Baseline Jet Energy Corrections for the CMS Experiment at the CERN LHC

Master Thesis

Ilias Zisopoulos
Supervisor : Prof. Niki Saoulidou

National & Kapodistrian University of Athens
Department of Physics
February, 2021

Abstract

Jets are the experimental signature of quarks and gluons produced in the final state of high energy scatterings, as the ones that take place in the CMS experiment at the Large Hadron Collider. Numerous physics analyses that not only aim to study Standard Model (SM) Processes, but also to search for physics beyond the SM, use jets, thus creating the need for high accuracy in the estimation of their energy and the associated uncertainties. The factorized Jet Energy Corrections (JEC) that are utilized by the CMS experiment aim to provide exactly that. In this thesis the baseline jet energy corrections that are derived from simulation and constitute the first and main step of the JEC chain, will be presented. These corrections are divided into two levels; the pileup offset corrections (L1) whose goal is to remove the excess jet energy that is caused by additional proton - proton collisions different from the main hard scatter, and the relative and absolute corrections (L2L3) that aim to refer the jet energy at reconstruction level to the respective one at generator level , thus proving an estimation of the true jet energy without the detector resolution and reconstruction effects being considered. Before discussing the jet energy correction estimation and showing the final results, the theoretical aspects of Quantum Chromodynamics and jet formation will be presented. Furthermore, a description of the CMS detector layout will be provided, along with the methods used in order to reconstruct and cluster jets in the CMS experiment.

Key words: Experimental High Energy Physics, Particle Physics

Evaluation Committee

Niki Saoulidou, Associate Professor, Physics Department, NKUA

Vassilis Spanos, Associate Professor, Physics Department, NKUA

Dimitris Fasouliotis, Associate Professor, Physics Department, NKUA

Περίληψη

Ο Μεγάλος Επιταχυντής Αδρονίων LHC (Large Hadron Collider) στο Εργαστήριο CERN παρέχει συγκρούσεις πρωτονίου-πρωτονίου στα 13 TeV, η υψηλότερη ενέργεια μέχρι σήμερα, καθιστώντας τα πειράματα ικανά να ελέγξουν και να διερευνήσουν το Καθιερωμένο Πρότυπο (ΚΜ) με μεγάλη ακρίβεια. Οι αδρονικοί πίδακες είναι οι πειραματικές υπογραφές των κουάρκ και γκλουονίων που παράγονται στην τελική κατάσταση υψηλοενεργειακών σκεδάσεων, όπως αυτές που πραγματοποιούνται στο πείραμα CMS (Compact Muon Solenoid) στον LHC. Ένα μεγάλο πλήθος αναλύσεων φυσικής που στοχεύουν όχι μόνο στην μελέτη υπογραφών του Καθιερωμένου Προτύπου, αλλά και στην εύρεση φυσικής πέραν αυτού, χρησιμοποιούν τους αδρονικούς πίδακες, κάνοντας επιτακτική την ανάγκη για μεγάλης ακρίβειας μετρήσεις της ενέργειας τους και των συστηματικών αβεβαιοτήτων. Στην διπλωματική αυτή εργασία θα παρουσιασθούν οι βασικές διορθώσεις ενέργειας που παράγονται χρησιμοποιώντας προσομοιώσεις (Monte Carlo) και αποτελούν το πρώτο και κύριο μέρος της αλυσίδας των διορθώσεων ενέργειας. Οι διορθώσεις αυτές χωρίζονται σε δύο επίπεδα, αυτές που αφαιρούν την επιπρόσθετη ενέργεια που οφείλεται σε δευτερεύουσες αλληλεπιδράσεις πέραν της κύριας, και αυτές που στοχεύουν στην εξίσωση της ενέργειας σε επίπεδο ανακατασκευής με την αντίστοιχη στο επίπεδο του γεννήτορα που δεν λαμβάνει υπ' όψιν τις ατέλειες του ανιχνευτή και της ανακατασκευής του γεγονότος. Πριν συζητηθούν οι διορθώσεις της ενέργειας, θα παρουσιασθούν τα βασικά χαρακτηριστικά της Κβαντικής Χρωμοδυναμικής και του σχηματισμού των αδρονικών πιδάκων. Επιπλέον, θα περιγραφεί η ανιχνευτική διάταξη του πειράματος CMS και θα συζητηθούν οι τρόποι με τους οποίους ανακατασκευάζονται και ομαδοποιούνται οι αδρονικοί πίδακες στο CMS.

Λέξεις Κλειδιά: Πειραματική Φυσική Υψηλών Ενεργειών, Σωματιδιακή Φυσική

Τριμελής Εξεταστική Επιτροπή
Νίκη Σαουλίδου, Αναπλ. Καθηγ. Τμήματος Φυσικής ΕΚΠΑ
Βασίλειος Σπανός, Αναπλ. Καθηγ. Τμήματος Φυσικής ΕΚΠΑ
Δημήτριος Φασουλιώτης, Αναπλ. Καθηγ. Τμήματος Φυσικής ΕΚΠΑ

Contents

1 Theory **5**

1.1 The Standard Model of Elementary Particles 5

1.2 Scattering Cross Sections 7

1.2.1 $e^- p \to e^- p$ Elastic Scattering 8

1.2.2 $e^- p \to e^- p$ Inelastic & Deep Inelastic Scattering 11

1.3 Quantum Chromodynamics 15

1.3.1 The QCD Lagrangian 15

1.3.2 Formation of Jets 17

1.3.3 $pp \to jets + X$ Deep Inelastic Scattering 19

1.4 Physics Beyond the Standard Model 22

2 The CMS Experiment **26**

2.1 Coordinate System & Kinematics 26

2.2 The CMS Detector 30

2.2.1 Magnet 30

2.2.2 Inner Tracking System 31

2.2.3 Electromagnetic Calorimeter (ECAL) 32

2.2.4 Hadron Calorimeter (HCAL) 35

2.2.5 Muon System 37

3 Jets in CMS **38**

3.1 The Particle Flow Reconstruction 38

3.2 The Anti–k_T Jet Clustering Algorithm 41

3.3 CHS and PUPPI Jets 42

4 Jet Energy Corrections **44**

4.1 Introduction 44

4.2 Ultra Legacy Simulations 45

4.3 Pileup Offset Corrections 49

4.3.1 Investigation of Offset 49

4.3.2 Derivation of the Offset Corrections 56

4.4 Relative and Absolute Corrections 62

4.4.1 Derivation of the Relative and Absolute Corrections 62

4.4.2 Closure and Sanity Checks on the Relative and Absolute Corrections 68

5 Summary and Outlook **72**

A Monte Carlo Samples **74**

B Examination of Pileup Offset **76**

C Comparison among Parametric Functions for the Derivation of Offset Corrections **83**

D Examination of Corrected Response **86**

E Examination of Corrected Response for detailed p_T^{Gen} bins **92**

1 Theory

1.1 The Standard Model of Elementary Particles

The Standard Model (SM) is the best theory that provides an insight into the fundamental structure of matter. All known particles interact with each other through the four fundamental forces in nature; the electromagnetic force, the strong and weak nuclear force and the gravitational force. Both electromagnetic and gravitational forces have an infinite range, while strong and weak forces dominate only at the level of the subatomic particles.

The strength of each force can be encapsulated in a dimensionless constant, named coupling constant. The order of magnitude for each constant can be summarized in the table below.

Strong	$\alpha_s \sim \mathcal{O}(1)$
Electromagnetic	$\alpha \sim \mathcal{O}(10^{-2})$
Weak	$\alpha_w \sim \mathcal{O}(10^{-6})$
Gravitational	$\alpha_g \sim \mathcal{O}(10^{-37} - 10^{-43})$

Table 1: The order of magnitude for the coupling constants of the four fundamental forces in nature.

The gravitational force is the weakest force by a lot of orders of magnitude and has not yet been able for physicists to implement it within the Standard Model. Thus the SM includes the electromagnetic, strong and weak forces and all their carrier particles, and explains well how these forces act on all of the matter particles. More specifically, the elementary particles that comprise matter in the known, observable, universe are 1/2 - spin particles called fermions, along with their corresponding antiparticles. Fermions interact with each other by exchanging force - carrier particles, named gauge bosons, which have an integer spin.

One of the essential ingredients of the SM is the Dirac equation of relativistic quantum mechanics that describes the dynamics of the fermions and their antiparticles:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \tag{1}$$

where ψ is the Dirac spinor of the fermion, m its mass and γ^μ (μ=1,2,3,4) are the four Dirac γ-matrices.

Fermions are divided into two categories; quarks and leptons. The former type of elementary particles interact via all 3 forces in the SM and are the constituents of baryons, as the proton and neutron, and mesons, while on the other hand leptons interact only with the electromagnetic and weak force and comprise particles such as the electron and neutrinos. Both groups consist of six particles that are paired in three generations according to their masses, with the first generation being the lightest and therefore comprising particles that make all stable matter in the universe. A list of such particles can be found in the table below.

Generation	**Quarks**	**Leptons**
1^{st}	up (u) down (d)	electron (e) electron neutrino (ν_e)
2^{nd}	charm (c) strange (s)	muon (μ) muon neutrino (ν_μ)
3^{rd}	top (t) bottom (b)	tau (τ) tau neutrino (ν_τ)

Table 2: Quarks and Leptons in the Standard Model of elementary particles.

As far as the bosons are concerned, each fundamental force results from the exchange of these force - carrier particles; the electromagnetic force is carried by the neutral charged photon (γ), the strong force by the gluon (g) and the weak force by the charged bosons $W^{\pm}$ and the neutral Z. All these particles have a spin of 1.

The underlying gauge symmetry in the quantum theory of the SM implies that all gauge bosons should be massless. Although the photon and gluon indeed are, the carriers of the weak force have been experimentally measured to have a mass of $\approx 80\ GeV/c^2$. This is an outcome of the spontaneous symmetry breaking that is caused by the Higgs mechanism. The Higgs particle is a scalar boson (spin 0) whose field breaks the electroweak symmetry and gives masses to $W^{\pm}$ and Z bosons. It was discovered in 2012 at the Large Hadron Collider and is therefore implemented in the SM.

Within the context of the SM, the interactions between particles are described by Quantum Field Theory. Hence, the electromagnetic interactions are explained by Quantum Electrodynamics that is associated with the $U(1)$ gauge symmetry and the strong force by Quantum Chromodynamics, related to the $SU(3)$ group. Regarding the weak interaction it has been discovered by Glasgow, Salam and Weinberg (GSW) that although at low energies it behaves quite differently than the electromagnetic one, at high energies both forces merge into one, thus providing the electroweak unification.

Ongoing theoretical models suggest that at even higher energies all three forces of the

SM could be unified into a single electronuclear force with many force carriers but a unified coupling constant. Such models, however, are difficult to be studied experimentally due to the impossibly high energies required.

The main focus of the next two sections will be the study of the gauge theory of Quantum Chromodynamics that describes the interactions between quarks through the strong nuclear force, since jets are the experimental signatures of exactly such interactions. Before that however, the kinematics behind the elastic and inelastic scatterings in Quantum Electrodynamics will be presented, as they provide necessary tools to consequently study the interactions between quarks.

1.2 Scattering Cross Sections

The means of studying the characteristics and properties of the SM and more specifically of QCD and the formation of jets, while also searching for undiscovered physics lie on collider experiments, such as the ones being performed at the Large Hardron Collider of CERN. In these experiments protons are being forced to collide at very high energies, where the dynamics of the hard scatter can be approximately described as a two-to-two process between massless partons. Therefore, it is very interesting to study some basic features of the theoretical aspect of such processes.

Before proceeding into details with respect to the elastic and inelastic scattering of particles, the Mandelstam variables are useful to be introduced. These kinematic variables are Lorentz invariant and express the four momentum transfer via the virtual boson in the corresponding Feynman diagram. Examining a two-to-two scattering $1 + 2 \rightarrow 3 + 4$ where each particle i has a four momentum $\mathcal{P}_i = (E_i, \vec{p}_i)$, the Mandelstam variables s, t, u are shown below:

$$\begin{aligned} s &= (\mathcal{P}_1 + \mathcal{P}_2)^2 \\ t &= (\mathcal{P}_1 - \mathcal{P}_3)^2 \\ u &= (\mathcal{P}_1 - \mathcal{P}_4)^2 \end{aligned} \tag{2}$$

Concerning the s variable, in the center-of-mass frame where

$$s = (\mathcal{P}_1 + \mathcal{P}_2)^2 = (E_1 + E_2)^2 - \cancel{(\vec{p}_1 + \vec{p}_2)^2} = E_{total}^2 \tag{3}$$

the total available energy is given by the $\sqrt{s}$ quantity.

The fundamental physics of any interaction between particles is contained within the cross section, which can be considered as the effective cross sectional area associated with

each target particle, when one considers a beam of initial particles with a certain flux. Therefore, the cross section for a process can be defined as:

$$\sigma = \frac{\text{number of interactions per unit time per target particle}}{\text{incident flux}} \tag{4}$$

where the incident flux $\mathcal{F}$ accounts for the relative motion of the particles. It can be easily proven (chapter 7.3 of [4]) that a general expression of the scattering cross section of a two-to-two elastic process is given by:

$$d\sigma = \frac{1}{\mathcal{F}} |\mathcal{M}_{fi}|^2 \, d\Phi_2 \tag{5}$$

where the incident flux is:

$$\mathcal{F} = 4\sqrt{(\mathcal{P}_1 \cdot \mathcal{P}_2)^2 - m_1^2 m_2^2} \tag{6}$$

$\mathcal{M}_{fi}$ is the Lorentz invariant matrix element and $d\Phi_2$ is the phase space of two particles that is calculated as:

$$d\Phi_2 = (2\pi)^4 \delta^{(4)}\Big(\mathcal{P}_1 + \mathcal{P}_2 - \mathcal{P}_3 - \mathcal{P}_4\Big) \frac{d^3\vec{p}_3}{2E_3(2\pi)^3} \frac{d^3\vec{p}_4}{2E_4(2\pi)^3} \tag{7}$$

As a result, the calculation of the differential cross section for a particular process lies on the determination of its matrix element $\mathcal{M}_{fi}$, since the phase space depends on the observation system and the kinematics of the problem, and not its physical characteristics.

