



POSTGRADUATE STUDIES PROGRAMME
IN PHYSICS

Department of Physics
School of Sciences
University of Ioannina

POSTGRADUATE STUDIES

**Physics with Gluon Jets at 13 TeV at the
CMS experiment**

Peter Kornezos
Registration Number: 811

Supervisor: John Strologas

Ioannina
November 11, 2024

Abstract

In this thesis we present a study quark and gluon jets, as a measurement of gluon-jet fraction in the inclusive multi-jet sample. First, we measure the double-differential inclusive-jet cross section, as a means to validate our jet data. Subsequently we measure the gluon-jet fraction in this data. We use integrated luminosity of 36.31 fb^{-1} , from the 2016 run of the LHC accelerator, collected by the CMS experiment.

Contents

1	Introduction	4
1.1	Overview of the SM	4
1.2	Quantum Chromodynamics	5
1.3	Color Confinement	6
1.4	Hadronic Jets	6
1.5	Hard Process	7
1.6	Showring and Hadronization	7
1.7	Importance of Simulation	9
1.7.1	Madgraph	9
1.7.2	Pythia and Herwig	10
1.7.3	Geant4 Detector simulation	10
1.8	The CMS Detector	10
1.9	CMS Coordinate System	12
1.10	Jet Reconstruction Algorithm	12
1.11	Partonic Origin of Jets	13
2	Analysis Strategy	15
2.1	Purpose of the Study	15
2.2	Discriminating Variables	15
2.3	QGL calculation	18
3	Data and Simulation samples	21
3.1	Data Samples and Requirements	21
3.1.1	Triggers	21
3.1.2	Jet Identification Requirements	23
3.1.3	MET filters	24
3.1.4	Prefire weights	24
3.1.5	Hot/Cold-channels	25
3.2	Merging of data	25
3.2.1	Use of Prescales to Merge the Data	25
3.2.2	Use of Effective Luminosities to Merge the Data	26
3.2.3	Comparison	27
3.3	MC Selection and Requirements	29
3.3.1	H_T -bins	29
3.4	Data-MC comparison	30

4	Double-Differential Inclusive Jet Cross Section Measurement	31
4.1	Unfolding	31
4.2	Result	33
5	Gluon-Jet Fraction Measurement	35
5.1	Discriminating Variables	35
5.2	Gluon-Jet Fraction Measurement	38
5.3	Unfolding of Gluon-Jet Fraction	39
5.4	Systematic Uncertainties	41
5.5	Results	42
6	Discussion	46

Chapter 1

Introduction

1.1 Overview of the SM

The Standard Model (SM) is our current best theory for the microcosm [1]. It describes three of the fundamental forces: the electromagnetic, the weak and the strong interactions. It consists of 12 fermions (six leptons and six quarks), four gauge bosons and one scalar boson. The scalar boson is the Higgs boson and the four gauge bosons are the photon, which carries the electromagnetic interaction, the W^\pm and Z bosons, which carry the weak interaction, and the gluons, which carry the strong interaction. The charged particles interact electromagnetically, the leptons and quarks also interact weakly, and the quarks also interact strongly. Massive particles interact with the Higgs boson. In extensions of the SM there is also the graviton, the carrier of the gravitational force, which has not been detected. In this thesis, we will concentrate on the quarks and the gluon, which are described by the Quantum Chromodynamics (QCD) [2], the part of the SM that covers the strong interactions. Fig. 1.1 shows a graphical overview of the SM.

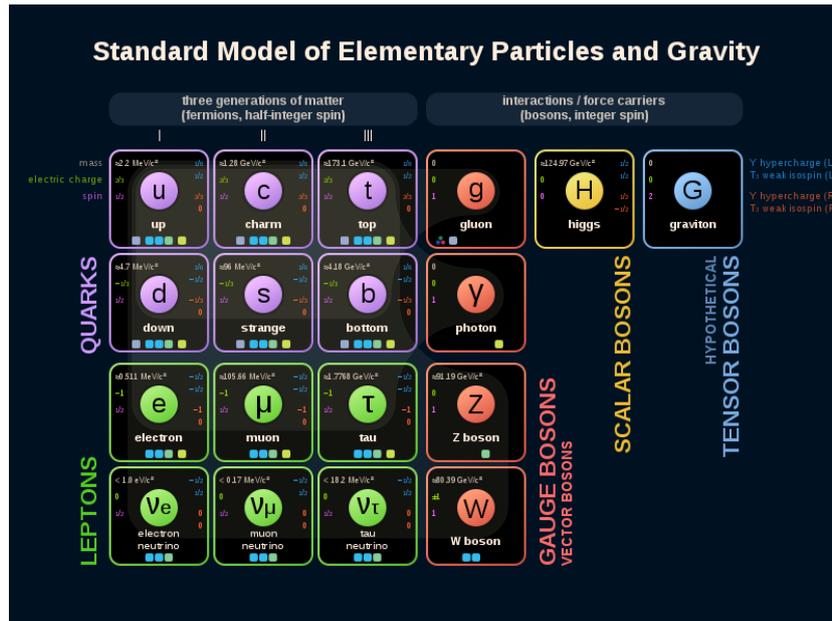


Figure 1.1: Graphical Representation of the SM.

1.2 Quantum Chromodynamics

Within the framework of QCD, particles interact strongly if they possess the quantum property of color. There are 3 basic colors: red, green and blue. A particle can also have different linear combinations of these colors. White, or colorless, is the combination of all three colors or the combination of color with the corresponding anti-color.

Quarks are fermions with a fractional spin of $\frac{1}{2}$. They come in three generations with an increasing order of quark mass. Each generation has two quarks one with a charge of $+\frac{2}{3}$ and one with $-\frac{1}{3}$. Another major property of quarks is their flavor: each one has a different flavor named after the quark itself. The strong and electromagnetic interactions conserve every flavor. On the contrary, weak interactions do not conserve them. The names of the quarks are down (d), up (u), strange (s), charm (c), bottom (b) and top (t). Each one could come in different colors. Quarks are combined in order to make other particles, named "hadrons". In addition to the quarks, the hadrons contain gluons that hold the quarks together. There are two categories of hadrons: the mesons (quark-antiquark pairs) and the baryons (triplets of quarks or antiquarks). There is also evidence [3] of the existence of tetraquarks made of 2 quarks and 2 anti-quarks and pentaquarks made up of 4 quarks and 1 anti-quark or 4 anti-quarks and 1 quark. The most common baryons in nature are the proton (uud) and the neutron (udd). Of course everything applies to the anti-quarks as well, and their antiparticles. In Fig. 1.2 we present the different quarks in their respective generation with their charge and flavor.

		Quark classification							
		q	Q	D	U	S	C	B	T
First generation	{	d	$-1/3$	-1	0	0	0	0	0
		u	$2/3$	0	1	0	0	0	0
Second generation	{	s	$-1/3$	0	0	-1	0	0	0
		c	$2/3$	0	0	0	1	0	0
Third generation	{	b	$-1/3$	0	0	0	0	-1	0
		t	$2/3$	0	0	0	0	0	1

Figure 1.2: Quarks properties.

The gluons are massless, neutral, spin-0 bosons. They only interact with colored particles, which are the quarks and themselves. They come in eight colors and have never been observed as free particles, so they cannot be white/colorless. In addition to flavor, strong interactions also preserve color. The strong interactions could be derived from the QCD Lagrangian.

$$\mathcal{L} = [i\hbar c\bar{\psi}\gamma^\mu\partial_\mu\psi - mc^2\bar{\psi}\psi] - \frac{1}{16\pi}G^{\mu\nu} \cdot G_{\mu\nu} - (q\bar{\psi}\gamma^\mu\lambda\psi) \cdot A_\mu \quad (1.1)$$

Where ψ is the quark field, γ^μ are the Dirac matrices, $G^{\mu\nu}$ is the asymmetric field tensor and A^μ is a gauge field associated with the gluons.

1.3 Color Confinement

Because free quarks or gluons have never been observed, the concept of color confinement was introduced, according to which, the quarks and gluons are always confined within hadrons and can never be observed as free particles. When a quark or gluon tries to get separated from its hadron new quarks and gluons are produced making new hadrons. This actually converts part of the kinetic and potential energy of the particle into its the new hadrons' mass. Color confinement does not prevent us from studying quarks and gluons. We can still perform experiments, such as hadron collisions, in order to reveal their inner structure and internal interactions.

1.4 Hadronic Jets

Hadronic jets are showers of particles, mostly hadrons, photons and leptons, usually forming a narrow cone. The most important thing about hadronic jets is that they are a direct result of QCD and by studying them we learn about the underlying physics that creates them [4]. In this thesis we study jets that are generated from proton-proton collisions at the Large Hadron Collider (LHC) at CERN. In this chapter we will discuss the very basic concepts of the creation of hadronic jets through proton-proton interactions.

1.5 Hard Process

In particle physics, one way to gain information about the smallest scales is by colliding particles. The higher the energy of the collision, the deeper we can probe inside the colliding particles. When we collide protons at high energies, interactions take place between the protons' constituents, the quarks and gluons, also known as partons. The interaction between the partons of different protons is called "hard process". Most of the times we only care about the strong interaction since it has a much higher coupling constant than the electromagnetic and the weak interactions. Nevertheless, sometimes we may include the other interactions, depending on the process and the desirable precision. The parton interactions give us an insight to the inner workings of QCD, with corrections coming from electroweak processes. The hard process can be represented by a simple Feynman diagram and calculated by using perturbation theory. We keep in mind that any outgoing partons in the Feynman diagram are not detected as such, but as hadronic jets, as we will describe below. In Fig. 1.3 we show an illustration of an example of the hard process.

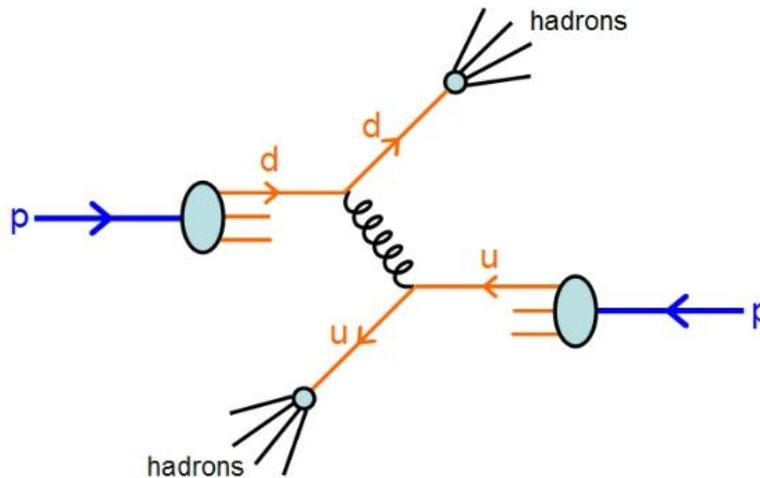


Figure 1.3: Illustration of the hard process.

1.6 Showering and Hadronization

The outgoing partons from the hard process eventually develop to hadronic jets through two sequential processes. First, we have the parton shower [5], where a lot of new partons are created. These particles are colored and cannot be free. Instead, they are combined to form bound hadronic states during the next process of hadronization. During the parton shower, new partons are created in two major ways. Through radiation of gluons from already existing partons and through quark-antiquark creation. The former can be initiated both from quarks and gluons, since gluons couple with both

of them, and the latter only from gluons splitting. The quark-antiquark pair can come into existence due to a particular property of the strong interaction, according to which the potential energy between two quarks is proportional to their distance, for small distances. As a result, when the distance between them and the potential energy is large enough, a quark-antiquark pair can be generated out of vacuum, converting the potential energy into its mass equivalent. Alternatively, a radiated gluon can spontaneously turn into a quark anti-quark pair. These processes turn some of the initial kinetic and potential energy into its equivalent of mass by making new partons.

After the parton showering, and during the hadronization stage, all the generated partons are grouped into pairs and triplets to produce mesons and baryons, respectively. These hadrons are white in color and thus free particles. Moreover electromagnetic and weak interactions also take place at this stage, which produce photons and leptons. From the leptons, neutrinos are impossible to detect with our detector and thus contribute to missing energy and momentum. Furthermore, the produced hadrons can potentially be highly unstable and quickly decay to other particles or be excited and emit photons by de-exciting. All of these particles, when detected, appear in the form of hadronic jets. The complexity of the aforementioned processes can be illustrated in Fig. 1.4.

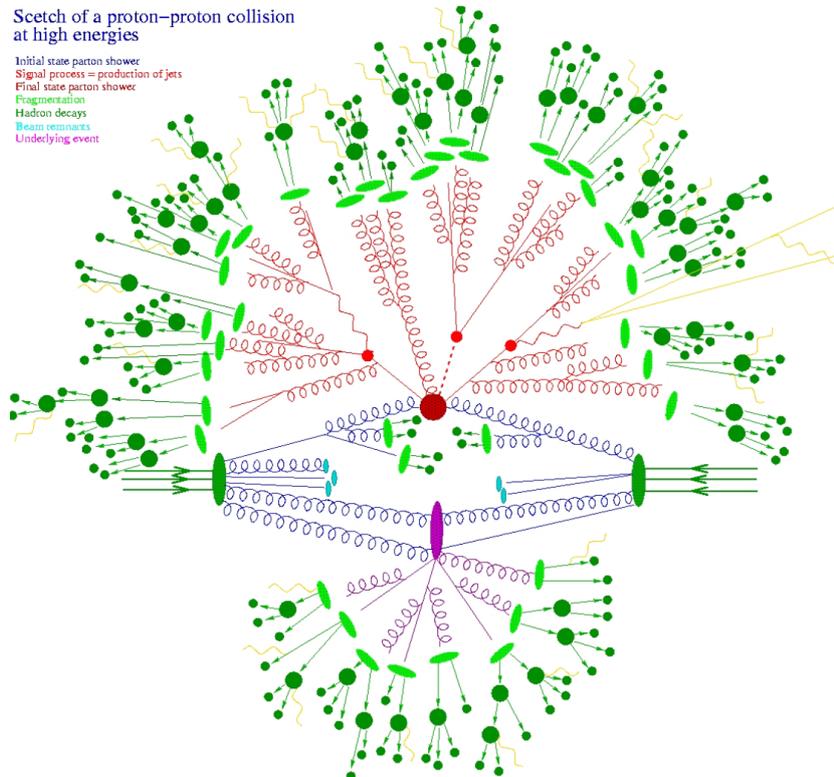


Figure 1.4: Illustration of the complexity of parton showering.

