TeV provides a test of the theory at two center-of-mass energies, and can serve as a cross check for PDFs derived with other datasets.

The results from this analysis allow us to test our understanding of the proton structure and the jet production process due to inelastic scattering. The PDFs used for comparison are derived from different experiments, primarily from hadron-hadron and hadron-lepton colliders, operating at a lower center-of-mass energy. These PDFs are evolved to make predictions at a higher interaction scale, corresponding to $\sqrt{s} = 2.76$ TeV and 8 TeV. The results show that these predictions are accurate within the uncertainty and allow for an improved estimate of the PDFs by adding an independent constraint.

Chapter 5

Calibration of FFTJet Algorithm

Parton scattering is a commonly occurring interaction in pp collisions at the LHC. The deep inelastic scattering process is of particular interest as it provides crucial information about the proton structure, and allows us to test theoretical predictions describing such interactions. The production of jets results from the hadronization of final-state partons created in the scattering process. Jets represent the energies and directions of the outgoing partons and therefore are widely used in physics analyses. As proxies to the originating partons, jets can provide insight into the particle interactions occurring in the high energy collisions [58, 59]. For example, in hadron-hadron collisions the parton-parton scattering can be analyzed in terms of jet cross sections, as measured in chapter 4, to infer the value of parton momentum distributions inside hadrons and the strong coupling constant. Another application of jets is to search for heavy as-to-yet undiscovered particles produced in the collisions which subsequently decay into partons. The studies in jet multiplicities and jet substructure provide useful information about the radiative properties of partons, the strong coupling and the color structure of interactions.

Due to the large number of uses for jets and wide breadth of phenomena studied with jets, the practical definitions of "jet" vary. In order to compare data from different analyses and experiments, some common standards for jet definition are adopted. These include universality between theory and experiment, finite and calculable cross sections to any orders in pQCD, and infrared and collinear safety. The theory of jet production and fragmentation is covered in section 3.3.

5.1 Purpose of this study

Jet reconstruction algorithms are important tools in data analysis in the high energy regime. With the rise in the center-of-mass energy, especially in hadron colliders, the complexity of the final-state increases. This is evident at the LHC where jet production from inelastic interactions is accompanied by an unprecedented amount of underlying event and pile-up. Therefore, the ability to discriminate between these secondary effects while identifying all the information about the primary interaction is crucial.

There are many jet algorithms in use at CMS. These include cone-based algorithm SISCone [60, 61], and sequential recombination algorithms, anti-k_T [45], Cambridge-Aachen (CA) [62] and k_T [63]. The SISCone, k_T and CA algorithms are sensitive to pile-up and underlying event, and therefore have limited applications at the LHC. In addition, the k_T and CA algorithms suffer from irregular jet shapes and variable jet areas, making it difficult to compare different results and derive calibrations. The most commonly used jet algorithm at CMS is anti-k_T. It shows the best performance among the sequential recombination algorithms, with the least sensitivity to pile-up and underlying event, and produces jets with near uniform jet areas. Despite the success, this algorithm has the following drawbacks: a) the process of sequential recombination is affected by any extraneous contributions to jet energy, b) energy deposits are always assigned to the most energetic jet instead of sharing between multiple jets c) jet identification and reconstruction are performed at fixed jet scales. The FFTJet algorithm is designed to address these problems.

The motivation behind the FFTJet algorithm is to perform precise jet identification which allows for substructure analysis with reduced sensitivity to pile-up and noise. This is implemented using a novel approach to jet reconstruction, the use of Gaussian filters with multiple scale capability. The reconstruction is divided into two steps in which jet identification and jet energy measurement are performed independently. The use of cones for energy measurement ensures that jet areas are fixed, and allows the energy deposits to

be divided between adjacent jets, proportional to the overlapping area. The purpose of this study is to implement the FFTJet algorithm at CMS. The study is divided into three parts. In the first part, the operating parameters of the algorithm for processing CMS data are determined. This is performed using MC simulation. In the second part, the jet calibrations are derived. This is an important step, since the jet energy is affected by several processes which makes it non-trivial to recover its original value. The calibration involves adjusting jet energies to correct for the effects of pile-up and energy response of the detector. The MC and data samples used for calibration include zero-bias , QCD MC, dijet and Z+jet. In the third part, an analysis is performed using a well studied process, semi-leptonic $t\bar{t}$ decay. In this process, a hadronically decaying particle (W boson) is produced in conjunction with other particles which provide very specific experimental signatures, i.e., lepton produced with two heavy-flavor jets. A relatively pure sample of semi-leptonic $t\bar{t}$ events is obtained, in data and MC, by selecting for these process-specific signatures. The reconstructed invariant mass and transverse momentum of the W boson are examined. A good agreement between the MC and data is required to confirm correct jet energy calibration.

The following section (section 5.2) covers the data processing at the detector level, with a brief description of the particle-flow algorithm. Next, the FFTJet algorithm and the calibration process is explained in detail in section 5.3. Finally, the validation of jet corrections is presented in section 5.4.

5.2 Jet reconstruction

Jet reconstruction is performed starting with the energy flow in the detector. At CMS, the energy flow is constructed using the following input objects from the detector: calorimeter towers, reconstructed tracks and the muon system. Three types of jets are constructed:

- *Calo* jets are reconstructed from calorimeter inputs using energy deposited in the electromagnetic and hadronic calorimeters.