1.2.1 $e^- p \to e^- p$ Elastic Scattering

In order to be able to describe and predict the theoretical cross sections of the deep inelastic scatterings between protons that take place in the LHC machine one should start at first from simpler processes. For that reason, the electron - proton scattering will be initially examined.

The precise nature of the electron - proton scattering depends on the wavelength of the virtual photon in comparison to the radius of the proton. At very low energies of the incoming electron the dominant process is the elastic scattering, which is described by the coherent interaction of a virtual photon γ and the proton as a whole, thus providing information regarding the global properties of the proton. In Fig. 1 the kinematics of such a process, along with the corresponding Feynman diagram are shown.

In such an elastic scattering the proton can be treated as if it was a point-like Dirac particle, and therefore, using the first order term in the perturbation expansion, the matrix element for this Feynman diagram can be written as:

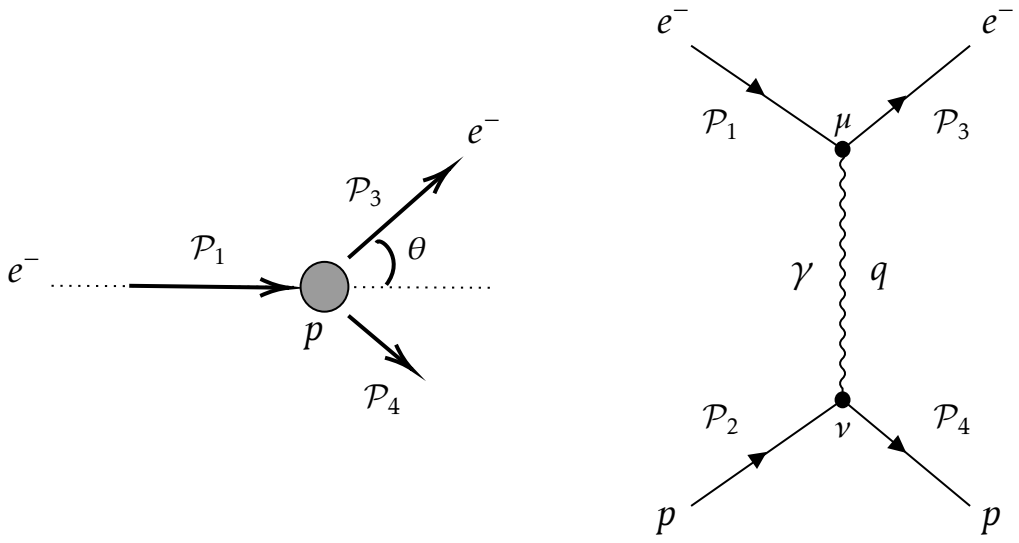

Figure 1: Elastic scattering of an electron from a proton at rest in laboratory frame and the corresponding Feynman diagram.

$$\mathcal{M}_{fi} = \frac{e^2}{q^2}\Big[\overline{u}(\mathcal{P}_3)\gamma^\mu u(\mathcal{P}_1)\Big]\, g_{\mu\nu}\, \Big[\overline{u}(\mathcal{P}_4)\gamma^\mu u(\mathcal{P}_2)\Big] \equiv \frac{e^2}{q^2} j_e \cdot j_p \tag{8}$$

where $q^2 = (\mathcal{P}_1 - \mathcal{P}_3)^2$ is the squared four momentum carried by the virtual photon in this t-channel $e^- p \to e^- p$ process, u are the Dirac spinors and j_e, j_p are the electron and proton currents respectively.

At the low energy limit of the $e^- p$ scattering the electron energy is sufficiently low so that the kinetic energy of the recoiling proton is negligible compared to its rest mass. Working on the laboratory frame and considering that the electron is non-relativistic one can obtain the *Rutherford scattering process*, where only the interaction between the electric charges of the electron and proton contribute to the scattering process, since there is no significant contribution from the magnetic (spin–spin) interaction. The differential cross section of this process is given by:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Rutherford}} = \frac{\alpha^2}{16 E_K \sin^4(\theta/2)} \tag{9}$$

where $d\Omega$ is the steradian of the scattered electron, $\alpha = e^2/4\pi$ is the electromagnetic coupling constant and $E_K = p^2/2m_e$ is the non-relativistic incoming electron energy.

When the electron is relativistic, namely $m_e \ll E \ll m_p$, the elastic scattering needs to be corrected, arriving to the *Mott scattering*. The cross section for this case is as follows:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} = \frac{\alpha^2}{4E^2 \sin^4(\theta/2)} \cos^2(\theta/2) \tag{10}$$

where E is the incoming electron energy.

When the energies of the electron - proton scattering increase, so the wavelength of the virtual photon is smaller than the radius of the proton, the proton can not be considered as a point-like particle anymore and the treatment that was earlier described needs to be modified with a form factor. This form factor accounts for the phase differences between contributions to the scattered wave from different points of the charge distribution. As a result the matrix element needs to be corrected as $\mathcal{M}_{fi} \rightarrow \mathcal{M}_{fi}\, F(\vec{q}^{\,2})$, where the form factor $F(\vec{q}^{\,2})$ is the 3D transformation of the charge distribution:

$$F(\vec{q}^{\,2}) = \int \rho(\vec{r}) e^{i\vec{q}\cdot\vec{r}} d^3\vec{r} \tag{11}$$

Consequently, the Mott scattering cross section (Eq. 10) is modified to:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} \rightarrow \frac{\alpha^2}{4E^2 \sin^4(\theta/2)} \cos^2(\theta/2) \left|F(\vec{q}^{\,2})\right|^2 \tag{12}$$

In the Rutherford and Mott processes (Eq. 9 and 10) the low energy limit $|\vec{q}| \ll m_p$ is considered, where the proton recoil can be safely ignored. However, for electron - proton elastic scattering at higher energies, where the contribution to the scattering process from the pure magnetic spin–spin interaction is not negligible, the kinematics of the scattering need to be modified. Therefore, working on the proton rest frame, and considering it as a point-like particle the cross section this time is given by:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4E_1^2 \sin^4(\theta/2)} \frac{E_3}{E_1} \left(\cos^2(\theta/2) + \frac{Q^2}{2m_p^2} \sin^2(\theta/2) \right) \tag{13}$$

where E_1, E_3 are the initial and final state electron energies respectively and $Q^2 \equiv -q^2$. It is important to state that both Q^2 and E_3 can be expressed in terms of the scattering angle θ, thus measuring the scattering angle of the electron, the entire kinematics of the interaction are determined.

Finally, Eq. 13 that describes the differential cross section for elastic $e^- p \rightarrow e^- p$ scattering assuming a point-like 1/2 spin proton needs to be corrected so that the finite size of the proton is accounted for. This is performed by introducing two form factors; one related to the charge distribution of the proton, $G_E(Q^2)$, and the other related to the magnetic moment distribution within the proton, $G_M(Q^2)$. Hence, it can be shown that the most general Lorentz invariant form for the electron - proton scattering via the exchange of a single photon is given by the *Rosenbluth formula*:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4E_1^2 \sin^4(\theta/2)} \frac{E_3}{E_1} \left(\frac{G_E^2 + \tau G_M^2}{1+\tau} \cos^2(\theta/2) + 2\tau G_M^2 \sin^2(\theta/2) \right) \tag{14}$$

where τ is given by $\tau = Q^2/4m_p^2$. There is no a priori knowledge of the form factors $G_E(Q^2)$ and $G_M(Q^2)$, so they are obtained from experimental data. In more detail, they can be inferred from $e^-p \to e^-p$ elastic scattering experiments by varying the electron beam energy and measuring the differential cross section at different scattering angles. When the virtual photon has a large Q^2 both $G_E(Q^2)$ and $G_M(Q^2)$ become small and the elastic scattering cross section falls rapidly with increasing Q^2.

1.2.2 $e^-p \to e^-p$ Inelastic & Deep Inelastic Scattering

For the above reason, high energy electron - proton interactions are dominated by inelastic scattering where the proton breaks up and the virtual photon interacts with the quarks inside the proton. Such $e^-p \to e^-X$ process can be viewed schematically in Fig. 2.

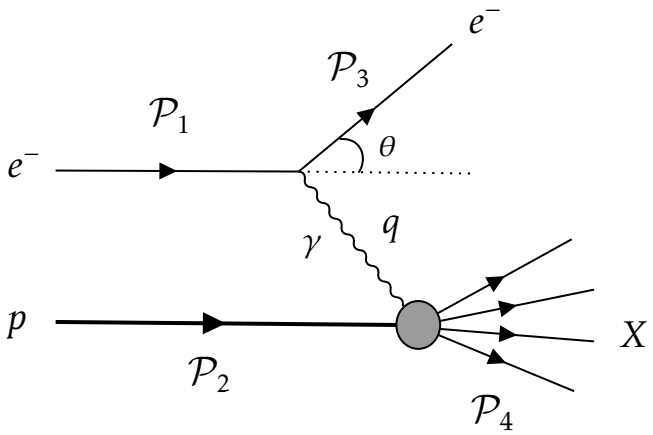

Figure 2: Electron - proton inelastic scattering.

In order to describe the inelastic scattering a few kinematic variables need to be introduced. As was already defined in the elastic scattering, the negative four momentum squared for the virtual photon is:

$$Q^2 = -q^2 = -(\mathcal{P}_1 - \mathcal{P}_3)^2 \approx 4E_1E_3\sin^2(\theta/2) \tag{15}$$

where the electron mass has been neglected, since the energies are sufficiently high. Next, a Lorentz invariant dimensionless quantity called Bjorken x is defined as:

$$x = \frac{Q^2}{2\mathcal{P}_2 \cdot q} \tag{16}$$

Due to the fact that the invariant mass W of the final state hadronic system is given by $W^2 = \mathcal{P}_4^2 \geq m_p^2$ the Bjorken x is always in the range $0 \leq x \leq 1$, with $x = 1$ giving the elastic scattering. Furthermore, the dimensionless inelasticity y, also constrained in the range $0 \leq y \leq 1$, is defined as:

$$y = \frac{\mathcal{P}_2 \cdot q}{\mathcal{P}_2 \cdot \mathcal{P}_1} \tag{17}$$

and expresses the fractional energy lost by the electron in the scattering process in the frame where the proton is initially at rest. The last variable to be defined is given by:

$$\nu = \frac{\mathcal{P}_2 \cdot q}{m_p} \tag{18}$$

and is convenient to work with since it is expressed in terms of energy rather than fractional energy loss.

Using the above kinematic variables, the Rosenbluth formula (Eq. 14) for the elastic scattering cross section can be rewritten as follows:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1 - y - \frac{m_p^2 y^2}{Q^2}\right)\underbrace{\frac{G_E^2 + \tau G_M^2}{1+\tau}}_{f_2(Q^2)} + \frac{1}{2}y^2 \underbrace{G_M^2}_{f_1(Q^2)}\right] \tag{19}$$

where the Q^2 dependence of the form factors $G_E(Q^2)$ and $G_M(Q^2)$ and τ are absorbed into two new functions, written as $f_1(Q^2)$ and $f_2(Q^2)$.

The above formula can be generalised to the inelastic scattering process, where this time the differential cross section must be expressed as a function of two independent kinematic variables instead of only Q^2. It can be shown that the most general expression for the cross section of the inelastic $e^- p \to e^- X$ scattering is given by:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1 - y - \frac{m_p^2 y^2}{Q^2}\right)\frac{F_2(x,Q^2)}{x} + y^2 F_1(x,Q^2)\right] \tag{20}$$

where the functions $f_1(Q^2)$ and $f_2(Q^2)$ have been replaced by the two structure constants of the inelastic scattering; $F_1(x,Q^2)$ and $F_2(x,Q^2)$. For the *deep inelastic scattering (DIS)* between the electron and the proton $Q^2 \gg m_p^2 y^2$, the differential cross section reduces to:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1 - y\right)\frac{F_2(x,Q^2)}{x} + y^2 F_1(x,Q^2)\right] \tag{21}$$

The structure constants were expected to be determined from the experimental data, which showed two remarkable features. The first observation is called *Bjorken scaling* and revealed that both $F_1(x,Q^2)$ and $F_2(x,Q^2)$ are almost independent of Q^2 allowing us to rewrite the structure constants as:

$$\begin{aligned} F_1(x,Q^2) &\to F_1(x) \\ F_2(x,Q^2) &\to F_2(x) \end{aligned} \tag{22}$$

The second feature of the structure constants showed that in the deep inelastic scattering regime, namely for Q^2 greater than a few GeV^2, they are not independent with each other, but satisfy the *Callan - Gross relation*:

$$F_2(x) = 2xF_1(x) \tag{23}$$

This can be explained by the fact that in the context of the deep inelastic scattering the electron is scattered from point-like constituents of the proton, namely quarks.

The cross section of the deep inelastic scattering (Eq. 21) can be also extracted via the quark model. In this quark model, the deep inelastic scattering cross sections are related to the cross section for the QED process of $e^-q \to e^-q$ elastic scattering. Hence, this process can be easily described by a Feynman diagram similar to the one depicted in Fig. 1 with the proton being replaced by the quark. Working the kinematics on the center-of-mass frame this time and neglecting the electron and quark masses, since we are in the high energy limit, the differential cross section of the $e^-q \to e^-q$ elastic scattering process can be expressed as:

$$\frac{d\sigma}{dq^2} = \frac{2\pi\alpha^2 Q_q^2}{q^4}\left[1 + \left(1 + \frac{q^2}{s_q}\right)^2\right] \tag{24}$$

where $Q_q e$ is the quark charge and $\sqrt{s_q}$ the center-of-mass available energy in the quark level system.