An important fact to mention is that the initial outgoing partons generally have high momentum with respect to the lab's inertial frame of reference.

The final-state hadrons that are generated will have a combined momentum that will be close to that of the original partons. As a result, the jets of hadrons will have a particular direction in space as opposed to having a spherical symmetry, this is why we say hadronic jets form narrow cones.

1.7 Importance of Simulation

Simulations are an integral part of modern physics since they can offer important insight on future experiments, provide experimental observables for theories, help understand the performance of detector systems, and facilitate their design. Most simulations are based on the Monte Carlo (MC) method. In High Energy Physics, we also use MC event generators to simulate the interaction between particles. Moreover some physical processes may not be possible to be calculated analytically and this is another benefit of MC simulations. Moreover, the simulations can help us understand the kinematics of signal and background processes in order to optimize our event selection.

1.7.1 Madgraph

Madgraph [6] can generate events for any high energy physics process by integrating matrix elements through adaptive integration (VEGAS [7]). We are not going to expand on the technical side of how one can use Madgraph. We discuss here the use of Madgraph in proton-proton collisions with jet final states. Generally, a proton-proton collision can result in a hard process with any number of final-state quarks or/and gluons. Different final states have different production cross sections and, due to the perturbative nature of jet production, we do not have to include all possible partonic final states. Madgraph only generates a partonic final state and not the free final-state hadrons or the reconstructed hadronic jets. The cross section $pp \rightarrow \text{partons}$ is determined by integration of the partonic matrix elements with the parton distribution functions (PDF), which correspond to the probability for an incoming colliding parton to have a particular fraction of its parent proton momentum. Because the integration is adaptive, it concentrates on kinematic phase-space regions where the integrand is larger, so it can be used as a particle generator. The production of partons (jets) in proton-proton collisions follows a steeply falling distribution as a function of the partons (jets) transverse momentum (p_T). As a result, it is more efficient to generate jets in bins of outgoing particles p_T , so that we maintain comparable statistics throughout the spectrum. The final merging of these separate MC datasets is done by applying weights on the events that are equal to the Madgraph-calculated cross section divided by the number of produced events (i.e., the inverse of luminosity).

1.7.2 Pythia and Herwig

Pythia [8] and Herwig [9] are programs for the event generation of high-energy collisions. We use it for the simulation of showering and hadronization because Madgraph deals only with the hard process. More specifically they take the partonic output of the hard process from Madgraph and by using MC methods implement the complete showering and hadronization. The result of this whole process is randomly generated events containing observable final-state particles. These particles are also called generated particles. Each event has its own weight, which is propagated from the Madgraph weight. The reason to have two different simulators for showering and hadronization is their different internal model. Pythia uses p_T -ordered showering of outgoing partons, and subsequent hadronization using the Lund string model [10] while Herwig uses angular-ordered showering and hadronization based on the cluster model [11]. Since there are certain aspects of these processes that are non-perturbative, in contrast to the hard process which is perturbative, having two models is necessary for calculating the respective model systematic error.

1.7.3 Geant4 Detector simulation

The output of Pythia's or Herwig's showering and hadronization is final-state particles, not all of, which will be observed. To complete the simulation of what we expect to see in the laboratory, we need to include a simulation of a detector. This is simulated by Geant4 which is a full material-radiation simulation which is the most accurate simulation for the CMS [12] available.

1.8 The CMS Detector

The CMS detector consists of five different layers. Listed from the inside (collision point) to the outside they are the Silicon Tracker, the Electromagnetic Calorimeter, the Hadronic Calorimeter, the Magnet and the Muon Detectors. Each of these layers plays its own part in the process of detecting a particle. Different categories of particles leave different trails in the detector and thus can be identified. Photons interact with the electromagnetic calorimeter but do not curve due to the magnetic field, while electrons both interact with the electromagnetic calorimeter and curve due to the magnetic field. Muons, on the other hand, being much heavier than electrons, are more penetrative and reach the outer part the Muon detectors. Similarly, charged and neutral hadrons interact with the hadronic calorimeter but the charged ones also curve due to the magnetic field. Charged particles also interact with the silicon tracker which can reconstruct their trajectories with high precision. All other particles are reconstructed from the above. Fig. 1.5 shows an overview of the CMS detector, while Fig. 1.6 demonstrates how different particles interact with the detector.

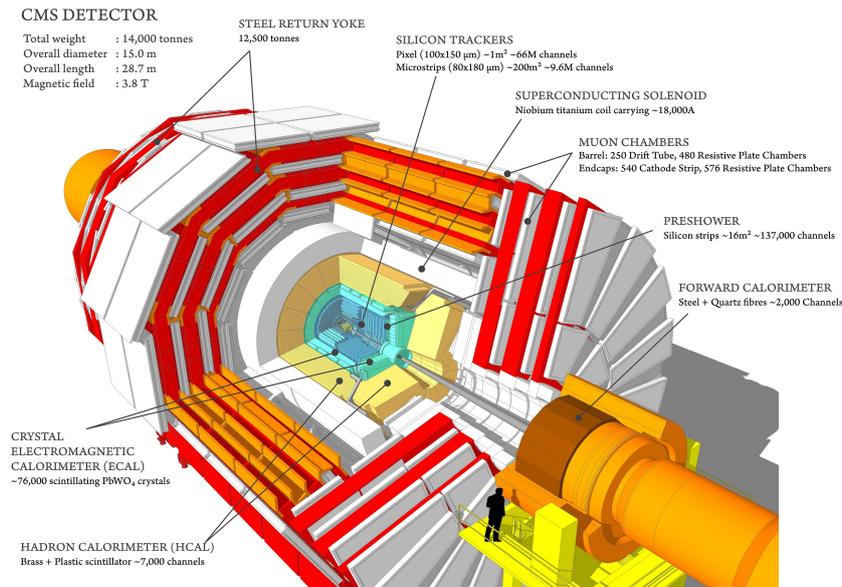


Figure 1.5: The CMS Detector.

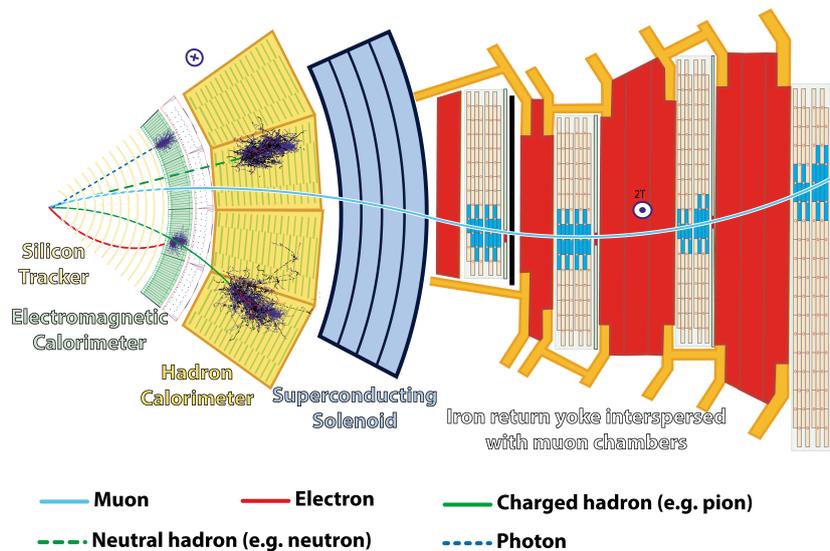


Figure 1.6: Representation of particle interactions at CMS.

The detector is able to detect stable particles that reach its layers and interact. Any highly unstable particle will decay before interacting and neutrinos will almost never interact, thus will not be detected. In this thesis we are going to consider only particles that can be detected by CMS.

The detector is not a perfect instrument. It has limited resolution, which naturally causes deviations between the detected and actual properties of particles. The differences between truth and observation can be studied with MC-generated events that are passed through a detector simulator.

1.9 CMS Coordinate System

Each detected particle has at least three parameters to define its momentum. These are the transverse momentum p_T , the azimuth angle ϕ and the pseudorapidity

$$\eta = -\ln[\tan(\theta/2)] \quad (1.2)$$

where θ is the polar angle, with the z -axis being towards the counterclockwise-beam direction. Sometimes, instead of the pseudorapidity η we use the rapidity

$$y = \frac{1}{2} \ln \left[\frac{E + p_z}{E - p_z} \right] \quad (1.3)$$

where E is the energy of the particle and p_z is the component of its momentum on the z -axis. The benefit of using the rapidity instead of the pseudorapidity is that the difference in rapidities Δy remains invariant under boosts along the z -axis. But the rapidity is harder and less accurate to define since it requires knowledge of the energy E of the particle. The difference between the rapidity y and the pseudorapidity η diminishes when the mass of the particle is a much smaller compared to its energy; for zero mass or infinite energy the two are identical. Fig. 1.7 shows the CMS coordinate system.

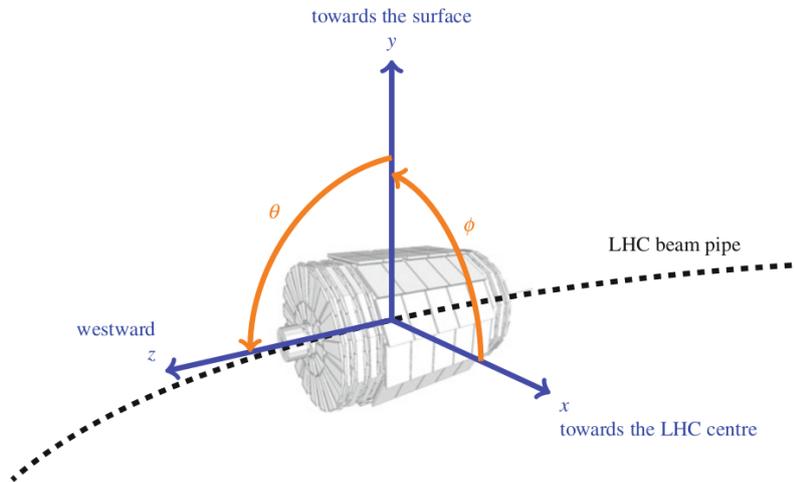


Figure 1.7: The CMS coordinate system.

1.10 Jet Reconstruction Algorithm

Unlike the individual detected particles that are very well defined, the definition of the hadronic jets, as a group of particles, depends on the algorithm used to cluster them. If particles are added or removed from a particular jet, the shape, direction and momentum of the jet alters. One of the most popular jet-reconstruction algorithms is the anti- k_T algorithm [13].

This algorithm treats all particles as pseudo-jets and when certain criteria are met it merges two pseudo-jets in one. More specifically, for all pairs of pseudo-jets the following quantities are calculated:

$$d_{ij} = \min(1/p_{T,i}^2, 1/p_{T,j}^2) \Delta R_{ij}^2 / R^2 \quad (1.4)$$

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \quad (1.5)$$

where p_T , y and ϕ have already been discussed, R is a user-defined parameter, which indirectly determines how wide the jets are going to be, and i and j are the indices of two pseudo-jets. Then it compares d_{ij} with

$$d_{iB} = 1/p_{T,i}^2 \quad (1.6)$$

and if $d_{ij} < d_{iB}$ then the i and j pseudo-jets are merged, otherwise they stay separated. Usually we will sum the 4-momenta of the two pseudo-jets to perform the merge (but other options are available). The result of the algorithm is that the highest- p_T particles group their lower- p_T neighbors around them to form a jet. This characteristic makes the anti- k_T algorithm preferable compared to the other algorithms, since it provides a better correspondence with the originating parton and also gives jets with more regular $y - \phi$ shape, as shown in Fig. 1.8.

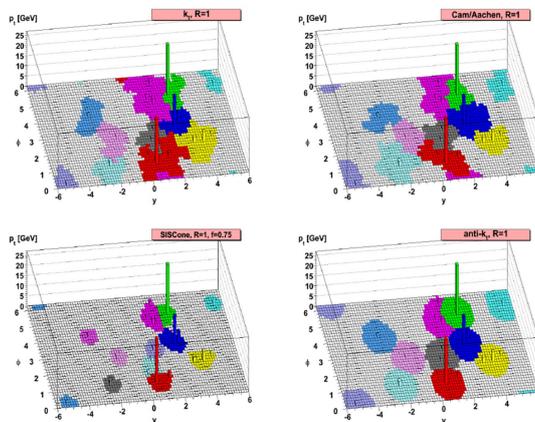


Figure 1.8: Comparison of different jet reconstruction algorithms.