- *JPT* (jet-plus-track) jets are reconstructed using information from the calorimeters combined with the tracking system, to take advantage of better energy resolution of charged particles reconstructed from the tracks.

- *PF* (particle-flow) jets are composed of particles which are reconstructed using information from all sub-detectors used in the experiment [64]. The particle-flow jets have better energy resolution and more stable energy response compared to the other two jet types. Hence they are the most commonly used jets for data analysis at CMS. In this study, the particle-flow reconstruction of the energy flow is used to calibrate FFTJet. The particle-flow algorithm is briefly described below.

5.2.1 Particle-Flow event reconstruction

The particle-flow algorithm aims to reconstruct all stable particles in the event using the optimal combination of all CMS sub-detectors, and to get the best possible particle identification and energy-momentum (**P**) resolution. This approach exploits the fact that different sub-detectors of CMS are optimized for measuring different types of particles. The particles in the event are divided into following categories:

- **Photons** are reconstructed using the electromagnetic calorimeter, owing to its good electromagnetic resolution.

- **Charged and neutral hadrons** are reconstructed using calorimeter deposits in the hadronic and electromagnetic calorimeters. The energy of charged hadrons is recalculated from the tracks, and the calorimeter deposits related to charged hadrons are excluded to estimate the energy of neutral hadrons.

- **Electrons** are reconstructed from the tracks and the energy deposits in the electromagnetic calorimeter.

- **Muons** are reconstructed as isolated or associated with the jets, using the combination of information from the trackers and the muon system.

- **Missing energy (E_T^{miss})** is defined as the negative vector sum of the transverse momenta of all reconstructed particles.

5.3 FFTJet algorithm

FFTJet algorithm takes a two step approach to jet reconstruction: i) pattern recognition and ii) energy reconstruction. The pattern recognition step is performed by convolving energy flow from the event with a Gaussian filter using the Fast Fourier Transform. The result of the convolution produces extrema called preclusters. In the second step, jet energies are reconstructed starting with the precluster positions and iterating until stable jets are found. The FFTJet algorithm operates with two scale parameters:

1. Pattern recognition scale (s) : 1σ width of the Gaussian filter.

2. Energy reconstruction scale (R) : the jet scale parameter.

The Gaussian filters are useful in multi-scale event reconstruction as they provide stable precluster definitions without introducing spurious structures at coarser scales [65, 66]. In comparison, the cone based algorithms suffer from an ambiguity in pattern recognition. In fig. 5.1, two energy deposits produce up to two preclusters with Gaussian filters (center and right diagrams), while the cone filter produces an additional precluster which does not correspond to any energy deposit. The preclusters constructed with Gaussian filters are stable against soft radiative emissions and collinear splitting of the particles produced in the hadronization. This makes FFTJet algorithm collinear and infrared safe. A Gaussian filter is only sensitive to the large, proximal energy deposits while being insensitive to the distant, low energy emissions. Therefore, these filters are good candidates for the detection of jet-like structures in the energy flow.

In the next section, the operating parameters of the FFTJet algorithm are derived for generic jet reconstruction. The optimal values of these parameters depend on the specific features of the physics process being investigated such as jet multiplicity, kinematics and

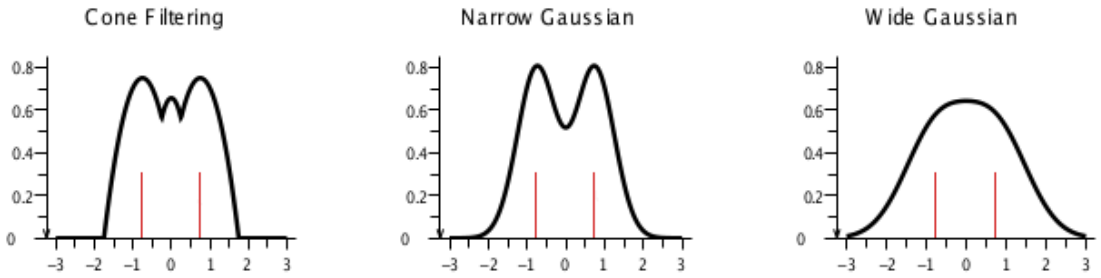


Figure 5.1: The result of convolution is shown in black with the original energy deposits in red. The extrema produced with the cone (left) and Gaussian filters (center and right) are shown [7]. An additional (third) extremum is seen in the cone filter which does not correspond to any energy deposit, unlike the Gaussian filter which produces one or two depending on the scale (s).

event topology. It is possible to fine tune the parameters based on the particulars of the analysis.

5.3.1 Pattern recognition

Pattern recognition is performed at the first stage of jet reconstruction to determine the location of jet-like objects. The energy deposits from the detector are plotted in a 2-D discretized η-ϕ grid and convolved with a Gaussian filter (fig. 5.2). The extrema in the resulting energy density are used as preclusters. The preclusters are produced by applying the Gaussian filter at different pattern recognition scales (s). Fifty scale values are used, in the range $s = 0.087$-0.6, in logarithmic intervals. At this stage, an energy cut is applied to the preclusters for noise suppression[1]. The preclusters that pass this preliminary selection are used in the energy reconstruction step. The calibration of the pattern recognition step involves optimizing the choice of scale and energy cut parameters.

5.3.2 Pattern recognition scale and E_T cut

The preclusters produced at the pattern recognition step are determined by two parameters: the scale (s) and the transverse energy (E_T) cut. The optimum values of these parameters are determined using a simulated $t\bar{t}$ process.