In the *quark - parton model* the basic interaction in deep inelastic electron – proton scattering is elastic scattering from a 1/2 spin quark within the proton, where the quark is considered as a free particle. In the frame where the proton has very large energies $E \gg m_p$ (infinite momentum frame) the struck quark is considered to have a four momentum of $\mathcal{P}_q = \xi\mathcal{P}_2$, where ξ is the momentum fraction that the quark carries. It can easily be proven that ξ is equal to Bjorken x. As a result, one can associate the kinematic variables for the underlying e^-q scattering process with those for the e^-p collision as follows:

$$\begin{aligned} s_q &= xs \\ y_q &= y \\ x_q &= 1 \end{aligned} \tag{25}$$

The differential cross section from Eq. 24 consequently can be expressed as:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2 Q_q^2}{Q^4}\left[(1-y) + \frac{y^2}{2}\right] \tag{26}$$

In order to transition from the e^-q process to the deep inelastic scatter of e^-p which is under examination, the *Parton Distribution Functions (PDFs)* need to be introduced. Since

quarks inside the proton interact with each other through the exchange of gluons, the dynamics of this interacting system will result in a distribution of quark momenta within the proton, which is described by the PDFs. Therefore, the up or down quark PDF for the proton $q^p(x)$ ($q = u$ or d quark) is defined such that $q^p(x)\delta(x)$ represents the number of q-flavored quarks within the proton with momentum fraction between x and $x + \delta x$. Hence, the cross section for elastic scattering from a particular flavour of quark i with charge Q_i and momentum fraction in the range $x \to x + \delta x$ is obtained from properly modifying Eq. 26 as:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[(1-y) + \frac{y^2}{2}\right] \times Q_i^2 q_i^p(x)\delta(x) \tag{27}$$

Finally, the differential cross section for the deep inelastic $e^- p$ is obtained by summing over all quark flavors:

$$\frac{d^2\sigma^{ep}}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[(1-y) + \frac{y^2}{2}\right] \sum_i Q_i^2 q_i^p(x) \tag{28}$$

Comparing the above formula with Eq. 21 leads to the parton model predictions for $F_1^{ep}(x, Q^2)$, $F_2^{ep}(x, Q^2)$:

$$F_2^{ep}(x, Q^2) = 2xF_1^{ep}(x, Q^2) = x\sum_i Q_i^2 q_i^p(x) \tag{29}$$

The PDFs can not be predicted from first principles since the theory of Quantum Chromodynamics (QCD) has a large coupling constant ($\alpha_S \sim 1$) that does not allow perturbation theory and need to be determined through experimental techniques. It is clear that the static model of quarks is overly simplistic and can not be used to determine the PDFs. Instead, the proton is considered to have 3 valence quarks and a sea of virtual gluons that give rise to $g \to q\overline{q}$ production. As a result, the PDFs need to be written as $q^p(x) = q_V(x) + q_S(x)$ and $\overline{q}^p(x) = \overline{q}_S^p(x)$.

Since the deep inelastic scattering $e^- p$ has been examined, the next step is to study the deep inelastic scatterings that take place inside the LHC between protons. Up until now, however, all the interactions that were examined involved the exchange of a virtual photon, namely they were electromagnetic processes. In order to examine the strong interactions between protons the basic features of Quantum Chromodynamics (QCD) need to be addressed. Nevertheless, as will be discussed in the subsequent section, the basic tools for the description of such processes have already been presented and only need to be modified properly to account for the characteristics of QCD. More detailed information regarding the cross section calculations that were presented in this section can be found in [1].

1.3 Quantum Chromodynamics

1.3.1 The QCD Lagrangian

Quantum Chromodynamics (QCD) is the quantum field theory of the strong interaction, that bears a lot of similarities with Quantum Electrodynamics (QED) but has some fundamental differences that will be briefly presented in this section.

In order to better understand the underlying symmetry associated with QCD, which is invariance under $SU(3)$ local phase transformations, one can start with the description of the much simpler $U(1)$ local gauge invariance of QED. More specifically, the free Lagrangian that gives rise to the Dirac equation (1)

$$\mathcal{L}_0 = i\overline{\psi}\gamma_\mu\partial^\mu\psi - m\overline{\psi}\psi \tag{30}$$

where ψ is the electron field spinor, is invariant under the transformation

$$\psi(x) \rightarrow e^{i\alpha(x)}\psi(x) \tag{31}$$

where $\alpha(x)$ depends on space and time in a completely arbitrary way, under the condition that the modified derivative D_μ is introduced:

$$D_\mu = \partial_\mu - ieA_\mu \tag{32}$$

where the vector field A_μ, called gauge field, is regarded as the physical photon field and transforms as:

$$A_\mu \rightarrow A_\mu + \frac{1}{e}\partial_\mu\alpha \tag{33}$$

Inserting a kinematic term of the physical photon field that involves the gauge invariant field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{34}$$

the QED Lagrangian is given by:

$$\mathcal{L}_{QED} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\overline{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{35}$$

In an analogous approach one can proceed to infer the structure of QCD from local gauge invariance, replacing $U(1)$ gauge group with the $SU(3)$ one. The main difference with the theory of QED for electrons is that quarks carry color as well as electric charge. There are three colors, r, g and b and are exchanged by 8 bicolored gluons, as an outcome

of the $SU(3)$ group, on the contrary to QED that has a single photon. Therefore, one can start with the free Lagrangian

$$\mathcal{L}_0 = \overline{q}_\kappa (i\gamma^\mu \partial_\mu \delta_{\kappa\lambda} - m\delta_{\kappa\lambda}) q_\lambda \tag{36}$$

where for simplicity only one quark flavor $q(x)$ is shown, and κ, λ denote the three color fields. As was done in QED one can demand invariance under local phase transformation of the form:

$$q(x) \rightarrow Uq(x) = e^{i\alpha_\alpha(x)T_\alpha} q(x) \tag{37}$$

where U is an arbitrary 3×3 matrix, and summation over the repeated suffix $\alpha = 1, ..., 8$ is implied. Furthermore, $\alpha_\alpha(x)$ are the group parameters, and T_α are a set of linearly independent traceless 3×3 matrices that are chosen to be

$$T_\alpha = \frac{\lambda_\alpha}{2} \tag{38}$$

where λ_α are the eight Gell - Mann matrices. The $SU(3)$ group, on the contrary to $U(1)$ is Non - Abelian, since not all the generators T_α commute with each other. Instead, it can be shown that they satisfy the commutation relation:

$$[T_\alpha, T_b] = i f_{\alpha bc} T_c \tag{39}$$

where $f_{\alpha bc}$ are the structure constants of the group, and are real constants.

To impose the local gauge invariance on the QCD Lagrangian one could follow exactly the same approach as in QED and define the eight gluon gauge fields G^α_μ that transform similar to the photon field A_μ:

$$G^\alpha_\mu \rightarrow G^\alpha_\mu - \frac{1}{g_s} \partial_\mu \alpha_\alpha \tag{40}$$

introducing at the same time the covariant derivative:

$$D_\mu = \partial_\mu + i g_s T_\alpha G^\alpha_\mu \tag{41}$$

where g_s is the gauge coupling for the strong interaction, similar to e in QED.

However, this strategy does not entirely work for a Non - Abelian theory like QCD, as it is not enough to ensure gauge invariance of the Lagrangian. In order to achieve this, the gluon gauge fields should instead transform as:

$$G^\alpha_\mu \rightarrow G^\alpha_\mu - \frac{1}{g_s} \partial_\mu \alpha_\alpha - f_{\alpha bc} \alpha_b G^c_\mu \tag{42}$$

and thus the gluon field strength tensor has an extra third term than its counterpart in QED:

$$G^{\alpha}_{\mu\nu} = \partial_{\mu} G^{\alpha}_{\nu} - \partial_{\nu} G^{\alpha}_{\mu} - g_s f_{\alpha bc} G^{b}_{\mu} G^{c}_{\nu} \tag{43}$$

As a consequence, the Lagrangian 36 for one quark flavor becomes:

$$\mathcal{L} = \overline{q}_{\kappa}(i\gamma^{\mu}\partial_{\mu}\delta_{\kappa\lambda} - m\delta_{\kappa\lambda})q_{\lambda} - g_s(\overline{q}_{\kappa}\gamma^{\mu}T^{\alpha}_{\kappa\lambda}q_{\lambda})G^{\alpha}_{\mu} - \frac{1}{4}G^{\alpha}_{\mu\nu}G^{\mu\nu}_{\alpha} \tag{44}$$

so summing over all flavors gives the final form of the QCD Lagrangian:

$$\mathcal{L}_{QCD} = \sum_{j} \overline{q}_{j,\kappa}\Big(i\gamma^{\mu}\partial_{\mu}\delta_{\kappa\lambda} - g_s\gamma^{\mu}T^{\alpha}_{\kappa\lambda}G^{\alpha}_{\mu} - m\delta_{\kappa\lambda}\Big)q_{j,\lambda} - \frac{1}{4}G^{\alpha}_{\mu\nu}G^{\mu\nu}_{\alpha} \tag{45}$$

where $q_{j,\kappa}$ represents the quark spinor of flavor j and color $\kappa, \lambda \rightarrow 1, 2, 3$ (r, g, b).

1.3.2 Formation of Jets

A remarkable feature of QCD that is a result of its Non - Abelian nature is the fact that gluons, as mediators of the strong force, carry color and are self interacting. This property is a direct consequence of the third term of Eq. 43. More specifically, if the QCD Lagrangian 45 is written in the symbolic form:

$$\mathcal{L}_{QCD} = "\overline{q}q" + "G^2" + g_s"\overline{q}qG" + g_s"G^3" + g_s^2"G^4" \tag{46}$$

then the first three terms have a QED analogue and describe the free propagation of quarks and gluons and the quark - gluon interaction respectively. However, the remaining two terms are only encountered in QCD and reflect the fact that gluons self interact, as seen in Fig. 3. These self interacting vertices result in *color confinement*, that will be discussed subsequently and play a crucial role in the formation of jets.

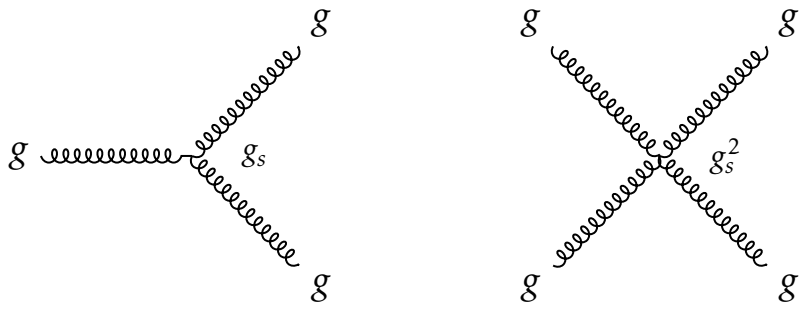

Figure 3: Gluon self interacting vertices. In the left-hand side is the Feynman diagram for the Lagrangian term $g_s"G^3"$ and in the right-hand side for the $g_s^2"G^4"$ term.

Although there is a plethora of experimental data that suggest the existence of quarks and gluons, in reality no free parton has been detected. The non observation of free quarks and gluons is explained by the color confinement hypothesis that suggests that colored objects are confined in color singlet states and cannot propagate freely.

As a result, when for example a quark - antiquark pair is pulled apart, they interact via exchanging a virtual gluon. Unlike the photon in QED however, this gluon carries color charge and self interacts thus squeezing the color field between the quarks into a tube. The energy stored in this tube is proportional to the separation of the quarks. Therefore, the energy increases linearly with the distance between the quarks, forcing these color objects to rearrange themselves into hadronic states that are colorless combinations without a color field between them. For similar reasons, since gluons also carry color charge they can not propagate over macroscopic distances like photons and are confined in colorless objects instead.

The fact that only color singlet states can be observed as free particles restricts the structure of possible hadronic states. In more detail, a bound qq state can not exist as the combination of two color triplets yields a sextet and a triplet ($3 \otimes 3 = 6 \oplus 3$). On the other hand, for a bound $q\overline{q}$ state the possible color wavefunctions form an octet and a singlet ($3 \otimes \overline{3} = 8 \oplus 1$) and hence the color wavefunction for mesons is given by the singlet state:

$$\psi^C(q\overline{q}) = \frac{1}{\sqrt{3}}(r\overline{r} + g\overline{g} + b\overline{b}) \tag{47}$$

When adding a third quark to form a state qqq, the combination of three colors yields a color singlet as $3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1$. So, the color wavefunction for baryons is given by the singlet state:

$$\psi^C(qqq) = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr) \tag{48}$$

while the combinations $qq\overline{q}, q\overline{q}\,\overline{q}$ do not produce color singlet states and as a consequence do not exist in nature.

The color confinement hypothesis results in the formation of jets when high energy quarks are produced in the final state of scattering processes. Since they can not propagate as free particles, high energy partons produce jets via the *hadronisation* process. A schematic view of the hadronisation process can be found in Fig. 4. When a $q\overline{q}$ pair is produced in a high energy interaction, the quark and antiquark separate at high velocities. Due to color confinement, the more they separate the more energy is stored in the color tube between them. At some point this energy is sufficient enough to produce a new $q\overline{q}$ pair, and this process is continued until all the quarks and antiquarks have sufficiently

low energy to combine to form colorless hadrons. The phenomenon where a quark or gluon has a finite probability to split into two partons due to this tube energy is called *parton branching*. The final result is the formation of two hadronic jets, one following the initial quark direction and the other the initial antiquark direction. These jets essentially comprise a narrow cone of mainly hadrons and other particles, whose momenta are collinear to the original parton that caused the hadronisation.