1.11 Partonic Origin of Jets

To determine the partonic origin of jets, in the context of the MC simulation, we use the “Ghost Particles Method”. In this method we make use of the generated partons in the simulation to determine the origin of a jet. We include all generated partons as particles with infinitesimal momentum in the list of particles used in the anti- k_T algorithm. The parton of the highest original momentum that is clustered in a jet defines the jet’s flavor. If no

parton is clustered in a jet, the jet's flavor is undetermined. In this thesis we want to discriminate between gluon or quark origin thus we will name them gluon-jets or quark-jets, in the context of MC. On the other hand, we don't have any parton information from data because of color confinement (free partons are not observed) and, as a result, we can not use the same method for data. This is why we use the Quark Gluon Likelihood (QGL) method [14], which is an assigned probability of a jet to have originated from a quark. To calculate the QGL value for each jet we use certain discriminating variables which have different distributions for quark-jets and gluon-jets. We can create quark-jet and gluon-jet distributions from our MC simulation where we have used the Ghost Particles Method to assign flavor to our jets. Then by comparing the values of a real jet with these distributions we can calculate the QGL. We will expand further on this topic later on.

Chapter 2

Analysis Strategy

2.1 Purpose of the Study

Our goal is to do jet-based physics using quark-jets and gluon-jets. The partonic origin of these jets is not an observable value but, as discussed, it can be determined in the MC simulation and correlated with certain observable values. One of the first measurements that we can do is to determine the gluon-jet fraction for detector-level jet as well as particle-level jets. To determine the detector-level gluon-jet fraction we can take a straight forward approach of using a maximum-likelihood fit on the QGL or discriminating variable distributions. In order to calculate the particle-level gluon-jet fraction we have to unfold the result from detector-level. By unfolding separately the differential cross section for quark-jets and gluon-jets, we calculate the particle-level gluon-jet fraction. After establishing the gluon-jet fraction expected by the SM, we will be able to expand the studies in the future and search for new physics using signatures that include gluon and quark-jets in the final states. Next sections present the discriminating variables distributions for MC and the extracted QGL using these variables.

2.2 Discriminating Variables

As discussed in section 1.11, we use certain discriminating variables in order to calculate the QGL value for jets. There are many such variables but we use a set of three which has been shown to have good performance. These are experimentally measurable and, based on MC, have a different distribution when separating for quark-jets and gluon-jets. First we have the jet particle multiplicity (QGMul). Next we have the negative logarithm of the minor axis of the ellipse defined by the jet (QGAX2) which is calculated as follows.

We make a 2x2 matrix M with the elements being

$$M_{11} = \sum_i p_{T,i}^2 \Delta\eta_i^2 \quad (2.1)$$

$$M_{22} = \sum_i p_{T,i}^2 \Delta\phi_i^2 \quad (2.2)$$

$$M_{12} = M_{21} = -\sum_i p_{T,i}^2 \Delta\eta_i \Delta\phi_i \quad (2.3)$$

which has the eigenvalues $\lambda_{1,2}$ where $\lambda_1 > \lambda_2$. The quantity σ_2 is defined to be

$$\sigma_2 = (\lambda_2 / \sum_i p_{T,i}^2)^{1/2} \quad (2.4)$$

Finally, we have a measure of how evenly the p_T is distributed between jet particles (QGpTD), defined as

$$p_t D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \quad (2.5)$$

If a particle has significantly more p_T than all of the rest combined then QGpTD is close to 1 and if all particle have the same p_T then QGpTD has the value of $1/\sqrt{QGMul}$. In Fig. 2.1 we present an example of the three discriminating variables

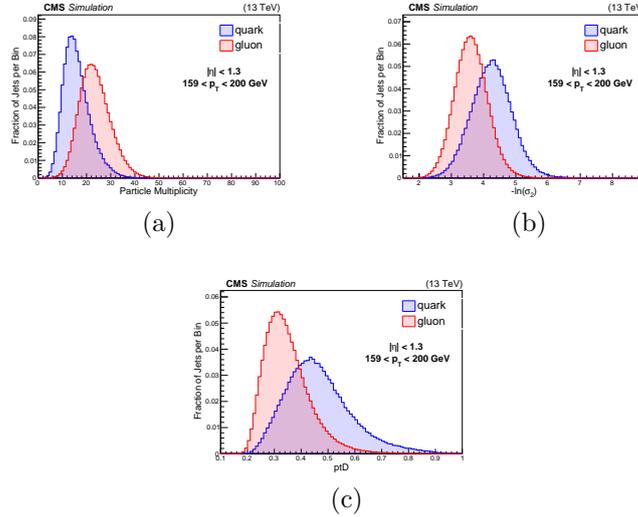


Figure 2.1: The quark-jet (blue) and gluon-jet (red) distributions of particle multiplicity (a), measure of jet width $-\ln(\sigma_2)$ (b), and fragmentation function $p_T D$ (c) for different p_T bins.

It is clear from these plots that there is separation between quark-jet and gluon-jet distributions and thus we can use them to discriminate between

the two kinds of partons. In Fig. 2.2 we present how the separation between these variables changes as a function of p_T .

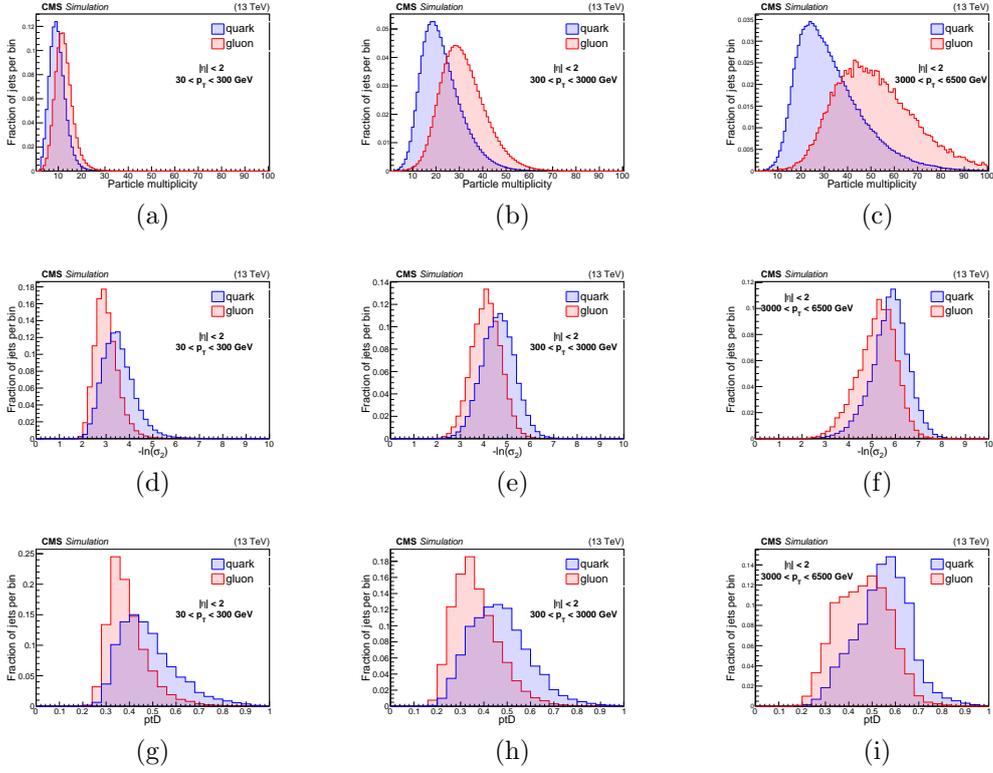


Figure 2.2: The quark-jet (blue) and gluon-jet (red) distributions of particle multiplicity, measure of jet width $-\ln(\sigma_2)$, and fragmentation function $p_T D$ for different p_T bins.

In Fig. 2.3 we present how the separation changes as a function of η .

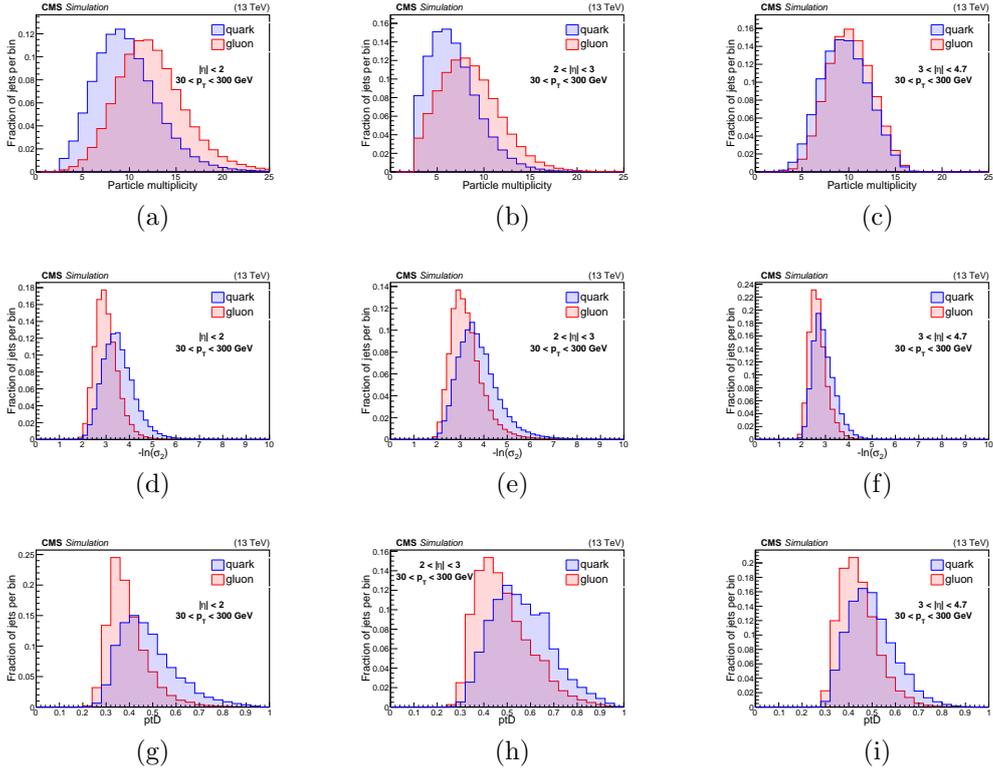


Figure 2.3: The quark-jet (blue) and gluon-jet (red) distributions of particle multiplicity, measure of jet width $-\ln(\sigma_2)$, and fragmentation function $p_T D$ for different η bins.

From all of those plots it is clear that the MC predicts different distributions for these three variables between quark and gluon jets.

2.3 QGL calculation

The QGL [14] is the assigned probability that a jet has originated from a quark. We calculate it by using the probability density plots, denoted as P , of each discriminating variable shown above. The QGL is calculated as follows

$$QGL = P_{quark} / (P_{quark} + P_{gluon}) \quad (2.6)$$

$$P_{quark} = P_{QGMul,quark} \times P_{QGAx2,quark} \times P_{QGpTD,quark} \quad (2.7)$$

$$P_{gluon} = P_{QGMul,gluon} \times P_{QGAx2,gluon} \times P_{QGpTD,gluon} \quad (2.8)$$

Meaning that the probability for the jet to have originated from a quark is the probability density of quark-jets over the sum of the probability densities of quark and gluon jets. In Fig. 2.4 we present the QGL for different p_T and y bins.

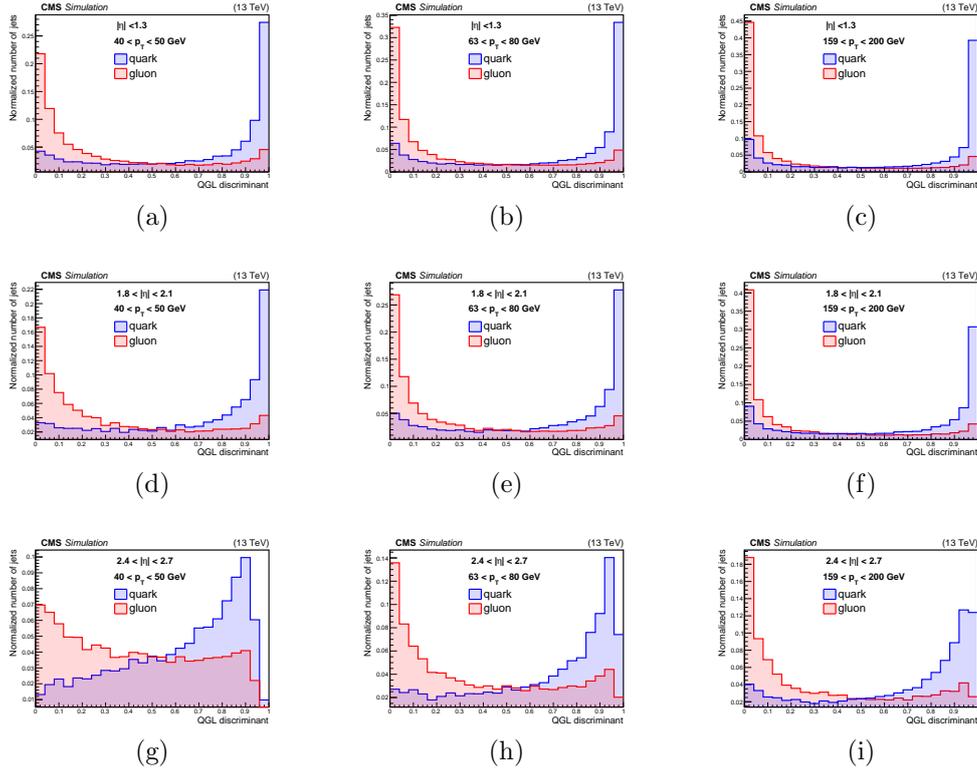


Figure 2.4: QGL discriminant for different p_T and y bins. Discrimination is better for higher p_T and lower η .