[1]The cut is applied to the precluster magnitude which is related to the precluster energy by eq. 5.1.

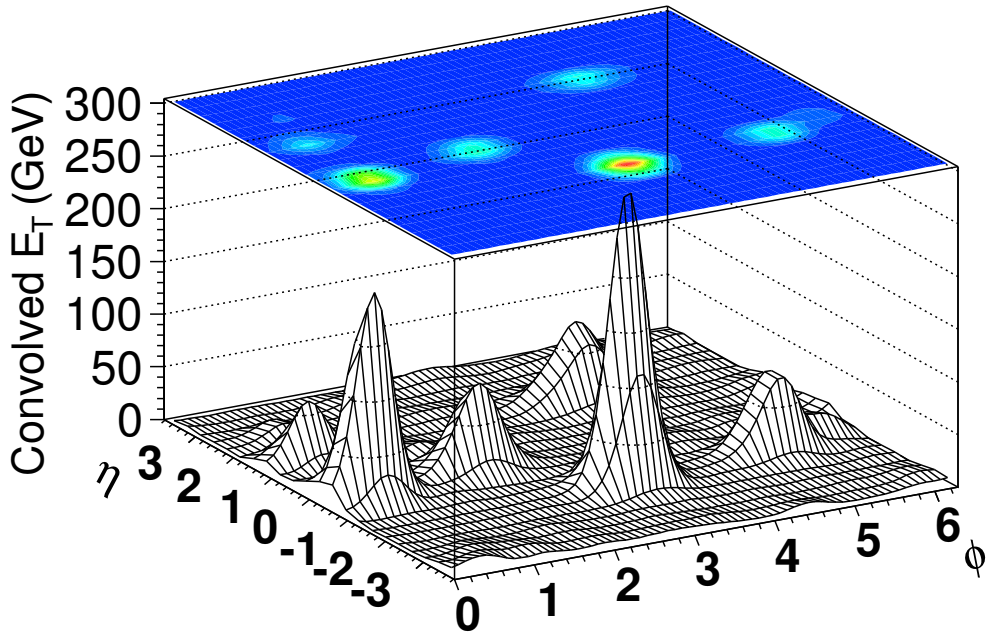


Figure 5.2: The result of convolving energy deposits from the CMS detector with a Gaussian filter. The simulated event is the decay of $t\bar{t}$ into six jets which appear as extrema.

First, the behavior of preclusters is studied as a function of the pattern recognition scale. The preclusters are matched to the partons using the MC information in the event record. The preclusters matched to partons (dR $<$ 0.25) are considered as *signal*; and the unmatched preclusters are considered as *noise*. The average number of *signal* and *noise* preclusters behave differently as a function of scale. The number of *noise* preclusters falls steeply as the event is reconstructed at higher scales; while the number of *signal* preclusters remains constant at low scales ($<$0.22) and falls slowly a higher scales ($>$0.22) (fig. 5.3). The loss of *noise* preclusters is attributed to the fragmented nature of low energy deposits from the noise in the detector while the concentrated energetic deposits of a signal jet persist through the increasing scales. When the pattern recognition scales get high enough (s $>$0.35) two or more neighboring preclusters are subsumed into a single large precluster leading to the loss of *signal* preclusters. The transverse energy represented by a precluster is given by the following relationship [67]:

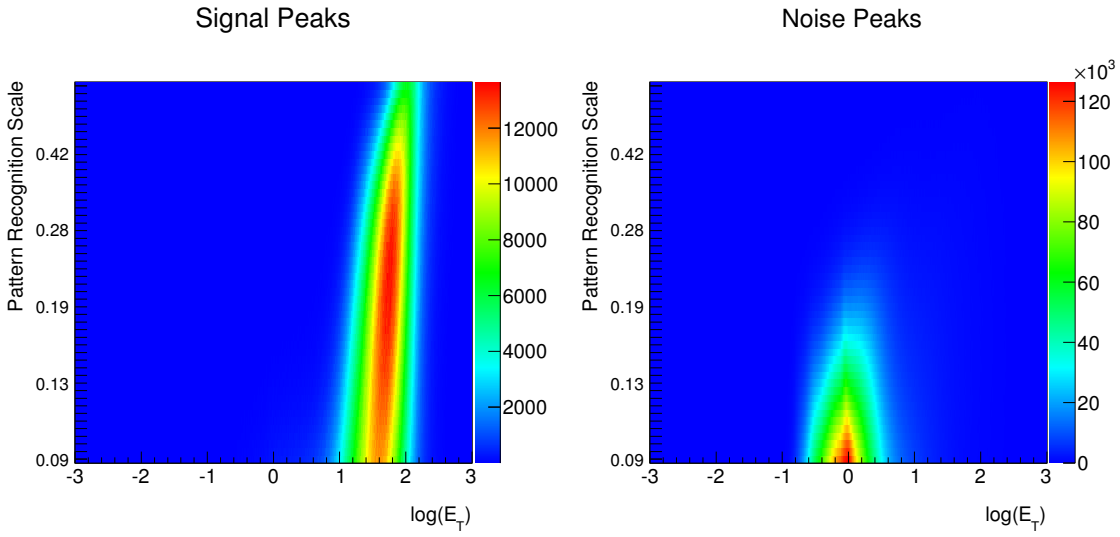


Figure 5.3: The number of *noise* and *signal* preclusters (color scale) as a function of the pattern recognition scale and the transverse energy cut (eq. 5.1).

$$E_T = A\,s^2\,m(s) \tag{5.1}$$

where $m(s)$ is the precluster magnitude, i.e., the maximum convolved energy[2] associated with the precluster, s is the scale and A is the normalization constant. The preclusters with low magnitude can be suppressed by applying a cut on the E_T of the precluster so that only sufficiently energetic preclusters are reconstructed into jets. As in the case of scales, if the E_T cut is too low or too high, the resulting preclusters from the pattern recognition process suffer from either excessive noise or loss of signal efficiency.