In high energy experiments, as the ones at the LHC, where QCD processes take place, hadronic jets are the experimental manifestation of high energy quarks and gluons that are produced in the final state. For that reason by measuring and studying jets in hadron colliders, information regarding the properties of the original partons can be obtained. Therefore, analyses with jets in final states are very important both for studying the characteristics of QCD and searching for exotic physics beyond the Standard Model.

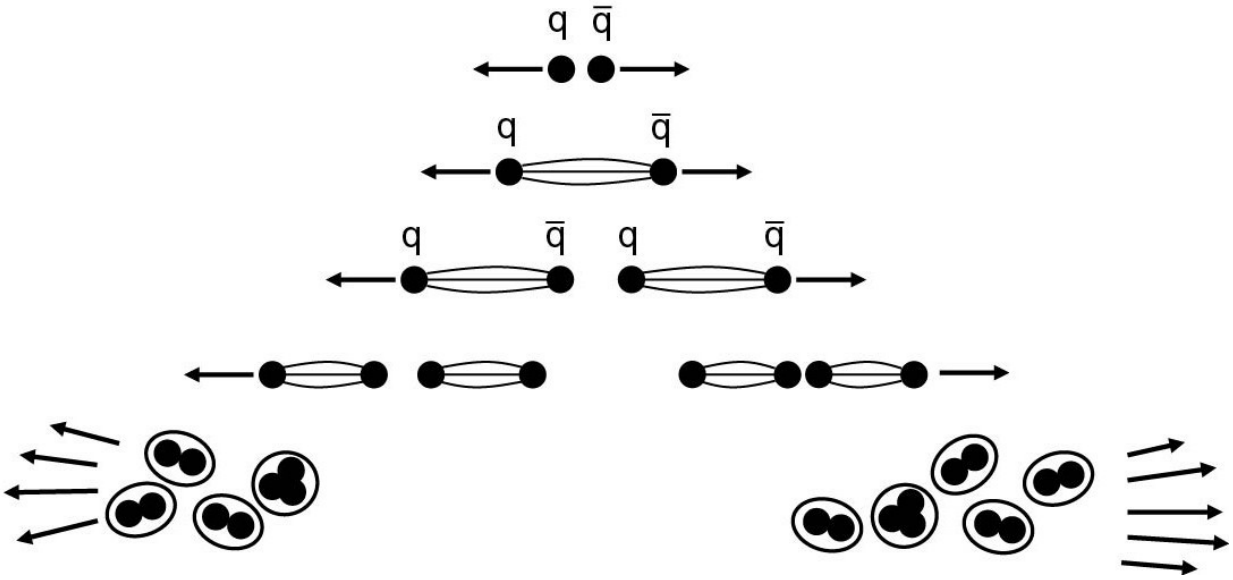

Figure 4: Schematic view of the hadronisation process that leads to the formation of hadronic jets. [1]

1.3.3 $pp \rightarrow jets + X$ Deep Inelastic Scattering

Having discussed the main properties of QCD and how jets are formed one can proceed with studying the deep inelastic scattering between protons that take place in the LHC machine, with the help of the formalisation of the $e^- p$ deep inelastic scattering that was discussed in section 1.2.2. There, in the context of the quark - parton model it was mentioned that the basic interaction of the electron - proton DIS is the elastic scattering

[1] Taken from Chapter 10.4.2 of [1] (Figure 10.9)

of the electron from a quark within the proton that can be considered as free. The reason why this assumption was made was due to the *asymptotic freedom* phenomenon.

The coupling between a photon and a charged fermion in QED, referred to as e up until this point throughout section 1.2 is not the electron charge that is measured experimentally, rather than a bare electron charge e_0 in the lowest order Feynman diagram. The experimental charge is the effective strength of the interaction which results from summing over all higher order diagrams. In QED the running coupling constant $\alpha(q^2) = e^2(q^2)/4\pi$ has a dependence on the momentum transfer of the virtual photon, and more specifically it increases only slightly with increasing q^2, with $\alpha(q^2 = 0) \approx 1/137$. Even at higher q^2 though, the coupling constant is still very small and perturbation theory can be applied.

In QCD on the other hand, the running coupling constant is large, namely $\alpha_s \sim \mathcal{O}(1)$, which prohibits any application of the perturbation theory. Fortunately however, due to the gluon self interactions the running coupling constant in QCD decreases as the momentum transfer q^2 increases. As a result, at high energy interactions, namely $|q| > 100$ GeV, which is the typical scale for modern high energy collider experiments, the strong coupling becomes $\alpha_s \sim 0.1$, which is sufficiently small to allow perturbation theory to be used. This is the property of QCD known as asymptotic freedom and is the key that allows us to treat quarks as free particles inside the proton at very high energy interactions. It is worth stating here also that α_s may be small at high q^2 but unlike QED it is not too small for higher order terms to be neglected. As a consequence, QCD calculations for the processes at LHC are always beyond the leading order (LO).

The proton - proton deep inelastic scattering can be examined by considering that a quasi free quark within the first proton interacts with a quasi free quark from the second proton by exchanging a gluon. Therefore the main focus is to study $qq \to qq$ processes, which can be described with respect to the $e^- q \to e^- q$ that were examined in section 1.2.2 with some modifications.

In more detail, the Feynman diagrams for the $e^- q \to e^- q$ and $qq \to qq$ with quark colors $ik \to jl$ in Fig. 5 are very similar. The matrix element for the QCD process for the particular combination of quark colors $ik \to jl$ can be obtained by the already calculated QED matrix element by replacing $-Q_q \alpha \to \alpha_s$ ($Q_q e$ is the quark electric charge) and inserting a *color factor* $C(ik \to jl)$. This color factor accounts for the 3^4 possible color combinations between the four quarks and 8 possible gluons that can be exchanged. It is given by:

$$C(ik \to jl) = \frac{1}{4} \sum_{\alpha=1}^{8} \lambda_{ji}^{\alpha} \lambda_{lk}^{\alpha} \tag{49}$$

where λ_{ji}^{α} is the ji element of the Gell - Mann matrix with index α. All the possible color

combinations are accounted for by the color averaged sum of squared matrix elements

$$< |\mathcal{M}|^2 >= \frac{1}{9} \sum_{i,j,k,l=1}^{3} |\mathcal{M}(ik \to jl)|^2 \tag{50}$$

which can be derived by the averaged color factor that is easy to prove that is given by:

$$< |C|^2 >= \frac{1}{9} \sum_{i,j,k,l=1}^{3} |C(ik \to jl)|^2 = \frac{2}{9} \tag{51}$$

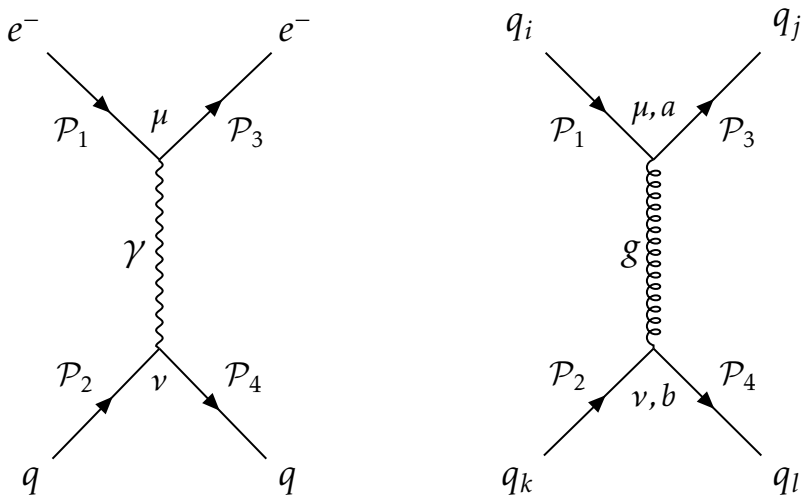

Figure 5: Feynman diagrams for $e^-q \to e^-q$ (left) and $qq \to qq$ with quark colors $ik \to jl$ (right).

Furthermore, color factors can be defined for single vertex processes in the parton branching process. The color factors for quark and gluon splitting are:

$$\begin{aligned} C(g \to gg) &= 3 \\ C(g \to q\bar{q}) &= \frac{4}{3} \\ C(q \to gq) &= \frac{1}{2} \end{aligned} \tag{52}$$

As a result, gluons systematically shower more than quarks.

Finally, in order to calculate the differential cross section of the $qq \to qq$ scattering one can use the $e^-q \to e^-q$ cross section of Eq. 24 and just replace $Q_q^2\alpha^2 \to \alpha_s^2$ and multiply with the averaged color factor:

$$\frac{d\sigma}{dQ^2}(qq \to qq) = \frac{4\pi\alpha_s^2}{9Q^4}\left[1 + \left(1 - \frac{Q^2}{\hat{s}}\right)^2\right] \tag{53}$$

where $Q^2 = -q^2$ and $\sqrt{\hat{s}}$ is the center-of-mass energy of the colliding qq system. In the $e^- p$ elastic scattering only one variable was enough to calculate the cross section; the scattering angle, while in the deep inelastic scattering two variables were needed; the momentum transfer Q^2 and the unknown momentum fraction x of the struck quark. Here, there are two unknown momentum fractions x_1 and x_2, with $\hat{s} = x_1 x_2 s$, where $\sqrt{s}$ is the center-of-mass available energy in the proton - proton colliding system. Hence, the differential cross section of the process can be written as:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha_s^2}{9Q^4}\left[1 + \left(1 - \frac{Q^2}{x_1 x_2 s}\right)^2\right] f(x_1, x_2) dx_1 dx_2 \tag{54}$$

where $f(x_1, x_2)$ is the sum over the products of the relevant parton distribution functions for the scattering process $qq \to qq$. Therefore, the cross section becomes:

$$\frac{d^3\sigma}{dQ^2 dx_1 dx_2} = \frac{4\pi\alpha_s^2}{9Q^4}\left[1 + \left(1 - \frac{Q^2}{x_1 x_2 s}\right)^2\right] f(x_1, x_2) \tag{55}$$

In order to calculate the cross section for the $pp \to jets + X$ scattering the above formula is not enough, since $qq \to qq$ is not the only process that contributes to the overall proton - proton scattering. There are a lot more Feynman diagrams that contribute to two jets in the final state and need to be summed to obtained the total cross section. In addition, as was mentioned before, higher order terms in the perturbation process need to be taken into consideration. Therefore, it is clear that the cross section calculations for hadron - hadron deep inelastic scattering at the LHC can be become quite complex.

1.4 Physics Beyond the Standard Model

The Standard Model has explained numerous phenomena and is consistent with the experimental data that have been processed so far. However, there are still many open questions it has not been able to answer and therefore it is not the final theory in particle physics. Dark matter and energy, supersymmetry and string theory are some of the most famous theoretical models that try to complement the SM.

Collider experiments, where deep inelastic scattering between hadrons take place, are possible candidates to provide physicists with traces of new, exotic particles outside the SM. There is a plethora of theoretical models that suggest the existence of such particles that are produced in high energy collisions and give rise to quarks and gluons in the final state. Hence, there are many physics analyses that study jets; for example almost 60% of all analyses in the CMS experiment use jets in the final state.

A few interesting theoretical models of partonic resonances where new particles decay into a pair of jets and appear as dijet resonances are discussed in [3].

For instance, many theoretical models propose that quarks are composite objects, named excited quarks and denoted as q^*. Through the quark - gluon fusions these particles could be produced in the LHC and subsequently decay into a pair of an ordinary quark and a gause boson, thus leading to a dijet signature.

Another example is the Randal and Sundrum gravity model that suggests the existence of spin-2 gravitons that appear as Kaluza - Klein excitations of the gravitational field and under certain conditions can decay into partons and leave an experimental signature via jets.

Dark matter mediators that arise from an interaction between quarks and dark matter, string resonances that originate from the Regge excitations of quarks and gluons, axigluons and colorons, new gauge bosons $W^{'}$, $Z^{'}$, are also some of the exciting new physics models that could be discovered in the LHC by searching for dijet resonances.

One very interesting theoretical model, presented in [5], that will be briefly discussed in this section proposes the existence of a diquark scalar plus a vectorlike quark that give rise to ultra heavy resonances, with four jets in the final state.

In more detail, in this model a complex scalar field S_{uu} that is a color - sextet with charge 4/3 is considered. The only allowed renormalizable couplings of S_{uu} to Standard Model fermions is with right handed up quarks (u_R). The partial decay width of $S_{uu} \to uu$ in Leading Order (LO) of the purturbation theory is given by:

$$\Gamma(S_{uu} \to uu) = \frac{y_{uu}^2}{32\pi} M_S \tag{56}$$

where y_{uu} is a dimensionless coupling and M_S is the diquark mass.

On the other hand, as far as its production $uu \to S_{uu}$ in the LHC is concerned, the LO cross section is given by the following formula.

$$\sigma(pp \to S_{uu}) = \frac{\pi}{6s}\ y_{uu}^2 \int_{M_S/s}^{1} \frac{dx}{x}\ f_u(x, M_S^2)\ f_u\left(\frac{M_S^2}{sx}, M_S^2\right) \tag{57}$$

where $\sqrt{s}$ is the center-of-mass energy and $f_u(x, Q^2)$ is the PDF of the up quark carrying momentum fraction x. The production of the antiparticle $S_{uu}^{\dagger}$ through the interaction $\overline{u}\,\overline{u} \to S_{uu}^{\dagger}$ has a much smaller cross section because the PDF of the anti - up quark is smaller than the respective of the up quark. Of course, as it was mentioned in the previous section, both the decay width and production of the scalar diquark need to be corrected with Next - to - Leading Order (NLO) terms.