A common method to measure the efficiency of discrimination between quark-jets and gluon-jets is the use of the Receiver Operating Characteristic Curve (ROC) and particularly its integral, which gives the Area Under Curve (AUC). We calculate these quantities for different p_T and η bins to get plots similar to Fig. 2.5

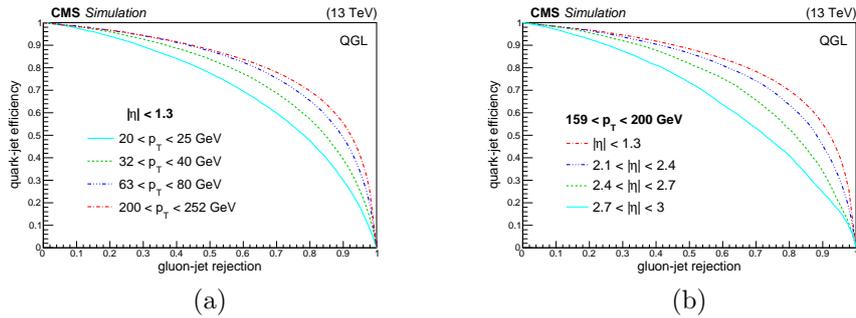
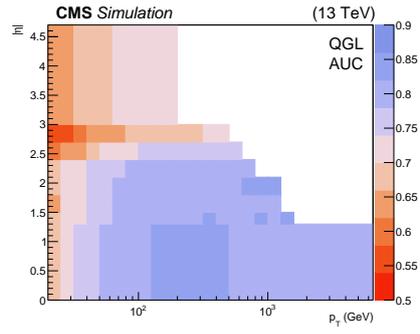


Figure 2.5: The ROC curves for different p_T bins and $|\eta| < 1.3$ or different η bins and $159 < p_T < 200(GeV)$. AUC is larger (discrimination is better) for higher p_T and lower η .

Finally, in Fig. 2.6 we present a 2D plot of the AUC for a grid of p_T and $|\eta|$ bins. We observe that we are more efficient at discriminating quark-jets from gluon-jets at low $|\eta|$ and high p_T of the jets.



(a)

Figure 2.6: AUC of ROC curves as a function of p_T and $|\eta|$. Larger AUC corresponds to better discrimination between gluon-jets and quark-jets.

Chapter 3

Data and Simulation samples

3.1 Data Samples and Requirements

In this section we present the data used for our studies. They are the so-called 2016 Ultra Legacy (UL) data, collected by the CMS experiment in 2016 and corresponding to integrated luminosity of 36.31 fb^{-1} . This dataset is composed of different eras, each era having different luminosity and different settings of the detector denoted by the abbreviations APV and nonAPV. We present this in Table 3.1. Because we are interested in jet production, we are using data collected with single-jet triggers described in section 3.1.1. We analyze jets that pass the jet identification requirements presented in section 3.1.2. Corrections and cleaning of the data are described in sections 3.1.3, 3.1.4, 3.1.5.

Era	Luminosity (pb^{-1})	Settings
B	5825.7	APV
C	2601.6	APV
D	4286.0	APV
E	4065.9	APV
Fe	2733.6	APV
F1	403.5	nonAPV
G	7653.2	nonAPV
H	8740.1	nonAPV

Table 3.1: Different Eras with luminosity and settings of the UL16 dataset.

3.1.1 Triggers

Triggers are essential pieces of equipment enabling the experimentalists to automatically decide which events to record for further analysis. Given the immense amount of data produced by particle collisions we would be unable to store them all and thus we have to decide in real time which events to keep. Triggers operate at multiple levels, subsequently stricter. The first level (L1) trigger, which is hardware-based, rapidly assesses basic event characteristics,

like energy deposits, to make real time decisions, significantly reducing the data flow to a rate of around 100 kHz by discarding uninteresting events. The High Level Trigger (HLT), which is software-based, performs a thorougher analysis on the events that passed the L1 trigger, further filtering events to achieve a rate of about 1 kHz, suitable for data storage. These triggers store data in buffers to provide time to analyze the incoming data and some of them are active only a fraction of the time ("prescaled" triggers). Different triggers fire at a preset energy deposition limit and for different limits the fraction of active time is different. This feature of our hardware gives us another problem to solve: how to correctly weight each event to counteract the trigger effects. This will be discussed in the analysis chapter. All in all, these triggers ensure that only the most promising events are recorded and stored for in depth analysis, thus optimizing the use of storage and computational resources while enabling the discovery of rare events. In Table 3.2 we show the used triggers with their respective effective luminosity and turn-on p_T which is the point at which they have reached almost 100% efficiency.

Trigger	Effective Luminosity (pb^{-1})	Turn-on p_T (GeV)
HLT_PFJet40	0.266973	64
HLT_PFJet60	0.726088	84
HLT_PFJet80	2.75846	114
HLT_PFJet140	24.1886	174
HLT_PFJet200	103.798	245
HLT_PFJet260	593.634	300
HLT_PFJet320	1772.11	362
HLT_PFJet400	5193.14	468
HLT_PFJet450	36310	507

Table 3.2: Effective luminosity for different jet triggers.

3.1.2 Jet Identification Requirements

There are three versions of Jet Identification (ID), each been progressively more strict: Loose, Tight, and TightLepVeto. These IDs are designed to balance between signal efficiency and background rejection. The Loose ID is the least strict, allowing a broader selection of jets with lower requirements, making it useful for preliminary analysis stages or when high signal efficiency is needed, even at the cost of higher background levels. The Tight ID is more strict, selecting higher quality jets with better isolation and identification, reducing background contamination. The TightLepVeto ID combines the strict criteria of the Tight ID with additional requirements to veto events with high lepton content, focusing on analyses where excluding leptons is necessary to reduce background from certain processes. For our analysis we are using the TightLepVeto cut (the Tight cut is applied to regions where TightLepVeto is not defined). In Table 3.3 we show each ID definition for different η ranges.

$-2.7 \leq \eta \leq 2.7$			
Jet ID	Loose	Tight	TightLepVeto
Neutral Hadron Fraction	< 0.99	< 0.90	< 0.90
Neutral EM Fraction	< 0.99	< 0.90	< 0.90
Number of Constituents	> 1	> 1	> 1
Muon Fraction	-	-	< 0.8
$-2.4 \leq \eta \leq 2.4$			
Charged Hadron Fraction	> 0	> 0	> 0
Charged Multiplicity	> 0	> 0	> 0
Charged EM Fraction	< 0.99	< 0.99	< 0.90
$2.7 < \eta \leq 3.0$			
Jet ID	Loose	Tight	
Neutral EM Fraction	> 0.01	> 0.01	
Neutral Hadron Fraction	< 0.98	< 0.98	
Number of Neutral Particles	> 2	> 2	
$ \eta > 3.0$			
Jet ID	Loose	Tight	
Neutral EM Fraction	< 0.90	< 0.90	
Number of Neutral Particles	> 10	> 10	

Table 3.3: Jet ID criteria for different η ranges.

3.1.3 MET filters

Missing Transverse Energy (MET) is the imbalance of transverse momentum in an event. This imbalance occurs because of undetectable produced particles, such as neutrinos, or from cosmic rays, mismeasured energy deposits or detector noise. Filters on MET are used to enhance the quality of experimental data by identifying and excluding events with outlying MET values. Even though we are not doing a MET-related study MET filters are mandatory for assuring the quality of data and reducing background noise. The used cut-off to accept an event is:

$$MET/Sum_E_T < 0.3 \quad (3.1)$$

with Sum_ET being the sum of the transverse energy.

3.1.4 Prefire weights

Prefire weights are correction factors applied to data to account for the pre-firing of detector components. This can occur when the detector records a signal from a particle slightly before the event of interest, due to timing mismatches or early triggering by unrelated particles. Prefiring can lead to the loss of energy measurements and affect the reconstructed event characteristics. Prefire weights are calculated based on detailed studies of the detector's performance. This ensures more accurate energy measurements and reduces systematic uncertainties leading to more precise results. In Fig. 3.1 we have the two prefire maps used which represent the prefire frequency for two different eras of UL16 data.

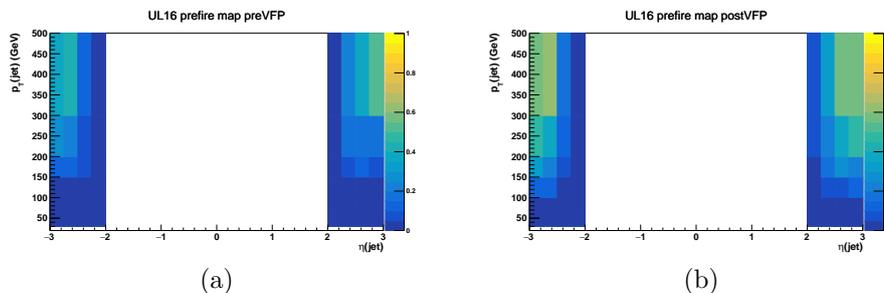


Figure 3.1: Two different prefire maps for different eras of UL16, (a) is preVFP or APV (b) is postVFP or nonAPV.

The prefire weight is calculated by the product of the inverse of the probability of each jet of the event not to have prefired:

$$W = \prod_i^{Jets} \frac{1}{1 - c_i} \quad (3.2)$$

Where c_i is the prefire probability stored in the prefire maps.

3.1.5 Hot/Cold-channels

Hot/Cold-channels refer to specific regions of the detector that produce abnormally high or low rates of signals. This can occur due to electronic noise, cross-talk between channels or even hardware malfunctions thus the corresponding signals are unrelated to the actual events we want to study. These channels can degrade data quality by introducing false signals and increasing background noise. To mitigate these effects, hot-channels are identified through calibration and monitoring. Our study does not focus on correcting signals from these regions and we rather exclude these data during our analysis. In Fig. 3.2 we have the two hot-channel maps used one from data and one from MC. We apply both data and MC maps on both data and MC data sets.

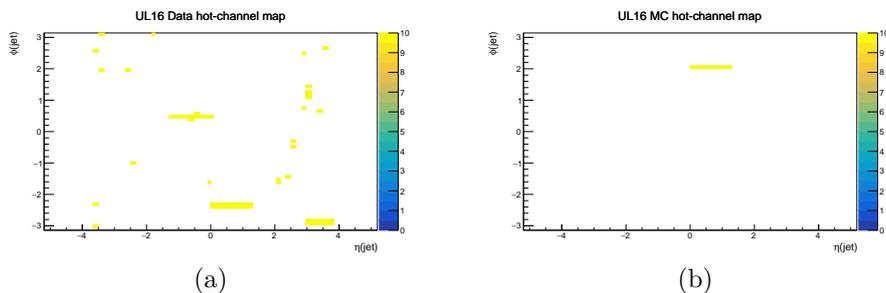


Figure 3.2: Hot/Cold-channels from data (a) and MC (b).

3.2 Merging of data

The use of triggers introduces effects that we have to account for. We have to figure out what is the correct weight for each event in order to counteract the effect of the trigger. We know that each trigger has a preset p_T limit but its effectiveness at correctly identifying a jet of that p_T is not 100% for jets that exceed that p_T . As a result, we have a higher minimum p_T requirement (Table 3.2) where the effectiveness is reaching 100% and so we can define the fully efficient window of each trigger as being the transverse momentum region where it has reached almost 100% efficiency up until the next trigger reached full efficiency. Having now defined the fully efficient window of each trigger there are two methods for accounting for the trigger effects and merge our data: the Prescales method and the Effective Luminosity method.

3.2.1 Use of Prescales to Merge the Data

In this method, we construct the p_T spectrum by weighting each jet in an event by a prescales weight. This weight accounts for the fact that the trigger that recorded this event was not always active. The prescale value used is that of a trigger that fired during the event, with the requirement that the leading jet in the event falls into its full-efficiency p_T window. Note that the highest- p_T trigger (HLT PFJet450) is not prescaled and always active to

record events. In Fig. 3.3 we present each individual part of the spectrum for each trigger as was defined for different rapidity bins, as well as the total spectrum resulted from the merging.