Next, a two dimensional optimization is performed for the scale and energy cut parameters. The jets obtained from the reconstruction process were matched to the initial partons and W boson and t quark were reconstructed. The relative mass resolutions of W and t are calculated and used as a figure of merit to gauge the performance of the reconstruction process. The pattern recognition parameters: scale and transverse energy cut, were varied by five steps each to get a 5×5 scan of the mass resolution as a function of scale and energy cut. The scale was varied from 0.12 to 0.23 in logarithmic intervals incremented by a factor

[2] the result of the convolution

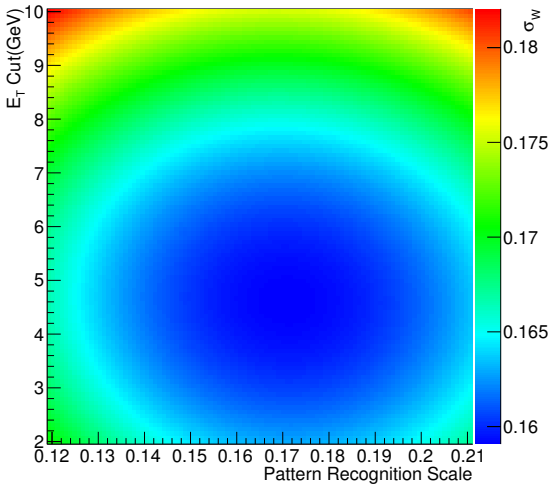


Figure 5.4: Relative W mass resolution (color scale) calculated with jets, plotted as a function of the pattern recognition scale and the transverse energy cut.

of 1.09; and the energy cut was varied from 2 GeV to 10 GeV in 2 GeV increments. The reconstructed W (and t) mass resolution worsens at low scales and at low energy cuts due to increase in *noise*. Similarly on the other extreme, at high energy cut value and at high scales, the loss of *signal* preclusters deteriorate the mass resolution. The optimum W mass resolution is found at scale 0.17, energy cut 4 GeV (fig. 5.4).

5.3.3 Energy reconstruction

The energy reconstruction step follows the pattern recognition. At this stage, jets are reconstructed from the positions of preclusters that are produced at the pattern recognition step. The optimal scale and energy cut derived from the calibration of pattern recognition are applied, and only the preclusters that pass are used for the energy reconstruction. The energy flow from the event is considered in η-ϕ coordinates and the jets are reconstructed by combining all the constituents inside the circular region within a given radius. The jet object is constructed from the four-vector sum of the constituents (eq. 5.2).

$$P_{jet} = \sum_i p_i \ , \quad \text{if } dR_i < R. \tag{5.2}$$

where,

dR_i is the distance in η-ϕ space of the ith constituent from the jet axis,

R is the jet scale parameter,

P_{jet} is the jet four-momentum,

p_i is the four momentum of the ith constituent.

In case of overlap between two jets, the common area shared by the jets is split in two equal regions, and all the constituents in each region are assigned to the closest jet. In order to get a stable jet, the energy reconstruction is repeated over several iterations. After the first reconstruction, the centroid is calculated as energy-weighted sum of the jet constituents. The jet center is shifted to the location of the centroid and the jet energy is recalculated. This process is repeated until the centroid locations from two consecutive iterations converge.

5.3.4 Jet energy corrections

The jet energy measurement from the detector is biased by several experimental factors such as non-uniform energy response of the calorimeter and multiple pile-up interactions in the event. The jet energy corrections are applied to account for these effects and correct, on average, the measured jet energy to the original particle-jet energy produced in the primary interaction [68]. These corrections are derived using simulated samples and verified using measurements in data. A small difference in jet calibrations is found between simulation and data, and rectified by applying residual corrections to jets reconstructed from data.

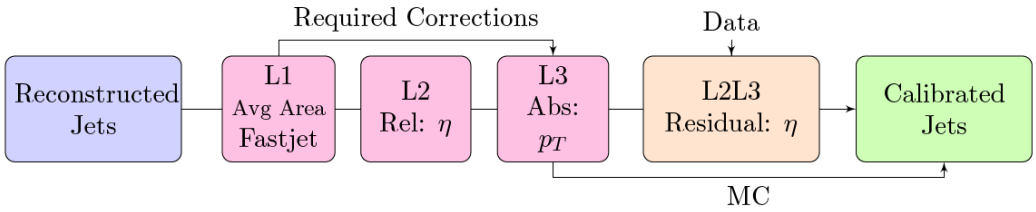


Figure 5.5: The jet corrections are applied in the following order: pile-up corrections (L1) and jet energy response corrections (L2L3). Jets reconstructed from data are corrected with an additional, residual factor (L2L3 residual); in MC this step is not applied.

The jets are calibrated using a factorized approach to jet corrections. A sequence of corrections is applied to independently correct for different effects listed below (fig. 5.5).