Besides the scalar diquark, a vectorlike quark χ is also introduced in this model, that has the same gauge charges as the right handed up quark. The partial width of the S_{uu} decaying into a pair of these quarks in LO is given by:

$$\Gamma(S_{uu} \to \chi\chi) = \frac{y_\chi^2}{32\pi} M_S \left(1 - \frac{2m_\chi^2}{M_S^2}\right)\left(1 - \frac{4m_\chi^2}{M_S^2}\right)^{1/2} \tag{58}$$

where y_χ is a dimensionless coupling constant and $m_\chi < M_S/2$ is the mass of χ.

Regarding the decay of the vectorlike quark χ, the only allowed occasion is $\chi \to ug$, namely it produces a pair of an up quark and a gluon.

Based on the production and decay of both S_{uu} and χ that were described previously, the process $uu \to S_{uu} \to \chi\chi \to (ug)(ug)$ leads to four jets in the final state, and more specifically a resonant production of a pair of equal mass dijet resonances. The corresponding Feynman diagram for this process can be found in Fig. 6.

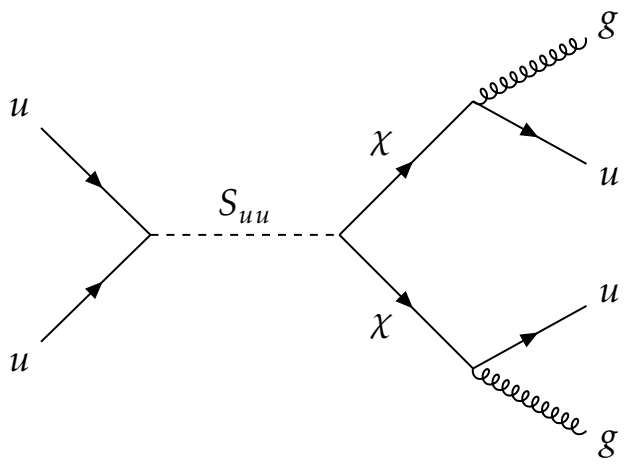

Figure 6: Feynman diagram for diquark production followed by decay into a pair of vectorlike quarks, each of them producing a gluon and an up quark.

The main background to the above signal is QCD production of four jets, where overall twenty Feynman contribute to the final state of two quarks and two gluons.

The CMS experiment has found an unusual event, reported in [6], that could be a possible candidate for an ultra heavy resonance decaying into a pair of massive particles. In Fig. 7 a three dimensional display of this event can be found. There are four jets reconstructed with the anti–k_T algorithm that will be described in section 3.2, that present a topology similar to the one expected of a diquark decaying into a pair of vectorlike quarks. More specifically, the four jets in the event have been grouped into two wide jets that have the same mass of 1.8 TeV and transverse momentum of 3.5 TeV and 3.4

TeV. The invariant mass of the dijet system that comprise these two wide jets has been measured to be 8 TeV.

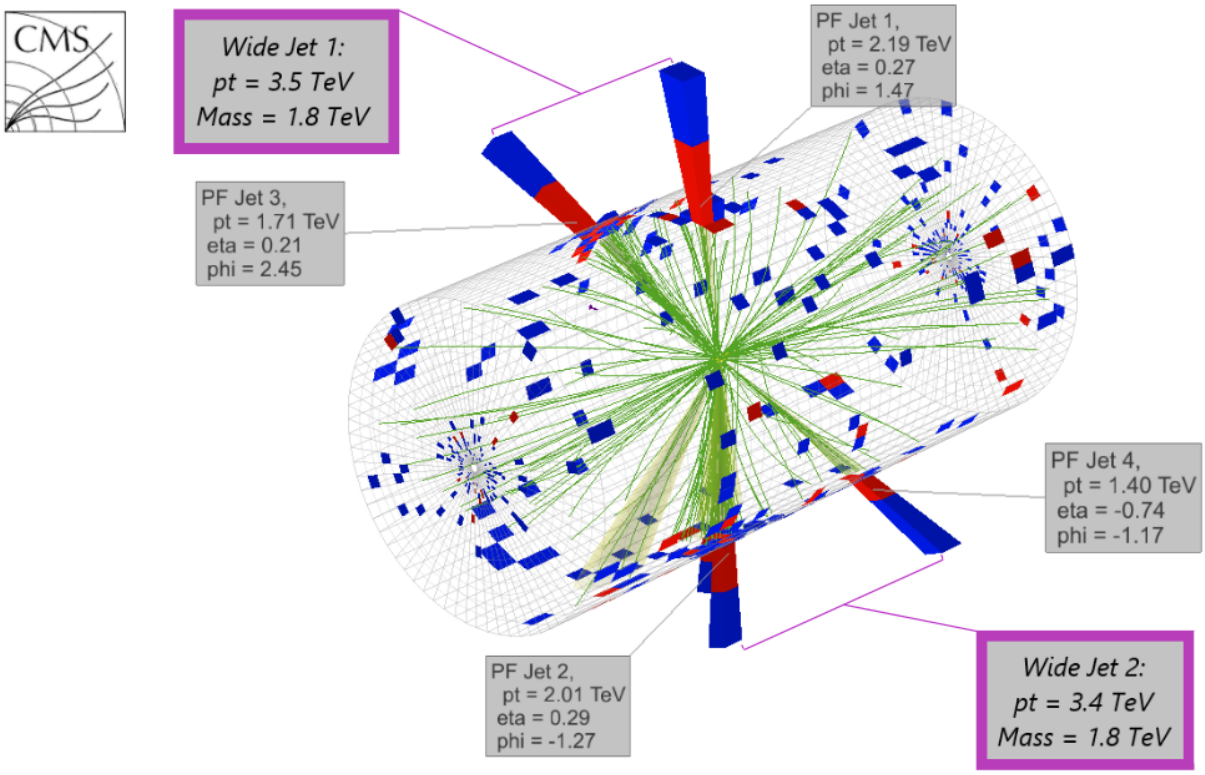

Figure 7: Three dimensional display of an event with invariant dijet mass of 8 TeV. With red color are the energy deposits in the electromagnetic calorimeter, with blue in the hadron calorimeter, with green are the reconstructed tracks of charged particles and with purple the grouping of the four jets into two wide jets.

Calculations that are presented in [5] show that the probability that such event is produced by QCD is 5×10^{-5}, motivating the search for a scalar diquark with $M_S = 8\ TeV$ and a vectorlike quark with $m_\chi = 1.8\ TeV$. Nevertheless, no firm conclusions can be derived unless more events in the high mass spectrum arise during the data collection in RunIII.

One of the most important aspect of physics analyses like the one that searches for ultra heavy resonances is to have the best possible accuracy in the measurement of the jet energy. The momentum and energy of jets as the ones in Fig. 7 need to be corrected after they have been measured, so as to account for any detector imperfections. This is the reason why the jet energy corrections that will be presented in this thesis are a necessity for all physics analyses that examine jets in the final state.

2 The CMS Experiment

The Standard Model has been thoroughly examined and proven to be consistent and accurate in the energy scale up to TeV. Therefore, it is crucial that it is studied for energies greater than 1 TeV, which is the main reason why the Large Hadron Collider (LHC) at CERN was built.

The LHC machine is the world's largest and most powerful particle accelerator, consisting of a 27-kilometer ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles. The LHC comprises thousand of magnets; 1232 dipole magnets 15 meters in length which bend the beams, and 392 quadrupole magnets, each 5–7 meters long, which focus the beams. The proton bunches are formed in the 26 GeV Proton Synchrotron (PS) and consequently are accelerated in the 450 GeV Super Proton Synchrotron (SPS) and then are transferred to the LHC. Inside the LHC the proton bunches are accelerated further up to the energy scales of a few TeV, and travelling at opposite directions are being forced to collide at 4 different locations, which correspond to where the CMS, ATLAS, Alice and LHCb are located.

Regarding the data collection for RunII, the LHC operated at center-of-mass energies of $\sqrt{s} = 13\ TeV$. With a bunch spacing of 25 ns, proton beams collide giving us a total of $\simeq$ 150 fb^{-1} of data for all three years (2016+2017+2018) combined, while the instantaneous luminosity in 2018 reached $\mathcal{L} = 2.06 \times 10^{34}\ cm^{-2}s^{-1}$.

One of the four detectors in the LHC, as mentioned above, is the Compact Muon Solenoid (CMS), which is a general purpose proton - proton detector. Most of the experiments that are carried out in CMS aim to detect traces of physics beyond the Standard Model, with around 60% of them using jets in final states.

In this chapter the coordinate system that is used in the CMS experiment will be presented and details regarding the CMS detector layout will be presented.

2.1 Coordinate System & Kinematics

The coordinate system that has been adopted by the CMS experiment, and has its origin at the collision point inside the detector, can be viewed in Fig. 8. The $z-axis$ is defined as the pp collision axis pointing towards the Jura mountains, while the $x-axis$ is set to point towards the center of the LHC machine and $y-axis$ points upwards so as to have a right-handed coordinate system.

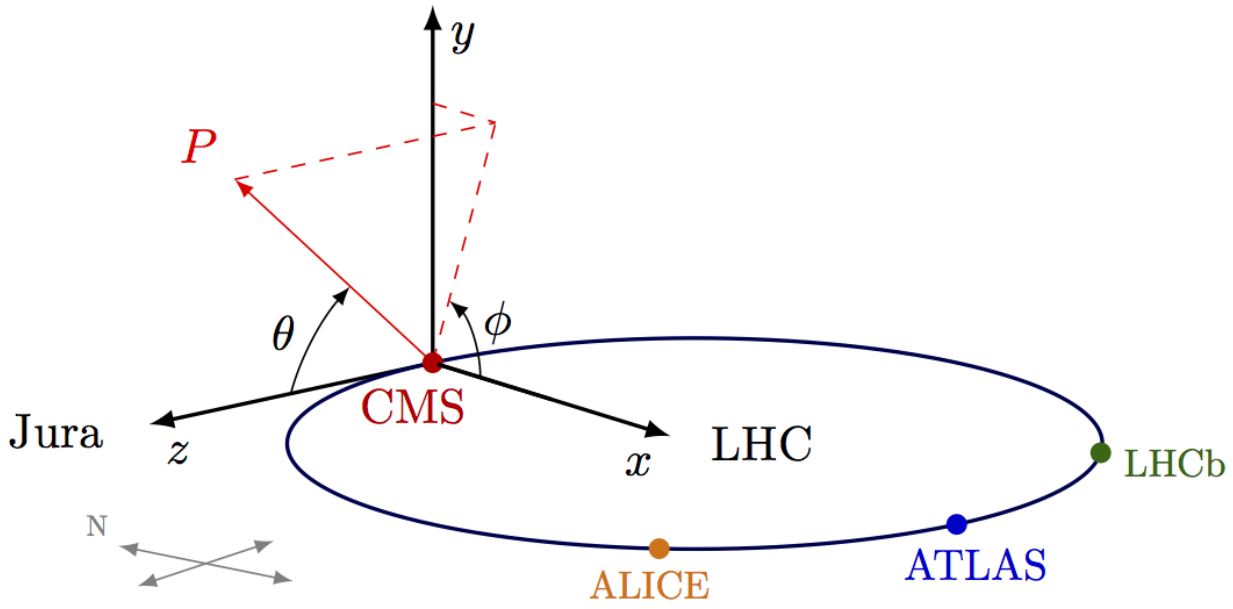

Figure 8: The coordinate system of CMS. [2]

Hence, the four vector of a particle that is produced in a pp collision and is measured in the CMS detector can be written as:

$$\mathcal{P}^{\mu} = (E, p_x, p_y, p_z) \tag{59}$$

where E is its energy and p_i is its momentum along i axis ($i = x, y, z$). Natural units are used, namely the speed of light c and the reduced Planck's constant $\hbar$ are set to 1.

Although the transverse components p_x and p_y are Lorentz invariant under a boost along the collision axis, the longitudinal ones E and p_z are not, thus making this description inconvenient. Therefore, it is very useful to define a four vector that has three Lorentz invariant transverse components under a boost transformation along $z-axis$ and one that is not. In more detail, the three invariant components are chosen to be the azimuth angle ϕ that is measured from the $x-axis$ in the $x-y$ plane, the transverse momentum p_T that lies in the $x-y$ plane, and the particle's mass m. These quantities are calculated as a function of the initial $\mathcal{P}^{\mu}$ components by the following formulas:

$$\begin{aligned} p_T &= |\vec{p}| \sin\theta \\ \phi &= \tan^{-1}\left(\frac{y}{x}\right) \\ m &= \sqrt{E^2 - |\vec{p}|^2} \end{aligned} \tag{60}$$

[2] Taken from https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates

The fourth component that will be considered is called rapidity and is defined as follows:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \tag{61}$$

As mentioned previously, this quantity is not invariant, so considering a boost of velocity β along the $z-axis$:

$$\begin{aligned} E^{'} &= \gamma(E+\beta p_z) \\ p_x^{'} &= p_x \\ p_y^{'} &= p_y \\ p_z^{'} &= \gamma(p_z+\beta E) \\ \gamma &= \left(1-\beta^2\right)^{-1/2} \end{aligned} \tag{62}$$

the rapidity will transform as follows:

$$y^{'} = \frac{1}{2}\ln\left(\frac{E^{'}+p_z^{'}}{E^{'}-p_z^{'}}\right) = \frac{1}{2}\ln\left(\frac{(E+p_z)(1+\beta)}{(E-p_z)(1-\beta)}\right) = y + \underbrace{\ln\gamma(1+\beta)}_{\text{constant}} \tag{63}$$

Furthermore, since $p_z = |\vec{p}|\cos\theta$ and $E = |\vec{p}|/\beta$, Eq. 61 can be rewritten as:

$$y = \frac{1}{2}\ln\left(\frac{1+\beta\cos\theta}{1-\beta\cos\theta}\right) \tag{64}$$