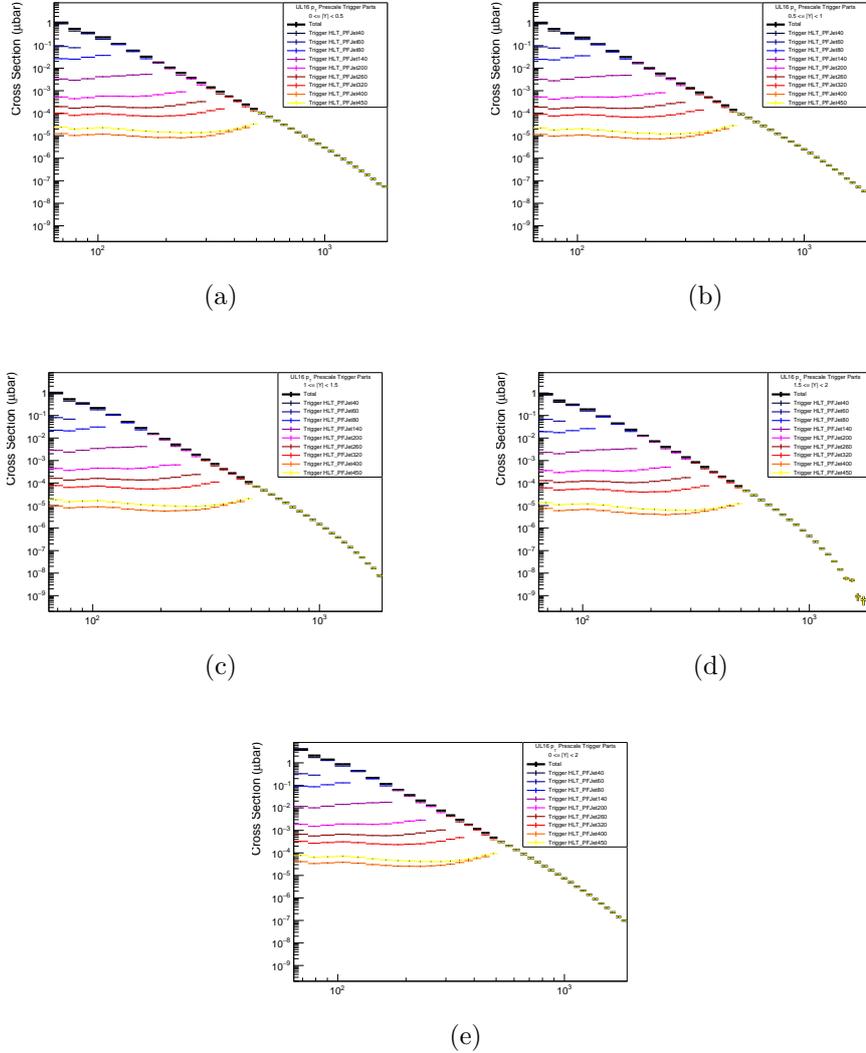


Figure 3.3: Individual parts of p_T spectrum for each trigger and the total spectrum for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

3.2.2 Use of Effective Luminosities to Merge the Data

In this method we treat the data collected by each trigger as an independent experiment. For each such experiment, we create a separate jet p_T spectrum. We subsequently build the entire p_T spectrum by concatenating only the fully efficient window of each trigger spectrum. By keeping only the fully efficient windows, which by definition do not overlap, we ensure that there is no double-counting. The way to weight the sub-spectra that correspond to these windows is by introducing the effective luminosity (Table 3.2). The

effective luminosity is calculated for each trigger and accounts for the fact that the trigger was not always active during data taking. In other words, it incorporates the prescales into it. For the highest- p_T trigger, which is not prescaled, the effective luminosity is the same as the total luminosity of the 2016 data taking. In Fig. 3.4 we show for all rapidity bins the spectrum for each trigger as well as the merged spectrum which is the result of concatenation of the fully efficient window of each trigger.

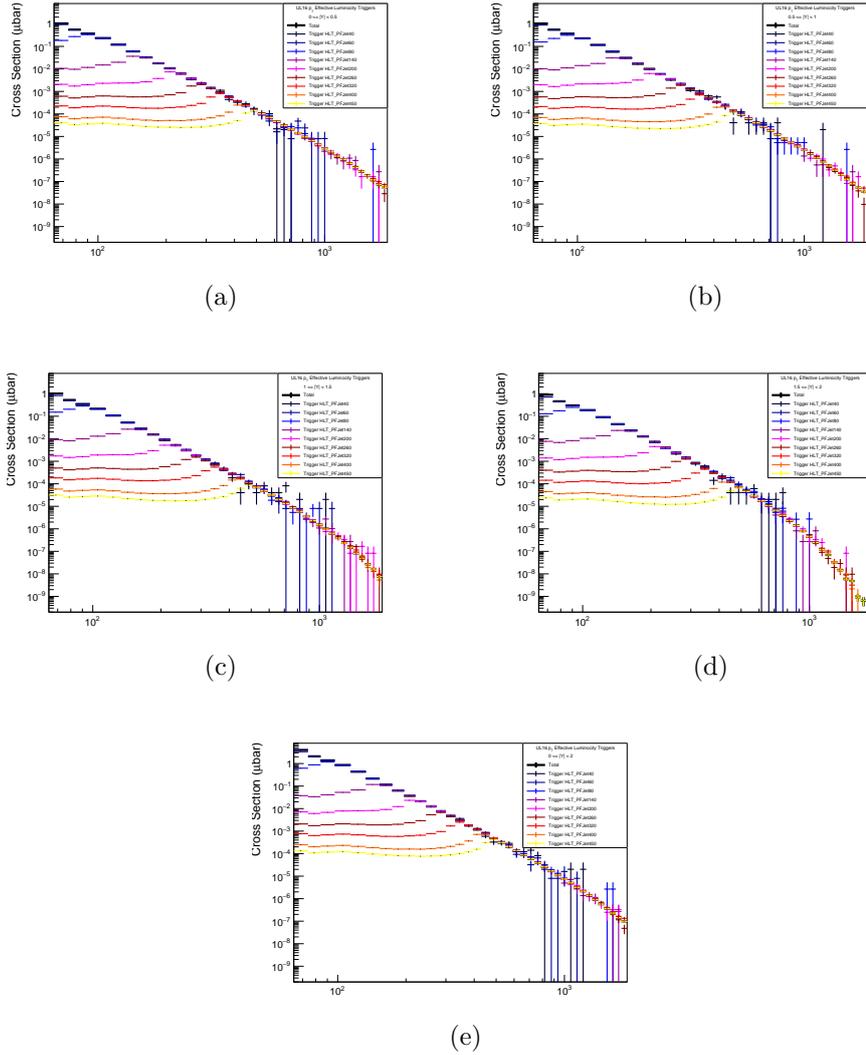


Figure 3.4: Trigger p_T spectra as if they were independent and the total spectrum for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

3.2.3 Comparison

Each method has each own advantages and disadvantages. For example the prescales method keeps about twice as many jets but since the effective luminosity method has more similar weights between different jets (on each

trigger) it has about 5% smaller statistical uncertainty. Moreover by using the prescales method if we keep an event we keep it in its entirety and thus we can define event wide properties such as for example an event been dijet. We present here the results for the p_T spectra for data for different rapidity bins with the two different methods applied Fig. 3.5.

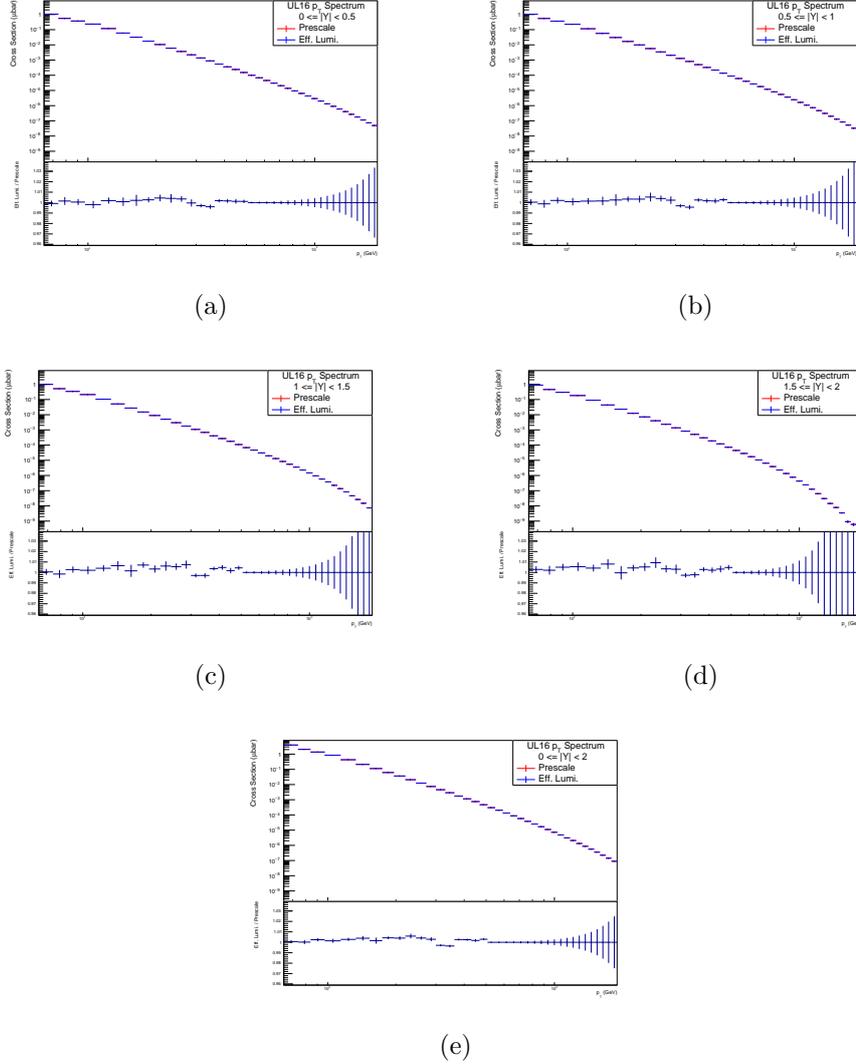


Figure 3.5: Prescale and effective luminosity p_T spectra (upper panel) and the ratio between them (bottom panel) for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

We can see that there is an agreement between these two methods. We observe that for $p_T > 507$ GeV, the two methods give identical results, as expected, because the data there are collected by a non-prescaled trigger. From this point onward we will keep using the effective luminosity method, for correcting for the trigger effects, except if specified otherwise.

3.3 MC Selection and Requirements

In this work the SM predictions come primarily from Madgraph+Pythia and secondarily from Madgraph+Herwig UL16 samples. Similarly to data selection, MC event selection is also very important for improving the quality of the dataset. First of all, we apply MET filters, hot-channels and Jet ID cut in the same way as we apply them for data. Prefire weights and trigger effects do not apply to MC, even if we could simulate these effects they are not relevant to our study. We utilize H_T -binned samples as presented in the next section.

3.3.1 H_T -bins

H_T refers to the scalar sum of the transverse momentum p_T of all outgoing particles in an event. It is often the case that we produce MC simulation events in different bins of H_T , in order to improve the statistics in higher H_T values. This is necessary because the production cross section reduces several orders of magnitude as a function of outgoing particle p_T . If we had used a single MC sample with all possible H_T values ("flat sample") it would have been very unlikely to produce events with jets with high transverse momentum thus making it impossible to achieve "good" statistics at high jet p_T . Instead, we separate event generation based on H_T -bins with each one having a reasonable span of cross-section and then merging them by using the luminosity of each bin, which is calculated as number of events generated in that bin over the cross-section in that bin. The cross-section is calculated by MC integration over the phase-space of possibilities. This H_T binning of course only applies to MC. In Table 3.4 we have each H_T -bin used with their respective luminosity weight. The merging of APV and nonAPV MC is done by weighting each one by the ratio of APV or nonAPV luminosity to the total luminosity, as determined in the data.

H_T -bin (GeV)	APV Luminosity-Based Weight (pb)	nonAPV Luminosity-Based Weight (pb)
50 - 100	4.87653	4.7819
100 - 200	0.326296	0.313335
200 - 300	0.030827	0.0376868
300 - 500	0.00624473	0.00624755
500 - 700	0.000639486	0.000590487
700 - 1000	0.000167105	0.000158555
1000 - 1500	7.9396e-05	8.8306e-05
1500 - 2000	1.40622e-05	1.27823e-05
> 2000	1.63078e-05	5.01464e-06

Table 3.4: H_T -bins used with their respective luminosity-based weight in (pb) for APV and nonAPV settings of the detector. This luminosity weight doesn't include the relative APV-nonAPV merging weight, which reflects the ratio APV/nonAPV in the data.

3.4 Data-MC comparison

In Fig. 3.6 we present the comparison between the data p_T spectra and the Pythia MC p_T spectra for the same rapidity bins. For MC there is no need to account for any trigger effects but the H_T -bin samples are merged as discussed in Section 3.3.1.

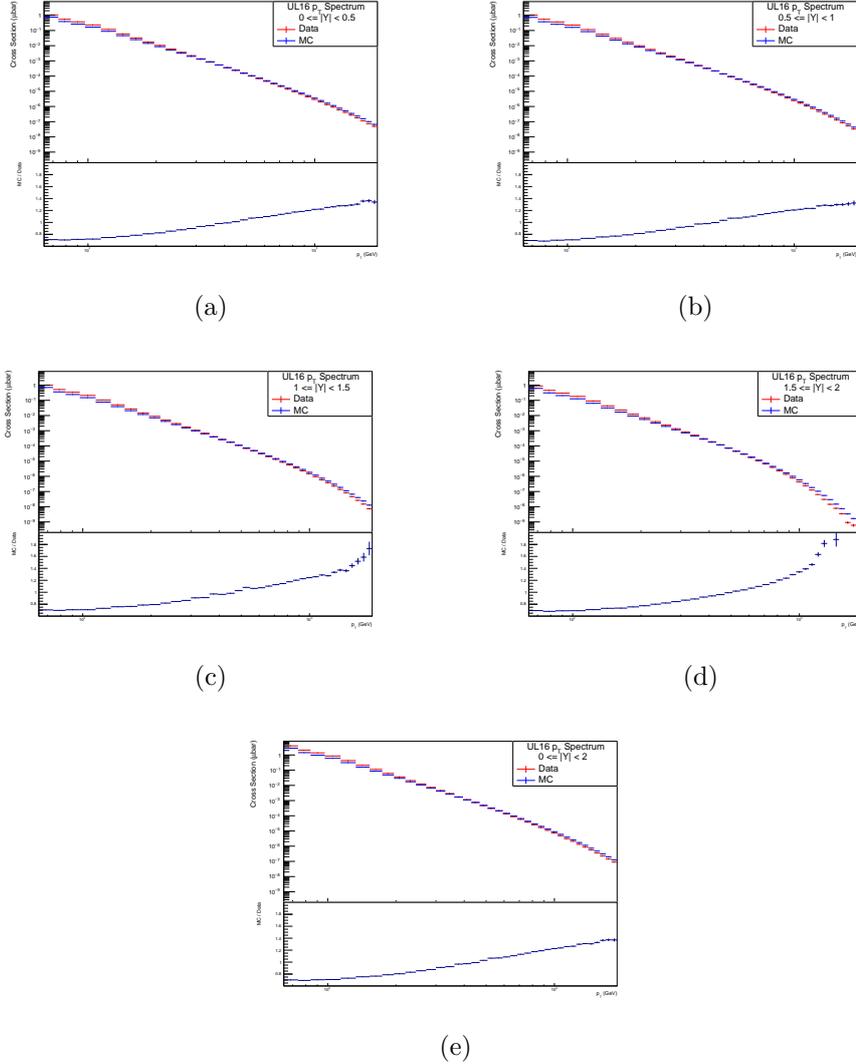


Figure 3.6: Comparison of Data and Pythia MC p_T spectra (upper panel) and the ratio between them (bottom panel) for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

We notice the disagreement between the measured detector-level jet spectrum and the predicted one by the MC. The reason is that our MC is a leading order (LO) generator and it doesn't fully describe the QCD effects that we observe in the data. The proper comparison is that of the unfolded experimental spectra (e.g., with the detector effects removed) and the spectra predicted by the perturbative QCD at next leading order (NLO) or next-to-NLO.