- Offset (L1): To correct for additional interactions which simultaneously occur with the primary (hard) interaction and contribute to the reconstructed jet energy.
- Relative (L2): To correct for the variable energy response as a function of jet pseudorapidity (η).
- Absolute (L3): To correct for the variable energy response as a function of jet transverse momentum (p_T). L2 and L3 are applied in the same step.
- Residual (Res): The L2 and L3 steps are derived with MC and applied to data. A residual correction is applied to jets in data to account for small discrepancies between jet calibrations in data and MC.

In this study, all jet corrections are derived as function of the jet scale parameter (R) in the range 0.2–1.0.

Offset corrections

The offset corrections are derived by estimating the energy density (ρ_{PU}) due to pile-up interactions. The number of pile-up interactions differs from event to event. Hence, a "pile-up estimate" quantity is calculated by analyzing all the energy deposits in the detector. A pre-determined calibration curve is used to map this "pile-up estimate" to ρ_{PU}. The pile-up energy contribution to the jet energy is calculated using ρ_{PU} and the area of the jet. The calibration curve is calculated as follows:

1. Pure pile-up sample is mixed with pure signal sample to simulate the data-taking conditions.
2. The energy deposits in the detector are smoothed with a narrow band Gaussian filter.

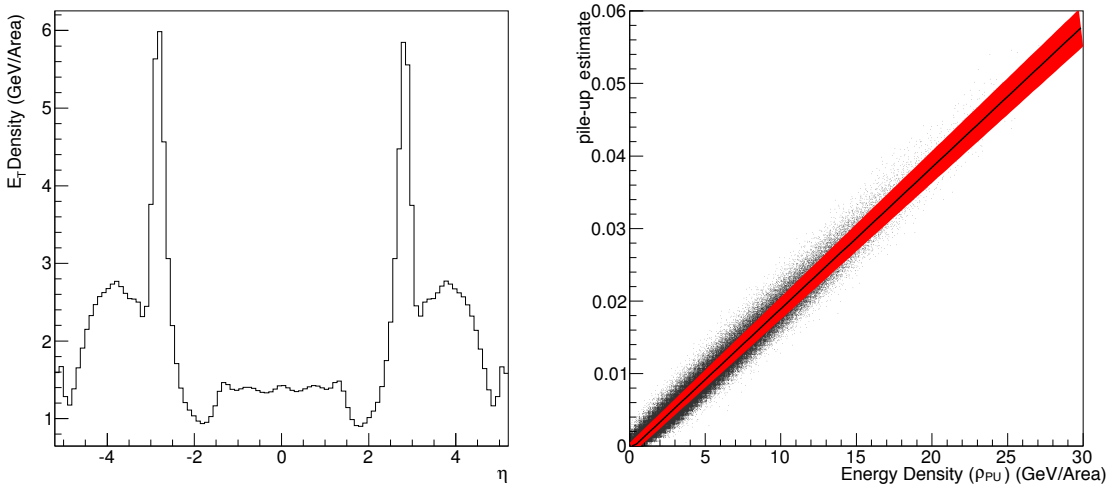


Figure 5.6: The average transverse energy density as a function of η is shown in the left figure. The correlation between the pile-up estimate (y) and the energy density (x) is shown in the right figure. The area is defined in $\eta - \phi$ space.

3. The deposits are scaled to correct for the η dependent response.

4. Finally, the deposit with 0.4 quantile energy is used as the "pile-up estimate". The calibration curve plotted in fig. 5.6 shows the dependence of the "pile-up estimate" on the known energy density ρ_{PU}. The solid black line indicates the central value with the uncertainty band in red.

L2L3 Corrections

The L2L3 corrections are applied to correct for jet energy response of the detector. The corrections are calculated using simulated QCD events which contain the particle and detector jets. The particle jets are reconstructed with the particles produced in the primary interaction, while the detector jets are reconstructed from the simulated energy deposits in the detector. Hence, the energy response of the detector is factored into the detector jet energy. The particle jets are matched to the detector jets and the correction factor is determined from the median value of the jet response ($p_T^{detector\ jet}/p_T^{particle\ jet}$) distribution. The correction factor is calculated as:

$$\text{Correction factor} = \frac{1.0}{\text{Jet response}}. \tag{5.3}$$

The jets were distributed into a two-dimensional grid of p_T and η bins with the ranges of 10 to 1000 GeV and -5.5 to 5.5, respectively. The correction procedure was repeated for ten discrete values of jet scale from 0.2 to 1.0 (fig. 5.7). The correction factor is applied as a function of jet p_T, jet η and jet scale. A detailed discussion on the derivation of jet energy corrections can the found in the references [46, 51, 68].