Apart from the rapidity, the pseudo rapidity η is also defined for massless particles with $\beta = 1$, namely $\eta = y(\theta, \beta = 1)$. As a result, pseudo rapidity can be calculated as:

$$\eta = \frac{1}{2}\ln\left(\frac{1+\cos\theta}{1-\cos\theta}\right) = \ln\left(\frac{\cos\theta/2}{\sin\theta/2}\right) = -\ln(\tan\theta/2) \tag{65}$$

For high energy particles with $E, p_T \gg m$ and $\beta \to 1$, such as the ones produced in the pp collisions at CMS, we obtain $y \simeq \eta$ and can safely replace rapidity with pseudo rapidity. By definition, the transverse plane is marked by $\eta = 0$, while $|\eta| \to +\infty$ signifies the collision axis. Practically, the value of pseudo rapidity in the experiments at CMS goes up to $|\eta| \lesssim 5$. In the end, the particle four vector that is being used is given by:

$$\mathcal{P}^{\mu} = (p_T, \eta, \phi, m) \tag{66}$$

Using these variables, the theoretical differential cross section not only for the $qq \to qq$ process (Eq. 55) but the overall $pp \to jets + X$ as well, can be easily converted so that it can be compared to the experimentally measured one. More specifically, assuming massless

jets in the final state that have $(p_T)_1 = (p_T)_2 \equiv p_T$, the four momenta of the colliding partons can be written as:

$$\begin{aligned} \mathcal{P}_1 &= \frac{\sqrt{s}}{2}(x_1, 0, 0, x_1) \\ \mathcal{P}_2 &= \frac{\sqrt{s}}{2}(x_2, 0, 0, x_2) \end{aligned} \tag{67}$$

and the conservation of momentum and energy implies that:

$$\begin{aligned} x_1 &= \frac{p_T}{\sqrt{s}}\left(e^{y_3} + e^{y_4}\right) \\ x_2 &= \frac{p_T}{\sqrt{s}}\left(e^{-y_3} + e^{-y_4}\right) \\ Q^2 &= p_T^2(1 + e^{y_4 - y_3}) \end{aligned} \tag{68}$$

where y_3, y_4 are the rapiditites of the final state jets. Therefore, the Jacobian of the transformation becomes:

$$\frac{\partial(y_3, y_4, p_T^2)}{\partial(x_1, x_2, Q^2)} = \frac{1}{x_1 x_2} \tag{69}$$

and the theoretical differential cross section can be converted as:

$$\frac{d^3\sigma}{dQ^2 dx_1 dx_2} \rightarrow \frac{d^3\sigma}{dp_T^2 dy_3 dy_4} \tag{70}$$

A final remark that needs to be addressed in this section is the significance of the transverse plane, since there the momentum conservation principle can be applied. Knowing that the overall momentum before the pp collision takes place is zero, then it is obvious that if $\vec{p}_T^{\,i}$ is the measured transverse momenta of each particle i, then it should be that:

$$\sum_i \vec{p}_T^{\,i} = 0 \tag{71}$$

If that is not the case then the overall missing momentum $\vec{p}_T^{\,miss}$ and consequently energy E_T^{miss} can be calculated, which would suggest flaws in the detector system or presence of neutrinos or even exotic particles beyond the Standard Model that do not interact with the detectors and therefore their energy were not measured. The same principle cannot be applied along the $z - axis$ since the momentum fractions that each parton takes from the initial colliding protons are not known a priori.

2.2 The CMS Detector

The overall layout of the CMS detector, that has a length of 28.7 m, a diameter of 15 m and a total weight of 12500 tons, is shown in Fig. 9.

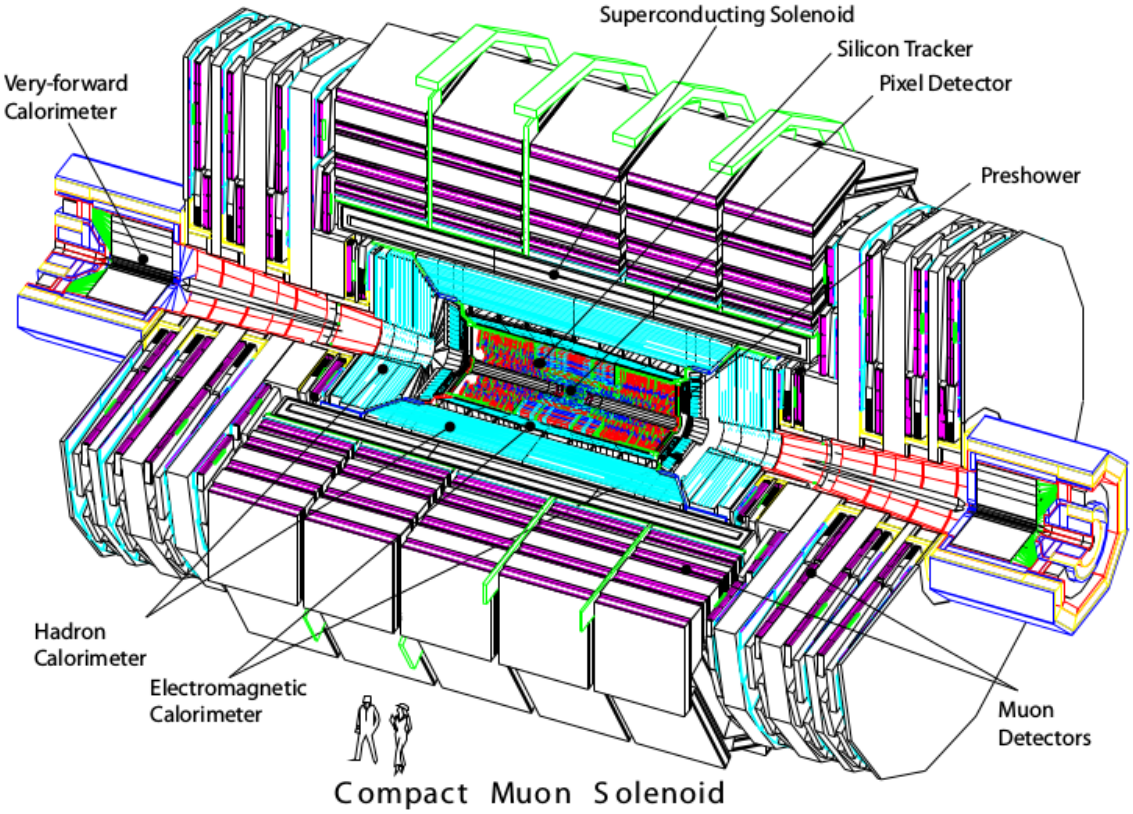

Figure 9: The CMS detector. [3]

At the heart of the detector lies a 13 m long and of 5.9 m inner diameter superconducting solenoid that has a nominal magnetic field of 4 T. This very high magnetic field allows 4 different muon detectors to be installed so as to provide both robustness and full geometric coverage. Furthermore, the bore of the magnetic coil is large enough so that it accommodates a silicon tracker along with layers of pixel detectors and additionally the calorimeters; an electromagnetic and a hadron one. Finally, in the very forward regions, along the collision axis an iron/quartz-fibre calorimeter is installed. Each of these parts of the detector will be briefly discussed in the subsequent sections.

2.2.1 Magnet

The magnet, as mentioned previously, is a large solenoid that contains the inner tracking system and the calorimeters. Its purpose is to bend the trajectory of charged particles

[3] Taken from Chapter 1.5 of [7] (Figure 1.2)

that are produced in the proton - proton collisions. Especially for muons, a powerful magnet is needed to bend their paths, which is the reason why the magnet operates with a large magnetic field of 3.8 T, has an electric current of around 18.5 kA and the stored energy in the magnetic field is around 2.3 GJ.

There are two main reasons why bending the trajectories of the charged particles is very important. Firstly, positively and negatively charged particles bend in opposite directions in the same magnetic field, thus helping us identify the charge of each measured particle. Furthermore, by measuring the radius of the circular motion of the particle due to the magnetic field its momentum can be calculated. More specifically, since the magnetic field lies on the $z-axis$ and supposing that a particle's momentum in the transverse $(x-y)$ plane is denoted as p_T, then the magnetic rigidity is defined as follows:

$$\frac{p_T}{q} = B \cdot R \qquad (SI) \tag{72}$$

where $q = |Q| \cdot e$, with $Q = \pm 1, \pm 2, ...$ is the charge of the particle, B is the magnetic field and R is the bending radius. Using that $[T] = V \cdot s \cdot m^{-2}$, the above relation can be converted appropriately so as to measure the particle's momentum in GeV/c:

$$p_T[GeV/c] = 0.3 \cdot |Q| \cdot B[T] \cdot R[m] \tag{73}$$

It is also worth remarking that for a specified magnetic field high momentum particles bend less than low momentum ones, which is why it is crucial to have a very strong magnet in order to bend even the most energetic particles. Assuming high precision measurement of the bending radius, the powerful magnet of CMS allows high accuracy in the measurement of the momentum of even high energy particles.

Inside the magnet the tracker and the calorimeters fit, while the muon detectors are interleaved with a 12-sided iron structure that surrounds the magnet coils and contains and guides the field. Made up of three layers this return yoke reaches out 14 metres in diameter and also acts as a filter, allowing through only muons and weakly interacting particles such as neutrinos. The enormous magnet also provides most of the experiment's structural support.

2.2.2 Inner Tracking System

As described in the previous section, in order to measure the transverse momentum of a charged particle, bending its trajectory with a magnetic field is not enough, since an accurate reconstruction of its path needs to be also performed so that the bending radius is measured as accurately as possible.

For that reason, the first layer of the CMS detector consists of a silicon tracker that provides this ability to reconstruct the paths of high energy muons, electrons and charged hadrons as well as detect tracks coming from the decay of very short-lived particles such as b-quarks. The tracking system needs to also be lightweight so as to disturb particles as little as possible, which is something that is achieved by taking position measurements so accurate that tracks can be reliably reconstructed using just a few measurement points, where each measurement is accurate to 10 μm. Furthermore, since the tracker is the innermost layer of the detector it receives the largest amount of particles, so its construction materials are carefully chosen to resist radiation.

The inner tracking system is entirely made of silicon and has a cylindrical shape with the outer radius extending to nearly 110 cm and its total length being approximately 540 cm. Closest to the interaction vertex where the particle flux is the highest pixel detectors are placed, with the size of each pixel being $\approx 100 \times 150\ \mu m^2$. In the intermediate region, at radius of 20 $cm < r < 55\ cm$ the particle flux is low enough to enable use of silicon microstrip detectors with a minimum cell size of 10 $cm \times 80\ \mu m$. In the outermost region of 55 $cm < r < 110\ cm$ the particle flux has dropped sufficiently to allow use of larger-pitch silicon microstrips with a maximum cell size of 25 $cm \times 180\ \mu m$. In the barrel part of the tracker ($|\eta| < 1.2$), which is separated into an Inner and an Outer Barrel, there are 3 layers of hybrid pixel detectors close to the interaction vertex while the forward region has 2 pixel and 9 microstrip layers in each of the 2 Endcaps ($0.9 < |\eta| < 2.4$). In order to avoid excessively shallow track crossing angles, the Inner Barrel is shorter than the Outer Barrel, and there are additional 3 Inner Disks in the transition region between the barrel and endcap parts, on each side of the Inner Barrel. The total area of the pixel detector is $\approx$ 1 m^2, whilst that of the silicon strip detectors is 200 m^2, providing coverage up to $|\eta| < 2.5$. The inner tracker comprises overall 66 million pixels and 9.6 million silicon strips.

As far as the part of the tracking system with the microstrip detectors is concerned, the Silicon Strip Tracker (SST) readout system is based on a front-end APV25 readout chip, analogue optical links and an off-detector Front-End Driver (FED) processing board. The APV25 chip samples, amplifies, buffers and processes signals from 128 channels of a silicon strip sensor. Some of these APV25 chips were destroyed throughout the data collection during 2016, leading to the "APV issue" that affected these datasets.

2.2.3 Electromagnetic Calorimeter (ECAL)

The purpose of the Electromagnetic Calorimeter (ECAL) is to accurately measure the energy of electrons, positrons and photons through the electromagnetic showers they produce when these particles interact with the calorimeter.

More specifically, when one of these particles of sufficiently high energy (greater than $\approx 1\ GeV$) enter the ECAL they interact with the material via the Bremsstrahlung radiation or the pair production, depending on the particle. Electrons and positrons lose energy through the Bremsstrahlung process by emitting a photon, while photons interact with matter via pair production, thus creating an electron and positron. The combination of these alternating processes result in a geometrical increase of such particles inside the calorimeter, as a function of its depth, that keeps feeding the shower until a certain threshold is reached when the particles are no longer energetic enough. This happens when energy losses of electrons and positrons other than Bremsstrahlung start to dominate, or when photons can no longer produce an electron-positron pair, which happens for photon energies $E_\gamma < 2m_e = 1022\ MeV$.

The main features of the electromagnetic showers can be described in terms of the radiation length X_0 which represents the average distance x that an electron needs to travel in a material to reduce its energy to $1/e$ of its original energy E_0 (Eq. 74). The radiation length also represents the 7/9 of the mean distance that a photon beam travels in the material in order for its intensity to be reduced to $1/e$ of the original intensity (Eq. 75).

$$\langle E(x)\rangle = E_0 e^{-\frac{x}{X_0}} \tag{74}$$

$$\langle I(x)\rangle = I_0 e^{-\frac{7}{9}\frac{x}{X_0}} \tag{75}$$

The radiation length depends only on the material the calorimeter is made of, where the showers take place.