Chapter 4

Double-Differential Inclusive Jet Cross Section Measurement

In order to measure the particle-level unfolded gluon-jet fraction we need to unfold the jet p_T spectra for quark-jets and gluon-jets separately. Before doing it separately we would like to unfold the non-parton-specific inclusive jet cross section and compare it to the published one by CMS [12]. This comparison will confirm the validity of our data set and the robustness of our analysis tools.

4.1 Unfolding

To unfold our spectra we are using the TUnfold [15] method, which implements a least-square minimization of the following:

$$\chi^2 = (\mathbf{Ax} - \mathbf{y})^T \mathbf{V}_{yy}^{-1} (\mathbf{Ax} - \mathbf{y}) \quad (4.1)$$

where \mathbf{A} is the response matrix, \mathbf{x} is the resulting particle-level spectrum, \mathbf{y} is the detector-level spectrum, and \mathbf{V}_{yy} is the covariance matrix. The response matrix that describes how detector-level jets correspond to particle-level jets in the simulation and a covariance matrix which captures the covariance of the different bins of the data spectrum. We are going to perform the unfolding simultaneously in p_T and y (2D unfolding). We have seen the data spectrum and how the simulation compares to it in the previous chapter so we will continue with the response matrix. In Fig. 4.1 we present the 2D response matrix for unfolding from detector-level to particle-level jets.

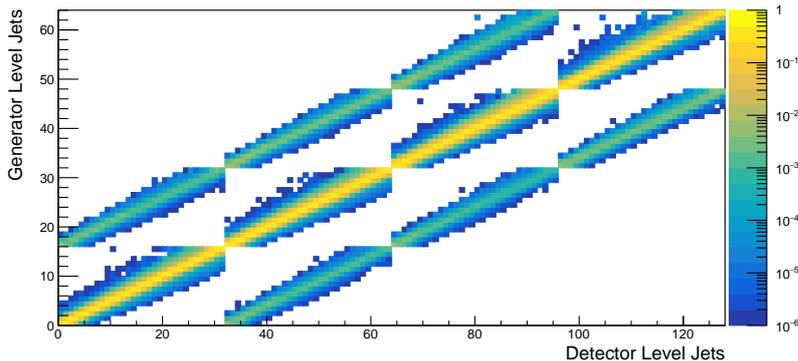


Figure 4.1: Response Matrix.

The bins of the response matrix do not show p_T or y values, as in a typical spectrum, but rather the index of the matrix. In Table 4.1 we present the p_T and y bins used. The particle-level p_T bins are larger as a form of implicit regularization (smoothing) of the unfolded result.

Detector-Level p_T bins (GeV)	64, 74, 84, 97, 114, 133, 153, 174, 196, 220, 245, 272, 300, 330, 362, 395, 430, 468, 507, 548, 592, 638, 686, 737, 790, 846, 905, 967, 1032, 1101, 1172, 1248, 1327
Particle-Level p_T bins (GeV)	64, 84, 114, 153, 196, 245, 300, 362, 430, 507, 592, 686, 790, 905, 1032, 1172, 1327
Absolute Rapidity bins $ y $	0.0, 0.5, 1.0, 1.5, 2.0

Table 4.1: p_T and y bins used for detector-level and particle-level bins, y bins are the same for both levels.

The response matrix is composed of 4 by 4 sub-matrices where each sub-matrix corresponds to a y bin. The detector-level is presented in the x-axis and the particle-level in the y-axis. Inside each sub-matrix the p_T bins are represented by the index. As a result, the diagonal sub-matrices correspond to jets that enter the same detector-level and particle-level y bin whereas the off-diagonal sub-matrices correspond to jets with a detector/particle-level y -bin mismatch. Because the generator-to-detector-level jet-matching algorithm requires that the difference in pseudorapidity should be less than 0.2, and because the pseudorapidity is highly correlated to the rapidity, there is no y migration further than one bin and, as a result, the second to off-diagonal sub-matrices have no content. This algorithm also has a limit on the allowed p_T migration but we ignore it because the response matrix is capturing these p_T migrations. Finally, the response matrix is normalized for each row which means that it gives the probability of the origin of each particle-level bin to different detector-level bins. Similarly to the response matrix the covariance matrix in Fig. 4.2 follows the same structure.

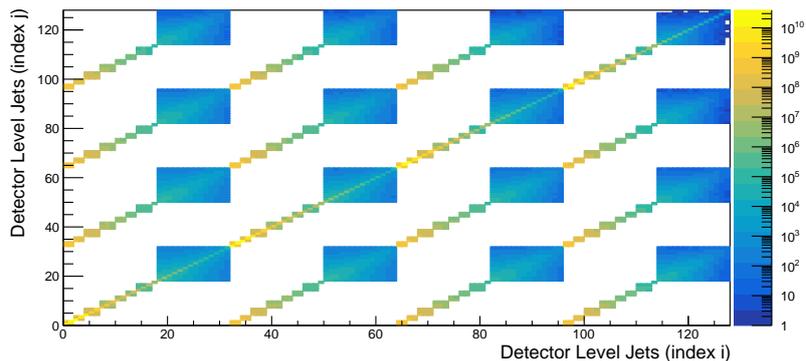


Figure 4.2: Covariance Matrix.

The first difference is that it tracks between detector-level bins i and j where i and j are the possible indices of the spectrum. The second being that all sub-matrices have values since we don't track between generator and corresponding detector-level jets but rather between different jets of the same event which can have any difference in y . And the last difference is the block diagonal form of each sub-matrix which correspond to each trigger's fully efficient window.

4.2 Result

In Fig. 4.3 we show the comparison between our unfolded double differential inclusive cross section measurement and the published one. Our spectra do not reach very high p_T values, due to statistics, because we analyze 2016 data, but also include two bins at lower transverse momenta, not published before.

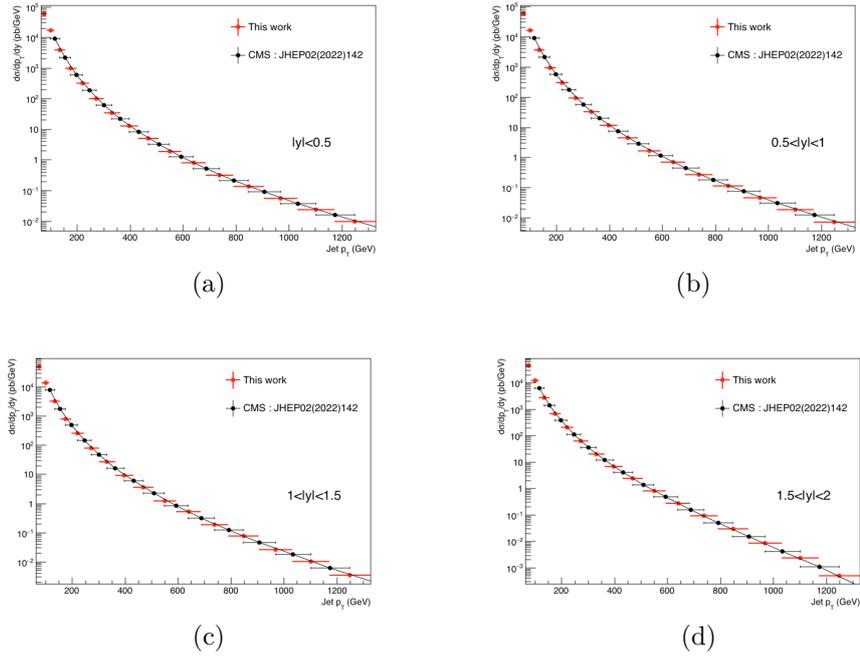


Figure 4.3: Comparison of our differential inclusive cross section to publicized for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2.

Even though we are using different p_T binning we can see the we have a good agreement, meaning that our method works well for unfolding the p_T spectra from detector to particle-level. This comparison also demonstrates that our event selection, data corrections and trigger-based sample merging were properly done.

Chapter 5

Gluon-Jet Fraction Measurement

5.1 Discriminating Variables

For the measurement of the gluon-jet fraction we are not going to use the QGL as done in the past [16] but instead we will use the discriminating variables directly. In Fig. 5.1 we show the comparison between the quark-jet and gluon-jet normalized distributions and that of the data for the three discriminating variables for different p_T bins, where all three distributions are normalized to unity. In Fig. 5.2 we see these distributions as a function of y . We notice that the data lies between the quark and gluon distributions. For that reason, we can extract the gluon jet fraction by maximizing a likelihood for all p_T and y bins.

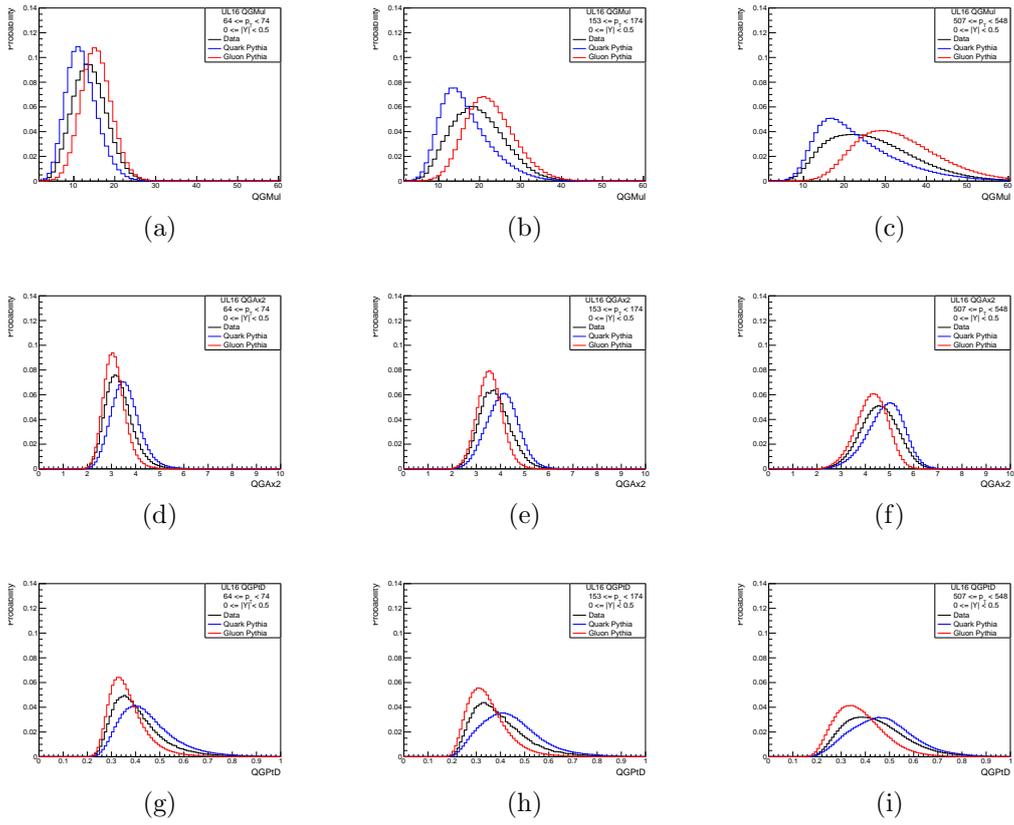


Figure 5.1: Discriminating variable for data and MC for $0 \leq |y| < 0.5$ as a function of p_T , (a) QGMul for $64 \leq p_T < 74$, (b) QGMul for $153 \leq p_T < 174$, (c) QGMul for $507 \leq p_T < 548$, (d) QGAx2 for $64 \leq p_T < 74$, (e) QGAx2 for $153 \leq p_T < 174$, (f) QGAx2 for $507 \leq p_T < 548$, (g) QGPtD for $64 \leq p_T < 74$, (h) QGPtD for $153 \leq p_T < 174$, (i) QGPtD for $507 \leq p_T < 548$.