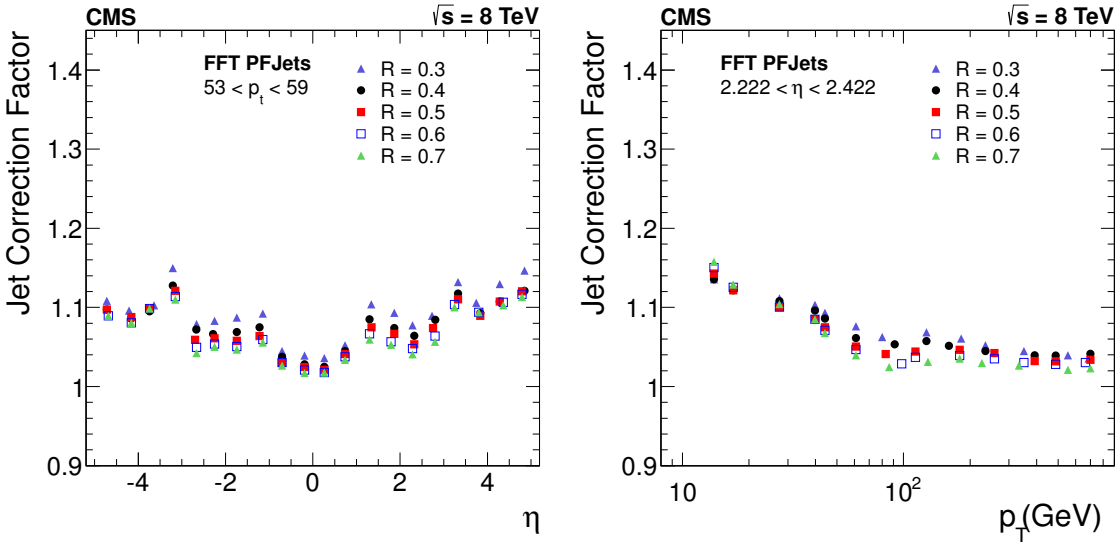


Figure 5.7: Jet energy correction factors plotted as a function of jet η (left) and jet p_T (right) for the jet scales 0.3–0.7.

Residuals Corrections

The residual corrections are derived with the MPF (Missing Energy Projection Factor) method which is based on the law of conservation of momentum. In an interaction producing two back-to-back particles, such as a dijet or a Z+jet process, the momentum of the two objects (reconstructed in the detector) should be perfectly balanced in the transverse plane. In these cases, any measured value of the missing energy E_T^{miss} can be attributed to the mismeasurement in the energy of the reconstructed objects. The two objects often do not have the same energy response. For example, the energy of the jet in the barrel region

($|\eta| < 1.3$) or Z decaying into two muons can be reconstructed with much better precision than the energy of the jet in the endcap or forward region. The well-reconstructed object is classified as the *tag* object and used to determine the energy response for the other (*probe*) object. The correction factors are derived by solving the following equations:

$$\overrightarrow{p}_T^{tag} + \overrightarrow{p}_T^{probe} = 0, \tag{5.4}$$

$$\mathscr{R}_{tag} \cdot \overrightarrow{p}_T^{tag} + \mathscr{R}_{probe} \cdot \overrightarrow{p}_T^{probe} = -E_T^{miss}, \tag{5.5}$$

where, $\mathscr{R}_{tag}$ and $\mathscr{R}_{probe}$ are the energy responses of the probe and tag objects. Solving the equations above yields the MPF response ($\mathscr{R}_{MPF}$),

$$\mathscr{R}_{probe} = \mathscr{R}_{tag} + \frac{E_T^{miss} \cdot \overrightarrow{p}_T^{tag}}{(\overrightarrow{p}_T^{tag})^2} = \mathscr{R}_{MPF}. \tag{5.6}$$

The ratio of MPF response in data and MC is used to get the final residual correction,

$$Residual\ Correction = \frac{\mathscr{R}_{MPF}^{Data}}{\mathscr{R}_{MPF}^{MC}}. \tag{5.7}$$

The residual corrections are calculated in the following steps:

- L2Residual corrections set the jet response of the detector equal to that of the barrel region ($|\eta| < 1.3$), by applying an η-dependent correction factor (fig. 5.8). The corrections are calculated with dijet events in which at least one jet is in the barrel.
- L3Residual corrections set the jet response in the barrel to the absolute scale, by applying a constant correction factor (table 5.1). L3 corrections are calculated using Z+jets events, with Z$\rightarrow \mu\mu$.
- The final residual correction is calculated as L2Residual$\times$L3Residual.

The L2res correction is derived using the Kalibri package developed by the University of

Hamburg CMS group [69]. The p_T dependent L3 correction were derived using a similar package, Calibration Framework (CALIBFW) developed by the Karlsruhe (KIT) group.

As in the case of L2L3 corrections, the residual corrections were derived as a function of jet scale within the range R = 0.2–1.0.

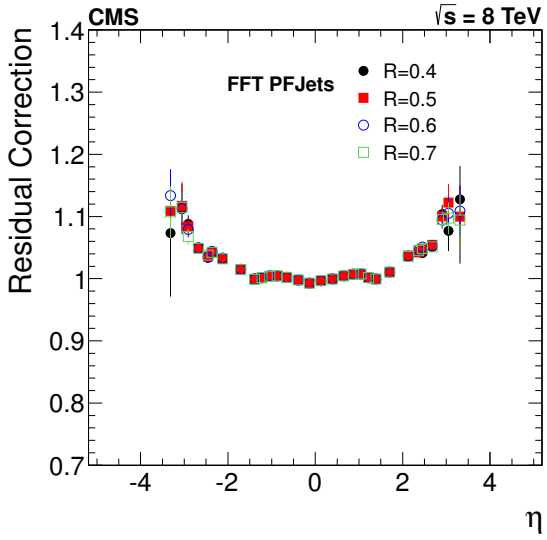


Figure 5.8: L2Residual corrections for jet scale R = 0.4 - 0.7.