Assuming a simple cascade model where each particle travels a distance of $\approx X_0$ before interacting, then at depth $x = nX_0$ ($n = 1, 2, ...$) inside the calorimeter the shower will contain 2^n particles with their average energy being $E(n) = E_0/2^n$, where E_0 is the energy of the initial particle. Introducing the critical energy E_C, for which the cascade process stops abruptly due to the particles not being energetic enough to continue with the Bremsstrahlung and pair production processes, the shower reaches its maximum depth for $E(n_{max}) = E_C$. As a consequence, the maximum depth of the cascade can be easily related to the initial energy of the particle as follows:

$$\begin{gathered} E_C = \frac{E_0}{2^{n_{max}}} \Rightarrow n_{max} = \frac{\ln\left(E_0/E_C\right)}{\ln 2} \Rightarrow \\ x_{max} = X_0 \frac{\ln\left(E_0/E_C\right)}{\ln 2} \quad \text{or} \quad E_0 = 2E_C e^{\frac{x_{max}}{X_0}} \end{gathered} \tag{76}$$

Since both the critical energy and the radiation length depend only on the material, the initial energy of the electron, positron or photon can be calculated by measuring the depth of the shower they create inside the ECAL, using the above formula.

Furthermore, as far as the transverse size of an electromagnetic shower is concerned, it can be considered that the entire shower is contained within a cylinder of radius $R = 2\rho_M$, where ρ_M is called Moliere radius and depends only on the material.

The electromagnetic calorimeter of the CMS detector is made of lead tungstate ($PbWO_4$) scintillating crystals that have short radiation ($X_0 = 0.89\ cm$) and Moliere ($\rho_M = 2.2\ cm$) lengths, thus providing compact EM showers. These crystals are also fast (80% of the light is emitted within 25 ns) and radiation hard (up to 10 $Mrad$).

The barrel section of the ECAL has an inner radius of 129 cm, contains 61200 crystals and covers the pseudo rapidity area of $|\eta| < 1.479$. Each crystal has a front face cross section of $\approx 22 \times 22\ mm^2$ and a length of 230 mm, corresponding to $25.8X_0$, so that the leakage of the electromagnetic showers at the end of the ECAL is as minor as possible. Additionaly, the ECAL also comprises 2 endcaps at a distance of 314 cm from the vertex and covering a pseudo rapidity range of $1.479 < |\eta| < 3.0$. In each endcap there are 7324 crystals that have an area of $28.6 \times 28.6\ mm^2$ and a length of 220 mm which corresponds to $24.7X_0$.

For additional spatial precision the CMS ECAL also contains a preshower detector that sits in front of the endcaps. The active elements of this device are 2 planes of silicon strip detectors, with a pitch of 1.9 mm, which lie behind disks of lead absorber at depths of 2 X_0 and 3 X_0. The aim of the preshower detector is to measure and distinguish between the low energy photos that are produced from neutral pions through the decay $\pi^0 \to \gamma\gamma$, and single high energy photons.

Moreover, one very important aspect of this ECAL is the fact that it is a hermetic, homogeneous calorimeter, where the shower developing medium is itself active, namely the material both produces the shower and measures the energy deposited. Therefore, the ECAL has an excellent energy resolution, compared to sampling calorimeters, which will be discussed in the next section. In more detail, the energy resolution of the ECAL can be parameterized as a function of the energy as:

$$\frac{\sigma(E)}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \tag{77}$$

where the first term is the "stochastic term" that includes the statistical fluctuations of the shower detection, the second is the "noise term" that is related to electronic noise and decreases as the energy increases, and the last one is the "constant term" that does not depend on the energy of the particles and relates to detector nonuniformities, imperfections

in the detector mechanical structure and readout system, temperature gradients, detector aging, radiation damage and so on.

Using test beams with electrons the energy resolution of the CMS ECAL has been measured as:

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E/GeV}} \oplus \frac{12\%}{E/GeV} \oplus 0.3\% \tag{78}$$

The stochastic term is indeed small for this homogeneous calorimeter; approximately 3% for particles with 1 GeV energy, and decreases even more for higher energies. At ≈ 50 GeV both the stochastic and electronic noise terms become negligible and the constant term dominates the energy resolution.

2.2.4 Hadron Calorimeter (HCAL)

The hadron calorimeter aims to measure the energy of charged and neutral hadrons, namely particles that are made of quarks and gluons, like protons, neutrons, pions and kaons, based on the hadron showers they produce inside the HCAL.

The hadronic cascades differ significantly from the electromagnetic ones, since the energy degradation of hadrons inside the HCAL proceeds through an increasing number of mostly strong interactions with the calorimeter material, meaning that the material is itself active and participates in the shower. More specifically, as the incoming hadrons interact with the calorimeter, energetic secondary particles are produced between interactions, whose momenta is a fair fraction of the primary hadron momentum. Additionally, in hadronic collisions with the material nuclei, a significant part of the primary energy is consumed in nuclear processes such as excitation, nucleon evaporation and spallation. These processes result in the creation of a hadron shower that is much broader than an electromagnetic one and extends deeper in the calorimeter. The incoming particles hits target nucleus in a hard scatter and generates a shower of hadrons that mostly contains pions ($\pi^{\pm}, \pi^0$) and kaons. Since neutral pions quickly decay into a pair of photons, inside the hadron showers there is also an electromagnetic component. As far as the hadronic component is concerned, the secondary particles keep multiplying, with the particle multiplicity scaling logarithmically as a function of the initial particle energy, until a certain threshold is reached. Since the lighter hadrons are pions, this threshold is reached when the secondary particles can no longer decay into pions, which happens for energies $E \lesssim 2m_\pi = 0.28\ GeV$.

Similarly to the radiation length in the electromagnetic showers, the hadron ones have a characteristic quantity λ_I called nuclear interaction length, which represents how deep

inside the material a particle can travel before interacting strongly and produce secondary particles. This length depends on the material that the shower takes place and for a specific material is much larger than the respective radiation length, namely $X_0/\lambda_I \ll 1$ (e.g. $X_0/\lambda_I \approx 0.033$ inside lead). For that reason hadron showers extend much deeper than the electromagnetic ones, which is why the HCAL is built in the most outer layers inside the magnet, further away from the interaction vertex. Therefore, the electromagnetic showers can be distinguished from the hadron ones since by the end of the ECAL they have already died off, while hadron showers have not yet deposited their energy. Finally, regarding the transverse plane, the hadron showers typically extend in it as far as one nuclear interaction length goes.

The HCAL of the CMS detector, contrary to the homogeneous ECAL, is a sampling (or heterogeneous) calorimeter, namely it has two alternating layers of a material that produces the particle shower (passive absorber) and another material that measures the deposited energy (active detector). The absorber material of the CMS HCAL is chosen to be brass since it has a reasonably short interaction length, while the active detector consists of plastic scintillator tiles. This way the HCAL is built to be compact, since a homogeneous HCAL would need to have a large depth in order to contain the hadron showers without leakages. The disadvantage, however, of this sampling calorimeter is that it has a much poorer energy resolution than the ECAL.

The hadron barrel part of the CMS HCAL consists of 32 towers covering the pseudo rapidity region $|\eta| < 1.4$, resulting in 2304 towers with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. There are 15 brass plates, each with a thickness of about 5 cm, plus 2 external stainless steel plates for mechanical strength. The hadron outer detector contains scintillators with a thickness of 10 mm and covers the region $|\eta| < 1.26$. Furthermore, the CMS HCAL comprises two endcaps, each consisting of 14 η towers with 5° ϕ segmentation, covering the pseudo rapidity region $1.3 < |\eta| < 3.0$. In order to ensure full geometric coverage, a forward calorimeter made of steel and quartz fiber is installed as well, in the most forward region $3.0 < |\eta| < 5.2$.

As far as the energy resolution is concerned, there are two main factors that degrade the HCAL performance compared to the ECAL. The first one is related to the intrinsic fluctuations of the hadron showers which are much higher than the electromagnetic ones. A significant part of the energy is spent on the binding energy of the material nuclei, which is eventually missed from the measurement since it will be deposited back in the calorimeter at a later time, when the nuclei de-excite and emit low energy neutrons, photons or other particles. Thus, this energy can not be matched with the particular particle that is measured and is lost. In addition, the resolution becomes worse due to the perfor-

mance of the HCAL itself, as it is a sampling calorimeter. Since, there are two alternating layers of materials, the sampling fluctuations are mainly caused by energy deposits in the wrong material. Besides that, there are also significant energy leakages from the rear of the calorimeter, as well as from the outer layer, since the ECAL+HCAL system reach $\approx 7\lambda_I$ in depth in the central η region of the barrel, while around $10\lambda_I$ are needed for 99% longitudinal containment.

Using test beams for single pions, the combined resolution of the CMS ECAL+HCAL system is:

$$\frac{\sigma(E)}{E} = \frac{110\%}{\sqrt{E/GeV}} \oplus 9\% \tag{79}$$

The resolution is indeed significantly worse compared to Eq. 78, making the energy measurement of the hadron showers less accurate than the electromagnetic ones.

2.2.5 Muon System

Muons are leptons like electrons and positrons, but 200 times heavier. As a consequence, they are expected to penetrate several meters of iron without interacting. Since none of the calorimeters of CMS can stop them, muon chambers have been placed at the very edge of the experiment, where they are the only particles likely to register a signal as all the other particles have already deposited their energy.

A muon particle is measured by fitting a curve to hits among the four muon stations, which sit outside the magnet coil and are interleaved with iron "return yoke" plates. By tracking its position through the multiple layers of each station, combined with tracker measurements the detectors precisely trace the particle's path. Therefore, its momentum can be also calculated by the bending radius using formula 73 of section 2.2.1. The field of the CMS magnet, as already discussed, is very powerful so that it provides the ability to bend even the paths of very high-energy muons.

In total there are 1400 muon chambers consisting of three types of gaseous sub detectors to identify and measure muons; 250 drift tubes (DT) and 540 cathode strip chambers (CSC) track the particles' positions and provide a trigger, while 610 resistive plate chambers (RPC) form a redundant trigger system, which quickly decides to keep the acquired muon data or not. The drift tubes are located at the barrel region, while the cathode strip chambers at the two endcaps. The resistive plate chambers can be found both at the barrel and the endcaps.

3 Jets in CMS

3.1 The Particle Flow Reconstruction

Outlining the previous chapter, the silicon tracker of the CMS detector provides the ability to accurately measure the tracks and momenta of charged particles, exploiting the large magnetic field of the solenoid, while the calorimeters measure the energy of particles; the ECAL with an excellent resolution, while the HCAL with a much worse one. As a consequence, the energy resolution of jets that comprise mostly hadrons is greatly degraded due to the poor performance of the hadron calorimeter.

In more detail, measurements on jet fragmentation have shown that on average 65% of the jet energy is carried by charged hadrons, 25% by photons and 10% by neutral hadrons. Therefore, using purely calorimetric measurements roughly 75% of the jet energy measurement is affected by the HCAL's resolution. Instead, a significantly improved event description can be achieved by correlating the basic elements from all detector layers in order to identify the final state particles and then combining the corresponding measurements to reconstruct the particle properties on the basis of this identification. This approach is called Particle Flow (PF) reconstruction. With this method the four vector of each visible particle is reconstructed and afterwards with the proper clustering the jet energy is calculated as the sum of the energies of the individual particles.

With this approach the charged hadrons are identified by simultaneously taking into consideration the tracks they leave while passing through the silicon tracker, one or more calorimeter clusters they produce in the HCAL and the absence of signal in the muon detectors. The combination of these measurements in the tracker and in the calorimeters provides an improved determination of the energy and direction of each charged hadron, dominated by the superior tracker resolution in that particular event. As far as the identification of photons and neutral hadrons is concerned, it is being performed by examining ECAL and HCAL clusters respectively, with no track signal. Then, their energy is measured via the ECAL and HCAL deposits. As a result, it becomes clear that since the momentum of charged hadrons are measured from the tracker, on average only 10% of the jet energy is being extracted from hadron calorimeter measurements, thus improving the overall accuracy.

The Particle Flow concept was developed by the ALEPH experiment at LEP and at first it was feared that it would be impossible to implement for high energy proton - proton collisions due to the innumerable particles that are produced in such interactions and also particles coming from secondary vertices that do not relate to the hard scatter. However, studies with Detailed Monte Carlo simulations demonstrated that the CMS detector

is adequate for PF reconstruction, due to its large magnetic field, fine-grained tracker, highly-segmented ECAL, hermetic HCAL with a coarse segmentation and excellent muon tracking system.

Although the PF reconstruction provides the most accurate way so far to measure the jet energy and its geometrical characteristics, a perfect level of performance cannot be achieved mainly due to the confusion term that arises in the energy resolution. This term is related to the associations of energy deposits with the correct particles not being perfect. This can lead for example to photon energy not being accounted for, when the calorimeter hits from a photon are not resolved from a charged hadron shower. Another aspect of the confusion term, on the other hand, could be the energy of some particles being double counted, when part of a charged hadron shower is identified as a separate cluster although it is already accounted for by the track momentum.