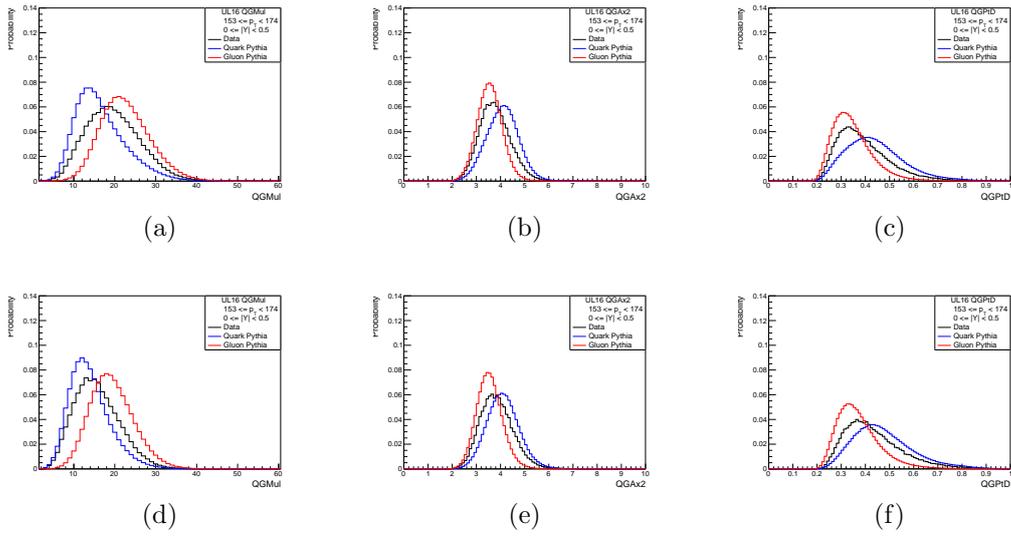


Figure 5.2: Discriminating variable for data and MC for $153 \leq p_T < 174$ as a function of rapidity, (a) QGMul for $0 \leq |y| < 0.5$, (b) QGAx2 for $0 \leq |y| < 0.5$, (c) QGPtD for $0 \leq |y| < 0.5$, (d) QGMul for $1.5 \leq |y| < 2$, (e) QGAx2 for $1.5 \leq |y| < 2$, (f) QGPtD for $1.5 \leq |y| < 2$.

In Fig. 5.3 we show a comparison between Herwig and Pythia MC

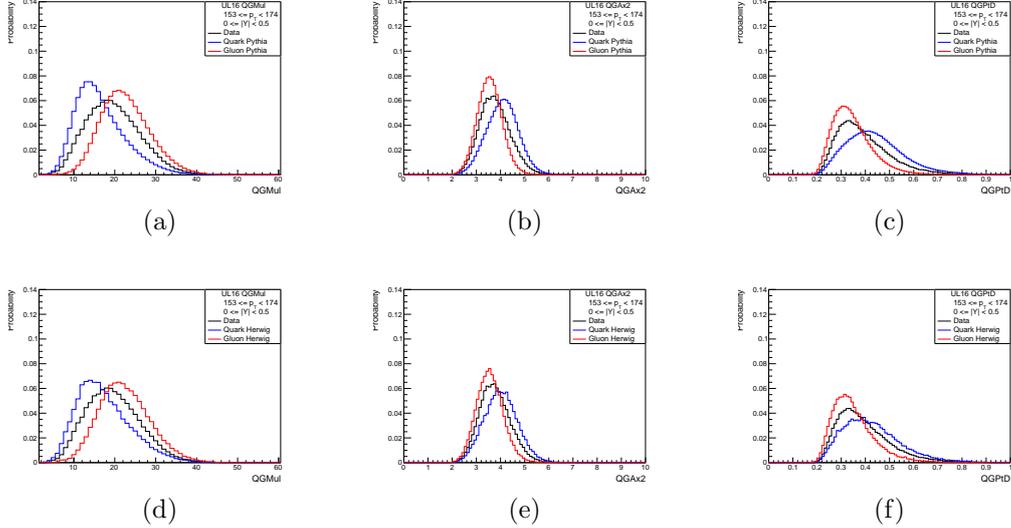


Figure 5.3: Comparison between Herwig and Pythia for $0 \leq |y| < 0.5$ and $153 \leq p_T < 174$, (a) QGMul for Pythia, (b) QGAx2 for Pythia, (c) for QGPtD Pythia, (d) QGMul for Herwig, (e) QGAx2 for Herwig, (f) for QGPtD Herwig.

5.2 Gluon-Jet Fraction Measurement

To make the measurement we perform a maximum likelihood fit of the three discriminating variable histograms, from data, with a combination of the quark-jet and gluon-jet distributions from MC. This can be written as:

$$H_{QGMul,data} = g * H_{QGMul,gluon-jets} + (1 - g) * H_{QGMul,quark-jets} \quad (5.1)$$

$$H_{QGAx2,data} = g * H_{QGAx2,gluon-jets} + (1 - g) * H_{QGAx2,quark-jets} \quad (5.2)$$

$$H_{QGpTD,data} = g * H_{QGpTD,gluon-jets} + (1 - g) * H_{QGpTD,quark-jets} \quad (5.3)$$

where g is the gluon-jet fraction of data. We separate jets based on p_T and y bins and for each bin we maximize the following likelihood.

$$L = \sum -\ln(\sqrt{2\pi}\sigma_i) - \frac{1}{2} \frac{(H_{data,i} - H_{MC,i})^2}{\sigma_i^2} \quad (5.4)$$

$$H_{MC,i} = g * H_{gluon-jets,i} + (1 - g) * H_{quark-jets,i} \quad (5.5)$$

Here the index i runs through all three of the discriminating variables bins. From the maximization of this likelihood we extract the detector-level gluon-jet fraction for a single p_T and y bin. We concatenate p_T bins to extract the entire gluon-jet fraction spectrum as a function of p_T for different y bins. In Fig. 5.4 we have used the two different MC templates, Pythia and Herwig, to extract the experimental gluon-jet fraction but also using the average of the two. The theoretical value is produced by weighting gluon-jets and quark-jets separately and taking the fraction of gluon-jets to the total.

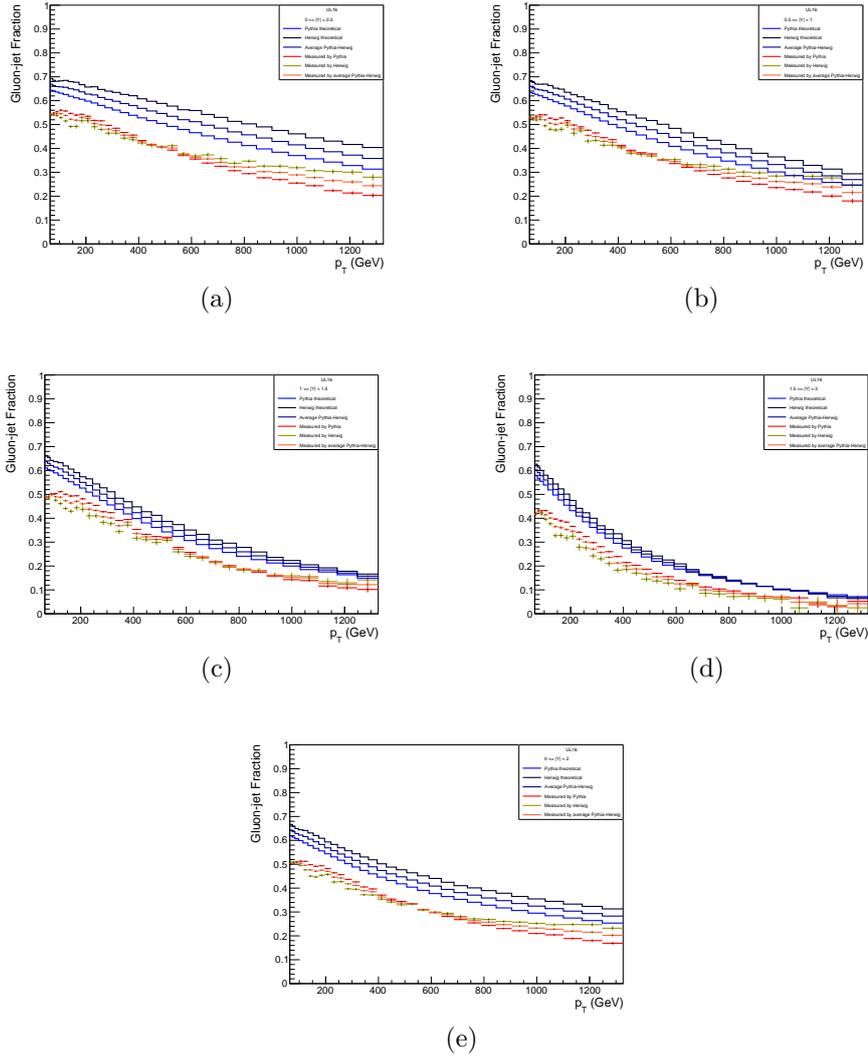


Figure 5.4: Detector-level gluon-jet fraction for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2 (red using Pythia templates, green using Herwig templates, orange using Pythia/Herwig average templates) as a function of p_T compared to the MC prediction (blue Pythia, black Herwig, dark blue average) predicted one. The band around average corresponds to model systematic error and the band around the MC corresponds to the difference between Pythia and Herwig.

This result shows a discrepancy between data result and MC prediction. The biggest factor for this is the fact that the MC does not fully describe the kinematics and the number of gluon-jets, compared to the data.

5.3 Unfolding of Gluon-Jet Fraction

Since we have demonstrated our ability to unfold a p_T spectrum we can apply it to calculate the particle-level gluon-jet fraction. We take the data p_T spectrum and multiply in by the measured gluon-jet fraction to get a

gluon-jet p_T spectrum and multiply in by one minus the gluon fraction to get a quark-jet p_T spectrum. Then, we can create two response matrices, one using gluon-jets and one using quark-jets. Here we are ignoring flavor migration between particle-level and detector-level in our simulation because it is very rare. In Fig. 5.5 we have the two different response matrices for unfolding gluon-jets or quark-jets.

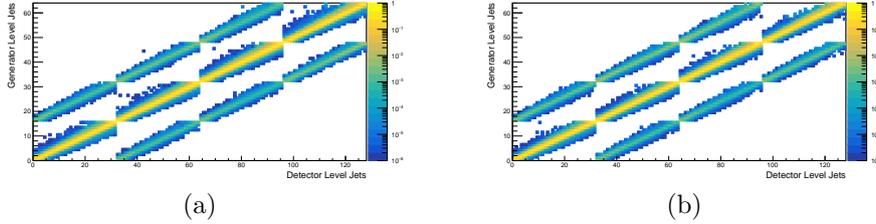


Figure 5.5: (a): Quark-jet response matrix, (b): Gluon-jet response matrix

This side by side comparison doesn't indicate the differences between the two matrices. This is why we are going to define two different matrices just for a side by side comparison. One is quark-jet minus gluon-jet truncated at zero and the other gluon-jet minus quark-jet truncated at zero (if negative it is set at 0). The first we are going to call quark advantage and the other gluon advantage. In Fig. 5.6 we have this comparison.

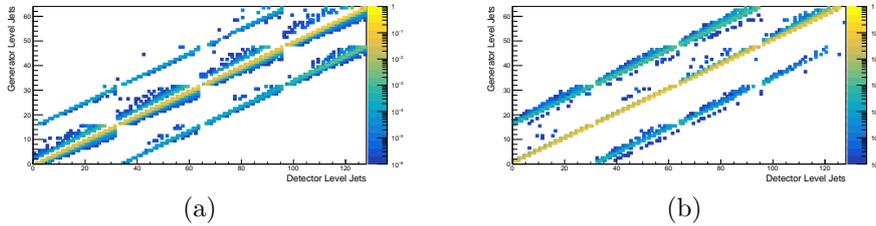


Figure 5.6: (a): Quark-jet advantage, (b): Gluon-jet advantage

Now it is clear that there is a difference in the migration patterns of quark-jets and gluon-jets. When there is no y bin migration quark-jets seem to migrate further than gluon-jets in terms of p_T , as demonstrate by the higher probability in the off diagonal regions of the diagonal sub-matrices. Moreover we notice that the pattern favors a higher detector-level p_T more than a lower one, compared to the particle-level p_T . The migration pattern changes when there is y bin migration with gluon-jets having a higher probability than quark-jets of migrating at lower detector-level p_T compared to their particle-level p_T .

Regarding the covariance matrix for unfolding gluon-jet we are going to multiply it by $g_i * g_j$ where i and j are the bin indices and for quark-jet unfolding we multiply by $(1 - g_i) * (1 - g_j)$ where $g_{i,j}$ are the experimentally extracted detector-level gluon-jet fraction. Finally to get the particle-level

gluon-jet fraction we divide the unfolded gluon-jet spectrum by the sum of the unfolded gluon-jet and quark-jet spectra.

5.4 Systematic Uncertainties

To calculate the systematic uncertainties we only focus on jet energy correction systematic (JEC), jet energy resolution systematic (JER) and method systematic. This is because we are calculating a ratio and there are cancellations between the numerator and the denominator. First the JEC systematic is produced by making use of the JEC uncertainty in data. We repeat the measurement three times, one for the nominal value of the JEC, one for the +1 uncertainty and one for the -1 uncertainty. The systematic error, in general, is asymmetric and is defined by the band around the nominal result and the ± 1 uncertainty results. This band can be expressed as a percent fraction of the nominal value and combined as independent errors to the rest of the systematic and statistical errors. Similarly the JER systematic is produced by making use of the JER uncertainty in MC. The method of calculating it is the same as for the JEC systematic. The method systematic is defined by the band around the measurement using average Pythia/Herwig templates and the measurements using Pythia or Herwig templates separately. Similarly to the rest of the systematic uncertainties, it is expressed as a percent fraction of the average value. In Fig. 5.7 we present the systematic uncertainty sources and the total uncertainty as percentages of the unfolded cross-section, for all p_T and y bins.