5.4 Jet Validation

The jet calibrations are validated by reconstructing a well-understood physics process, semi-leptonic $t\bar{t}$ decay ($t\bar{t} \to W^+W^- + b\bar{b} \to \mu\nu + q\bar{q} + b\bar{b}$). In this process there are four jets in the event, two originating from b quarks and other two from light u,d,s, and c quarks. The charged lepton present in the final state can be used to isolate this process from most of the background. In this study, the performance of the calibrated jets in MC samples and data is evaluated. The analysis is repeated with the anti-k_T jet algorithm and the results are compared with FFTJet. The data used for the study were collected during the 2012 run at the LHC at $\sqrt{s} = 8$TeV, with the CMS detector. The MC samples were generated using PYTHIA6 (datasets listed in table B.2). The rest of the section describes the selection requirements applied to isolate $t\bar{t}$ decay from the background processes.

Jet Scale	L3 Residual Correction
0.20	1.016 ± 0.002
0.25	1.017 ± 0.003
0.30	1.015 ± 0.002
0.35	1.014 ± 0.005
0.40	1.015 ± 0.004
0.50	1.018 ± 0.008
0.60	1.021 ± 0.008
0.70	1.032 ± 0.011
0.80	1.018 ± 0.010
1.00	1.018 ± 0.015

Table 5.1: L3 Residual correction factors as a function of jet scale (R).

The events were selected using high p_T, isolated muon triggers HLT_IsoMuon24 and HLT_IsoMuon30[3]. A veto was applied on any additional leptons (μ, e) in the event. A tight muon selection is applied, requiring the muon candidate to be a global muon (i.e., tracks in inner tracker and drift tubes) with a good track fit (χ^2/ndof < 10). A muon isolation criterion was applied requiring that any additional energy (charged + neutral hadronic energy), deposited within the radius of 0.3 around the muon must be less than 12% of the muon transverse momentum [70]. The Combined Secondary Vertex (CSV) algorithm was applied to detect jets originating from b quarks. The CSV algorithm uses displaced vertex and tracking information to build a likelihood-based discriminant to distinguish between jets from b quarks and those from charm or light quarks and gluons. A medium operating point was used to isolate b-jets from the light jets[4] [71]. Every event was required to have at least two b and two light jets. The two leading light jets are used to construct the hadronically decayed W ($W^{hadronic}$). The combination of $W^{hadronic}$ and a b-jet, yielding the

[3]Isolated muons with p_T greater than 25 and 30 GeV, respectively

mass closest to the top mass (173.2 GeV), is used to construct t. The following selection is applied:

1. Jet p_T >30 GeV, $\eta < 2.4$
2. Muon p_T >30 GeV, $\eta < 2.1$
3. Reconstructed top $133.2 < t_{mass} < 213.2$ GeV
4. $\Delta\phi$ (leading light jet, E_T^{miss}) >2.5

Reconstructed W mass and p_T are plotted in fig. 5.11 with FFTJet and anti-k_T algorithms. The average energy fraction carried by different types of jet constituents is shown in fig. 5.10. The contribution of QCD events was found to be negligible and is not included in these plots.

5.5 Results

The validation study tests the jet calibration scale derived for the FFTJet algorithm. The semi-leptonic $t\bar{t}$ decay process is reconstructed, and the results from MC and data are compared. A well-calibrated jet sample should yield identical results. The control variables, muon p_T, missing energy, and number of b-tagged and anti b-tagged jets, are plotted in fig. 5.9. The muon p_T and the missing energy spectra are correctly predicted by the simulation, while the number of b-tagged and anti b-tagged jets show a lower yield in data. These are followed by individual jet constituents, reconstructed by the particle-flow algorithm, in fig. 5.10. The average fraction of jet energy carried by each constituent in data is shown to be consistent with the simulation. Fig. 5.11 shows the invariant mass and p_T of the reconstructed W boson. This analysis is performed with the two algorithms, FFTJet and anti-k_T. The calibration scale for the anti-k_T algorithm is derived independently

[4] At the medium operating point of the CSV discriminant (0.697) the id- and mistag- efficiencies are 70% and 1%, respectively. Therefore, the probability of incorrectly assigning jet flavor is low, and the samples can be assumed to be relatively free of contamination.