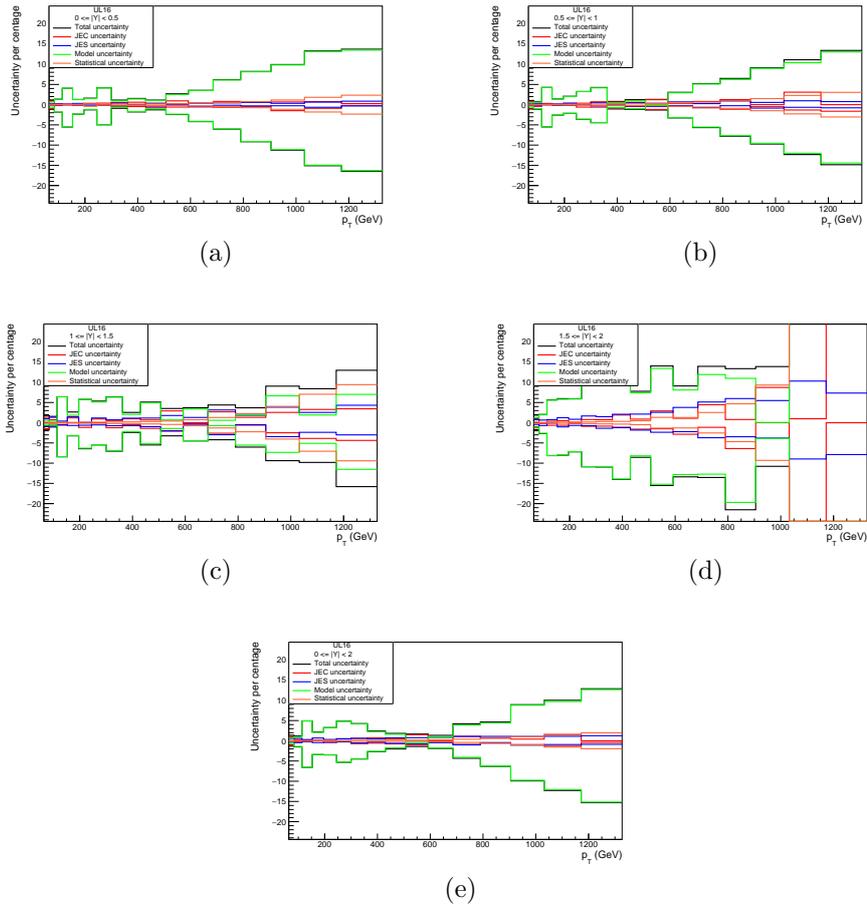
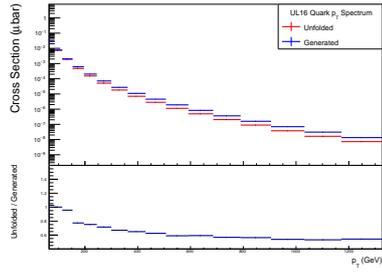


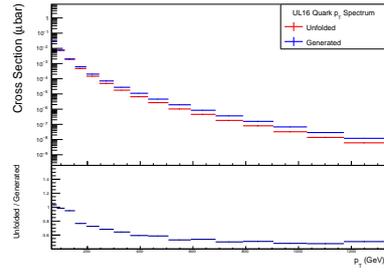
Figure 5.7: Different sources of uncertainty as a percentage for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

5.5 Results

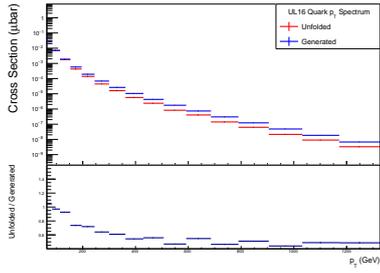
The measured and MC-predicted particle-level spectra for quark-jets and gluon-jets are shown in Fig. 5.8 and Fig. 5.9 respectively.



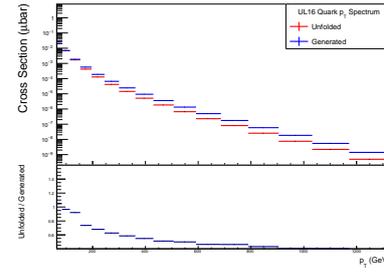
(a)



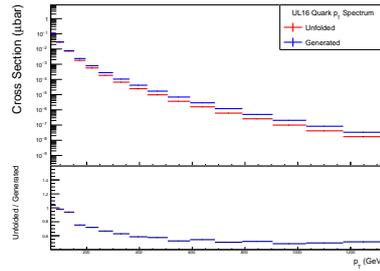
(b)



(c)



(d)



(e)

Figure 5.8: Measured and MC-predicted particle-level spectra (upper) and their ratio (bottom) for quark-jets and for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

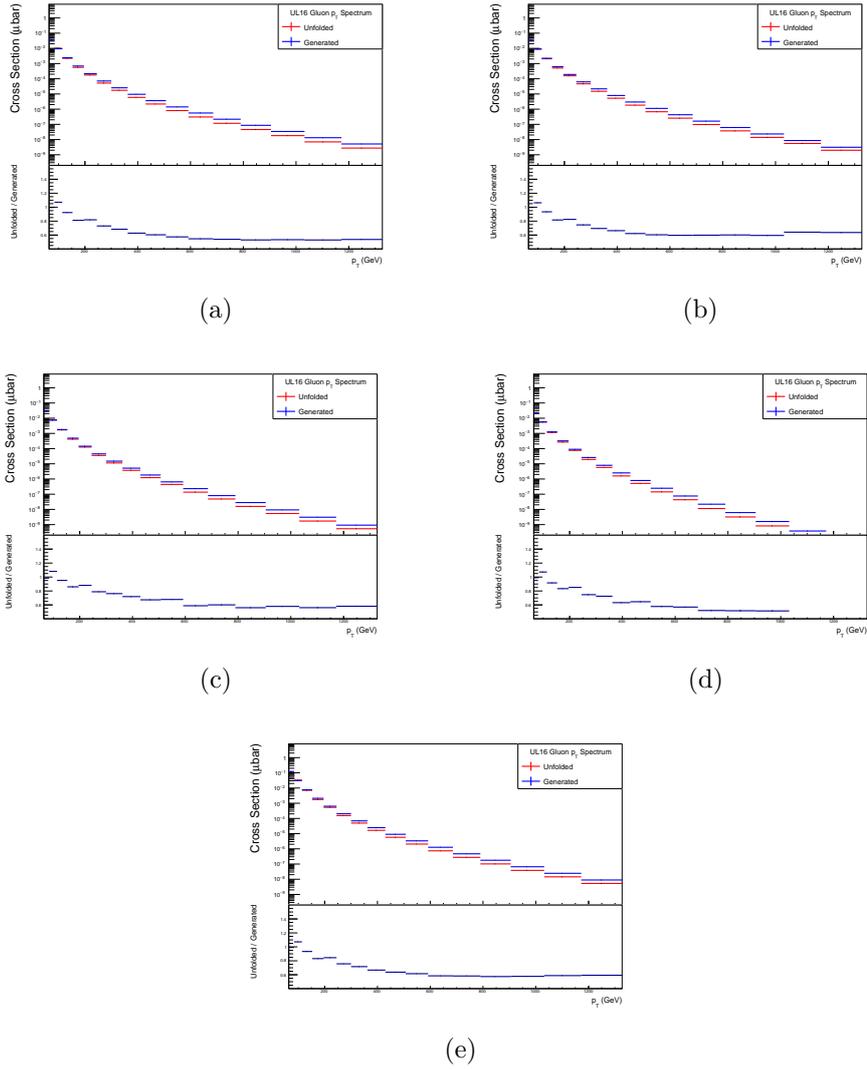


Figure 5.9: Measured and MC-predicted particle-level spectra (upper) and their ratio (bottom) for gluon-jets and for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

Finally Fig. 5.10 shows the final measurement of the unfolded gluon-jet fraction, calculated as the ratio of the unfolded gluon-jet spectrum over the sum of the unfolded gluon-jet and quark-jet spectra, with the yellow band representing the total uncertainty.

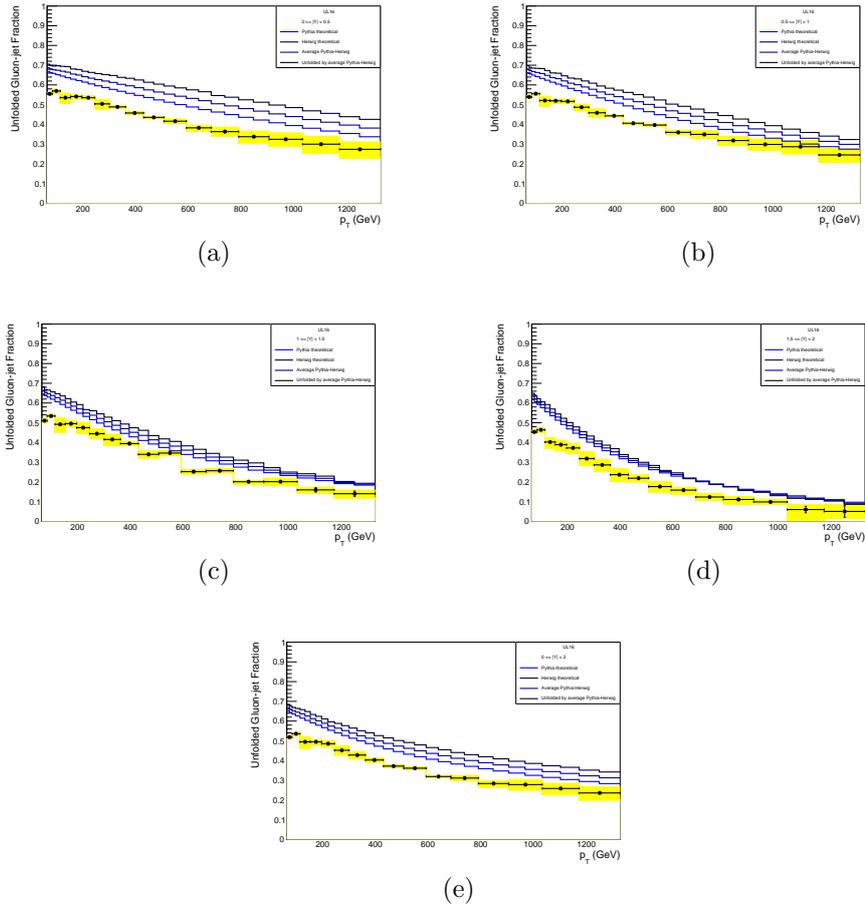


Figure 5.10: Theoretical and best fit gluon-fraction for different $|Y|$ ranges (a): 0 to 0.5, (b): 0.5 to 1, (c): 1 to 1.5, (d): 1.5 to 2, (e): 0 to 2.

Chapter 6

Discussion

We have successfully analyzed the UL16 dataset along with two corresponding MC samples and performed multiple parts of the analysis using various methods to ensure consistency in our results. While we have measured the unfolded gluon-jet fraction in inclusive multi-jet data, further work is required to refine this measurement. A more accurate result will be essential for future gluon-jet based physics explorations. Our results differ from Monte-Carlo simulation predictions, which could be due to several factors. The main potential reason for the discrepancy is the fact that our simulations are LO in QCD, so they don't capture the complexity of the gluon-jet production. For example, higher order or more outgoing partons could alter the gluon-jet content in our data. Future analyses at NLO and NNLO could provide a clearer understanding, as could addressing potential discrepancies between data and the models used in current simulations. Furthermore, our jet reconstruction or jet flavor attribution algorithms might introduce differences between data and simulation. Future improvements could include an infrared and collinear-safe definition, which would allow for direct comparison with NLO and NNLO QCD after they are corrected for non-perturbative effects. Finally, our studies showed differences between the Pythia and Herwig gluon-jet content, which originate from the different showering and hadronization algorithms. These differences become much smaller at larger rapidity bins.

Bibliography

- [1] D. Griffiths. *Introduction to Elementary Particles*. Wiley, (2020), <https://inspirehep.net/literature/806976> .
- [2] Gell-Mann, M. THE EIGHTFOLD WAY: A THEORY OF STRONG INTERACTION SYMMETRY. United States: N. p., (1961). Web, doi:10.2172/4008239 .
- [3] Greig Cowan and Tim Gershon. *Tetraquarks and Pentaquarks*. 2399-2891. IOP Publishing, (2018), arXiv:1808.04153 [physics.pop-ph] .
- [4] Perkins, D. (2000). *Introduction to High Energy Physics* (4th ed.), Cambridge: Cambridge University Press, doi:10.1017/CBO9780511809040 .
- [5] Stefan Höche. Introduction to parton-shower event generators. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders* 11 (2014), arXiv:1411.4085 [hep-ph] .
- [6] Alwall, J., Frederix, R., Frixione, S. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J. High Energ. Phys.* 2014, 79 (2014), <https://doi.org/10.1007> .
- [7] G. Peter Lepage. A New Algorithm for Adaptive Multidimensional Integration. *J. Comput. Phys.*, 27:192, (1978), doi:10.1016/0021-9991(78)90004-9 .
- [8] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, NishitaDesai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, (2015), <https://www.sciencedirect.com/science/article/pii/S0010465515000442> .
- [9] Bellm, J., Bewick, G., Ferrario Ravasio, S. *et al.* Herwig 7.2 release note. *Eur. Phys. J. C* 80, 452 (2020). <https://doi.org/10.1140/epjc/s10052-020-8011-x>
- [10] B. Andersson, “The Lund model”, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 7 (1998) 1, doi:10.1017/CBO9780511524363.

- [11] UA1 Collaboration, “A QCD model for jet fragmentation including soft gluon interference”, *Nucl. Phys. B* 238 (1984) 492, doi:10.1016/0550-3213(84)90333-X.
- [12] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004 .
- [13] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr (2008), doi:10.1088/1126-6708/2008/04/063 .
- [14] CMS collaboration, ”Jet algorithms performance in 13 TeV data”, CMS-PAS-JME-16-003 (2016).
- [15] S.Schmitt, *JINST* 7 (2012) T10003 [arXiv:1205.6201]
- [16] Kalliopi Manitará, ”Study of the Partonic Origin of Hadronic Jets at the CMS experiment of LHC”, MS thesis, University of Ioannina, 2020