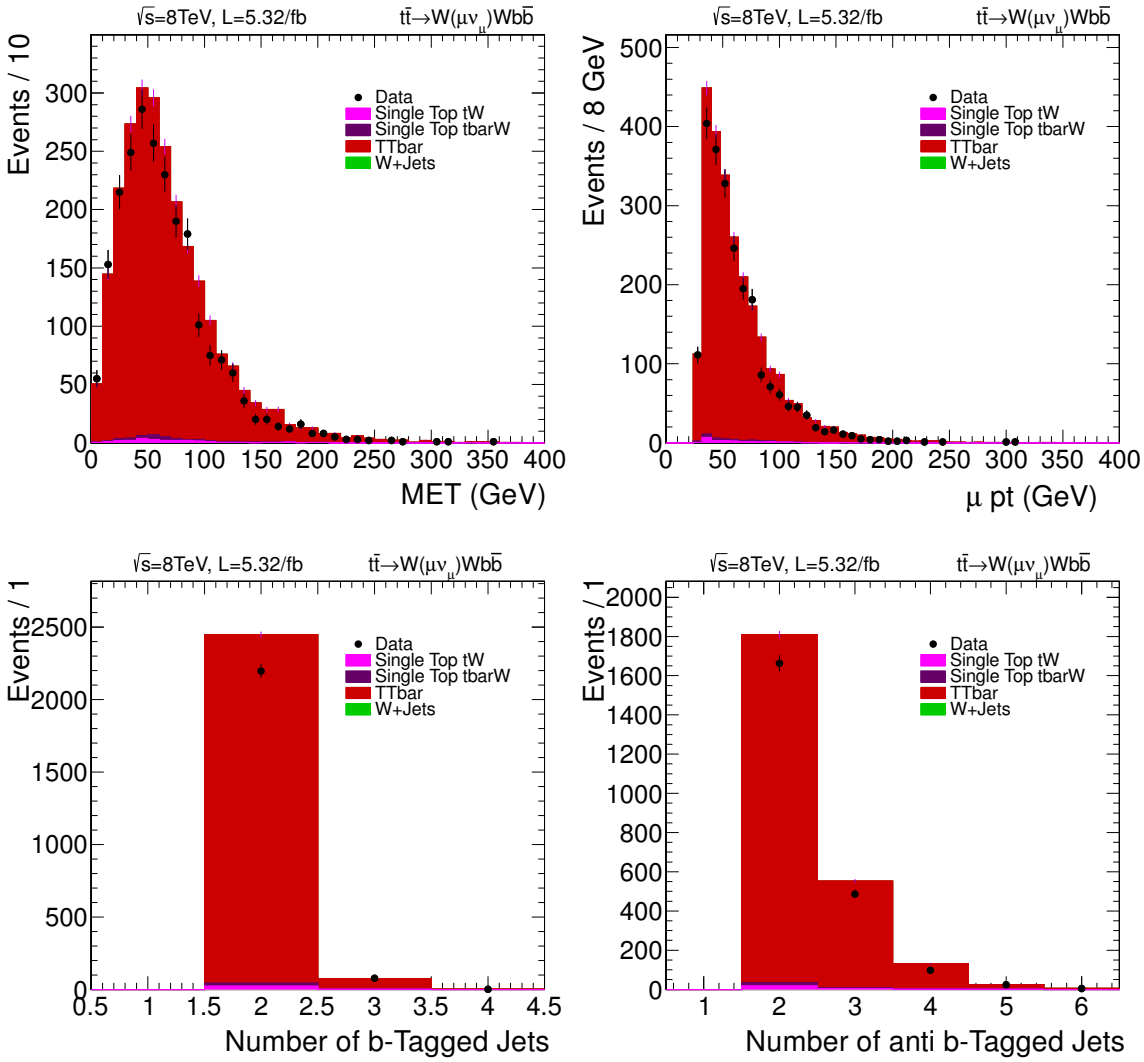


Figure 5.9: Selection variables for the $t\bar{t}$ sample are plotted: E_t^{miss} (MET), muon p_T, and the number of jets identified as b-jets and light-jets.

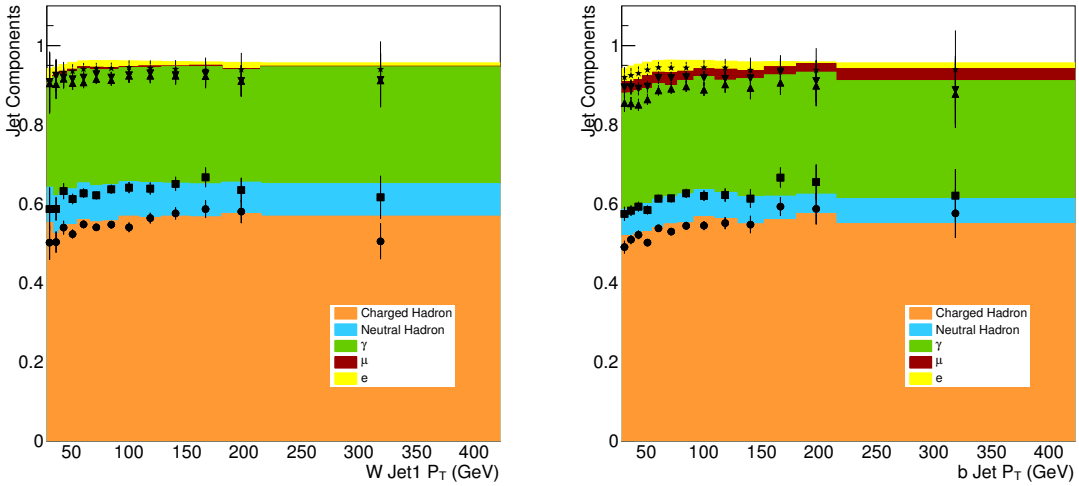


Figure 5.10: Jet constituents for leading light jet and b jet reconstructed using FFTJet algorithm.

by a CMS working group and used in several analyses. The two algorithms yield similar results, therefore we conclude that the calibration scale derived for FFTJet algorithm is correct.

5.6 Summary

The FFTJet algorithm has been implemented within the CMS software framework. The jet calibration scale is developed using the procedure recommended by CMS, and validated using a benchmark study. In this algorithm, the reconstruction process is performed by analyzing the energy flow with Gaussian filters to search for jet-like objects. The results are found to be identical to a standard jet algorithm. The reconstruction of the four-jet scenario in the $t\bar{t}$ decay did not show an improvement due to the advanced jet recognition technique employed by FFTJet. This could be explained by the relatively low number of pile-up interactions and jet multiplicities observed at $\sqrt{s}$ =8 TeV. However, the greatest advantage of FFTJet could be in the reconstruction of boosted objects and high jet multiplicity final states. In such cases, the performance of the sequential recombination algorithms is poor in distinguishing jets from the noisy background. In the upcoming run at the LHC, the