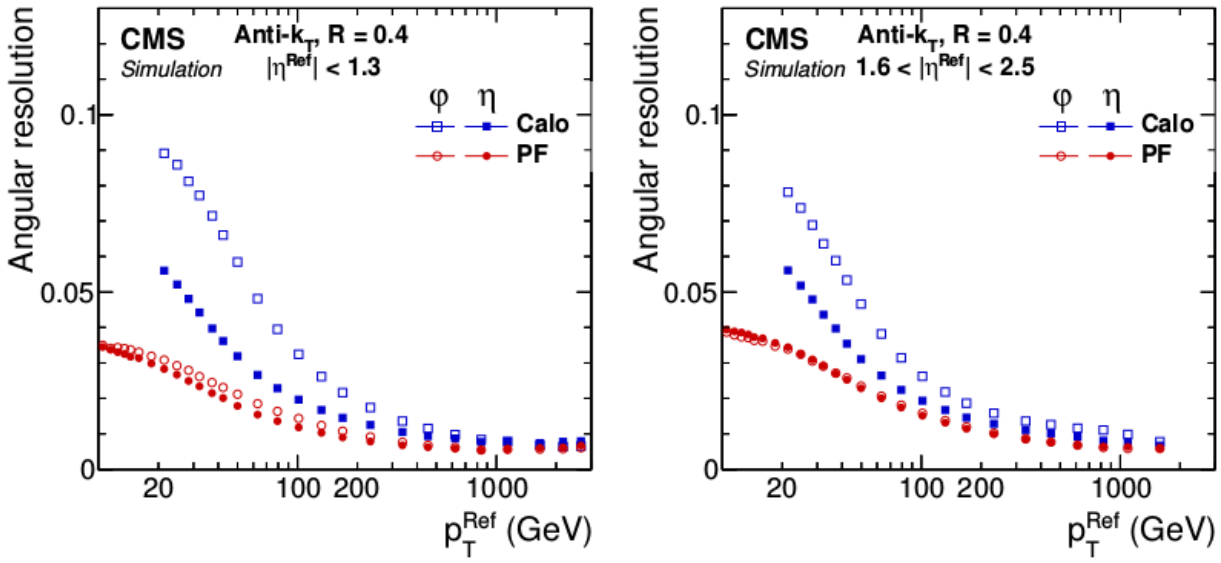

Figure 10: Jet angular resolution in the barrel region (left) and endcaps (right), as a function of the p_T of the Ref jet. The ϕ resolution is expressed in radians. [4]

The jet performance with respect to the Particle Flow algorithm has been quantified with a sample of QCD multijet events, as was more thoroughly studied and presented in [11]. Making use of the anti–k_T jet clustering algorithm which will be briefly explained in the next section, all measured particles can be reconstructed by the PF algorithm (PF jets) or the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo jets). Both of these categories of jets can be compared with the stable particles produced by the event generator in simulation (Ref jets). PF jets can be studied in the p_T region

[4] Taken from Chapter 5.1 of [11] (Figure 10)

down to 15 GeV, while Calo jets down to 20 GeV. For momenta lower than these values jets are considered to be unreliable and are rejected by the clustering algorithm.

In order to compare the reconstruction efficiency between PF and Calo jets, each one of these jets are being matched to the closest Ref jet in the (η, ϕ) plane with a $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.1$ and $\Delta R < 0.2$ criterion for PF and Calo jets respectively.

Afterwards, the angular and energy resolution can be examined and compared between these reconstructions. The angular resolution is defined as the gaussian width of the η or ϕ ratio between PF or Calo jets and Ref jets and is shown in Fig. 10 for the barrel region (left) and the endcaps (right). The angular resolution for PF jets is greatly improved mainly because of the precise determination of charged hadron directions and momenta, since the PF reconstructions combine information from all detector sub - systems. Moreover, with respect to the ϕ resolution, it is significantly degraded for Calo jets since the energy deposits of charged hadrons in the HCAL are spread along the ϕ direction by the magnetic field, thus increasing the resolution, especially at lower jet momenta, where the bending is more intense.

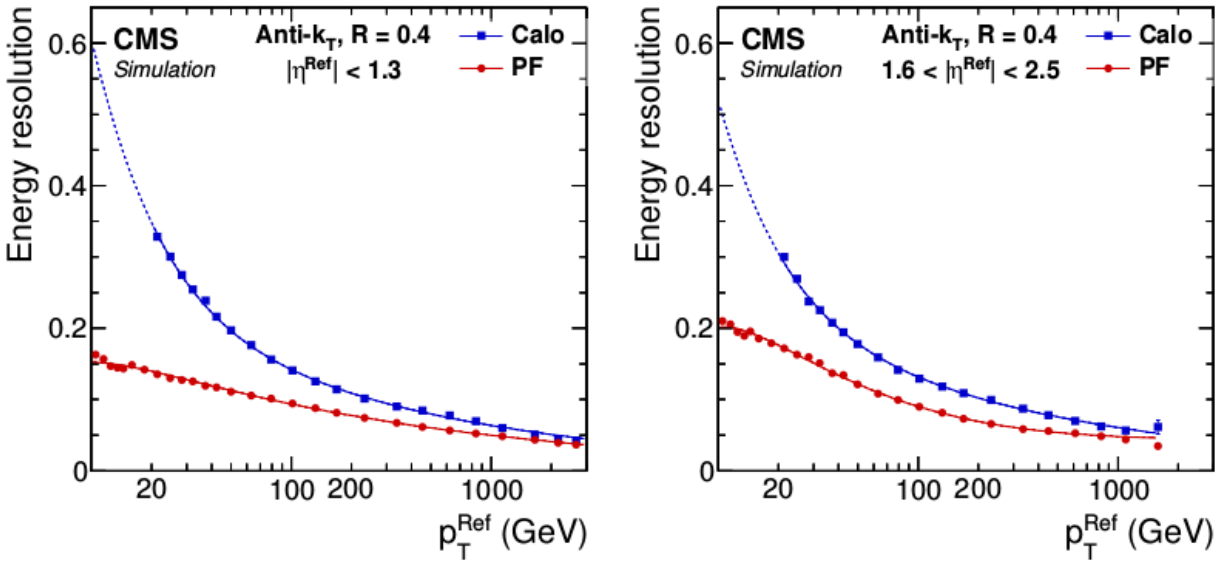

Figure 11: Jet energy resolution in the barrel region (left) and endcaps (right), as a function of the p_T of the Ref jet. [5]

Studies regarding the jet energy resolution have also been performed, after the application of corrections to both PF and Calo jets. This energy resolution, which is calculated as the gaussian width of the ratio between the corrected (PF or Calo) and Ref jets, is shown in Fig. 11 as a function of the p_T of Ref jets. It is obvious that the resolution for Calo jets

[5] Taken from Chapter 5.1 of [11] (Figure 13)

is significantly worse, especially as the jet momentum decreases. This is caused by the greatly degraded performance of the HCAL resolution (Eq. 79) when measuring hadron showers, that is already discussed in section 2.2.4. On the other hand, PF jets exploit the excellent accuracy of the tracker in order to measure the energy of the charged particles, therefore providing a much better overall jet energy resolution. For the highest jet energies the performance of Calo jets becomes almost identical to PF jets as the stochastic term of 79 becomes negligible.

For all the above reasons, PF jets are mostly used in physics analyses, so they will be the main focus of this thesis, regarding the extraction of their energy corrections.

3.2 The Anti–k_T Jet Clustering Algorithm

Using the Particle Flow approach the trajectory and energies of the produced particles in the proton - proton collisions can be measured with the highest possible accuracy. Consequently, these particles are clustered properly into jets (PF jets), using one of the jet clustering algorithms available. The hadron jets that will be studied in this thesis in Chapter 4 have been clustered with the anti–k_T algorithm, for which more detailed information than in this section can be found in [13].

The main idea behind the anti–k_T algorithm is that a distance d_{ij} between entities (like particles or pseudo jets) i and j, and d_{iB} between entity i and the beam are introduced. The clustering algorithm then proceeds with calculating these distances and identifying the smallest one. If the minimum is a d_{ij} distance then entities i and j are recombined into one, while if it is a d_{iB} the object i is considered as a jet and is removed from the list of entities. In any case, the distances are calculated again until no entities are left. These distances are generally described by the following formulas:

$$\begin{aligned} d_{ij} &= \min\Big((k_T)_i^{2p}, (k_T)_j^{2p}\Big)\frac{\Delta_{ij}^2}{R^2} \\ d_{iB} &= (k_T)_i^{2p} \end{aligned} \tag{80}$$

where $(k_T)_i$ is the transverse momentum of particle i, R is the radius parameter of the jets that we want to cluster and $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with y_i and ϕ_i being the rapidity and azimuth angle of particle i. The parameter p, that represents the relative power of the energy (k_T) versus the geometrical features (Δ_{ij}), is equal to -1 for the case of the anti–k_T algorithm. For $p = 1$ one can recover the inclusive k_T algorithm while for $p = 0$ the Cambridge/Aachen algorithm.

The radius parameter R that has been used for clustering the jets that will be examined in this thesis is $R = 0.4$ and $R = 0.8$. The former category of jets are called AK4 jets, while

the latter AK8. It is obvious that AK8 jets have a larger cone, therefore more particles are clustered within them, leading to larger jet energy. Usually, AK8 jets are used in high jet p_T physics analyses.

3.3 CHS and PUPPI Jets

One of the main challenges of the data collection of RunII is the high instantaneous luminosity which results in a large number of additional proton - proton collisions in each event that do not come from the primary hard interaction. This contribution is called pileup, and will be thoroughly discussed in Chapter 4.3. It is also important to state that the expected increase of the center-of-mass energy at $\sqrt{s} = 14\ TeV$ in RunIII will come at a cost, namely the pileup conditions will be even more evident.

For that reason, the accurate reconstruction of jet properties and shapes becomes very demanding. In Chapter 4.3 the method to remove this pileup contribution, as a part of the baseline jet energy corrections, will be extensively studied. In this section, two categories of jet collections will be presented that aim to treat the pileup energy while the jet reconstruction and clustering takes place, namely before any derivation of corrections.

The first technique that will be presented is the Charged Hadron Subtraction (CHS) which aims to reduce the effect of the in - time pileup (IT PU) that is related to additional proton - proton collisions taking place within the same bunch crossing. With this method, charged hadrons that are unambiguously associated to pileup vertices are removed from the event so that only the rest of the PF candidates participate on the jet clustering with the anti–k_T algorithm. As already discussed, charged hadrons are identified by the PF algorithm as a track, in combination with HCAL and ECAL hits. The CHS technique is relevant to the pseudo rapidity region of $|\eta| < 2.5$ where the tracker coverage extends. In each event the leading primary vertex (PV) of the hard scatter is selected by taking into account the sum of squares of the track transverse momenta ($\sum |p_T^{Track}|^2$), and then the subleading vertices are classified as pileup vertices. These subleading vertices, however, are required consequently to pass further quality criteria in order to be considered as good pileup vertices, that are related to the chi-square per degree of freedom χ^2/ndf of the track to a proto - vertex reconstructed without it. Finally, if the track from a charged hadron is associated to a good pileup vertex, it is considered as a pileup track, and is removed in the CHS procedure, while all other tracks, including those not associated to any PV, are kept.

A second algorithmic approach used to mitigate the pileup contribution is called Pile Up Per Particle Identification (PUPPI). This algorithm combines global information from

the event, such as the event - wide pileup density, and vertex information from charged tracks, with local information to identify pileup at the particle level. In more detail, a shape α is calculated for each particle, that attempts to locally distinguish parton - like radiation from pileup - like radiation by exploiting the differences in the p_T spectrum, since for the pileup particles it falls much faster. This quantity is calculated for the i-th particle by the following formula:

$$\alpha_i = \log \sum_{j \in \text{event}} \frac{(p_T)_j}{\Delta R_{ij}} \times \Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij}) \tag{81}$$

where Θ is the Heaviside step function, ΔR_{ij} is the distance between i and j in the $\eta - \phi$ plane and $(p_T)_j$ is the transverse momentum of particle j. Parameter R_0 defines a cone around each particle i, so that only particles within the cone enter the calculation of α_i. In addition, particles closer to i than R_{min} are discarded from the sum, with R_{min} effectively serving as a regulator for collinear splittings of particle i. More information regarding this shape can be found in [15], where $R_0 = 0.3$ and $R_{min} = 0.02$ have been chosen.

Making use of the fact that in the tracker coverage a charged particle can be identified as coming from a primary or a pileup vertex by combining tracks with calorimeter hits, the median and the RMS of the α values can be used for charged pileup as an event - level characterization of the pileup distribution. Afterwards a weight is assigned to each individual particle by comparing its α value to the median of the charged pileup distribution. Particles that have α values within a few standard deviations from the median are mostly pileup particles and are assigned small weights. If on the other hand the α value of a particle deviates far from the median of the pileup distribution then it gets a large weight. Under ideal circumstances the weight would be zero for pileup particles and one for particles coming from the hard scatter. Finally, these weights are applied in order to rescale the particles' four momentum and particles with small weights or very small rescaled p_T are removed from the event. This procedure leads to pileup corrected events where no further correction needs to be applied, as it is the case for CHS jets.

These two algorithms are the most commonly used in CMS analyses, making the examination of CHS and PUPPI jets very important. More information for the CHS and PUPPI techniques can be found in [14] and [15] respectively. Since, as mentioned in section 3.2, the radius parameter when clustering jets is usually $R = 0.4$ or $R = 0.8$, there are overall 4 different jet collections to be studied; AK4 and AK8 PF+CHS and PF+PUPPI jets. In Chapter 4 the baseline jet energy corrections for these jet categories will be presented.

4 Jet Energy Corrections

4.1 Introduction

It is widely known that there are numerous physics analyses that study jets in the final state in order to detect traces of particles beyond the Standard Model. Such analyses require high precision and accuracy in the measurement of energy and transverse momentum. However, due to imperfect detector conditions it is almost impossible to be able to reconstruct and measure the true energy of jets, since there is a plethora of factors that lead to loss of information, incorrect deposits of energy, deterioration of the detector response and so on. Therefore it is crucial to have a way to correct and calibrate properly the reconstructed jet energy so as to have the highest possible measurement accuracy in our datasets.

The energy of a jet that is measured at the detector level differs significantly from the obtained at the particle level; the latter being derived from simulation without taking into account the detector system. This disagreement, as mentioned above, mainly comes from the response of the detector and the performance of the algorithm used to cluster the jets, which in this case is the anti-k_T algorithm. In order to have the best possible performance in approximating the jet energy at the reconstructed level to the respective at the generator (true) level, a factorized approach is followed with respect to the jet energy corrections, where each factor corrects for a different source.

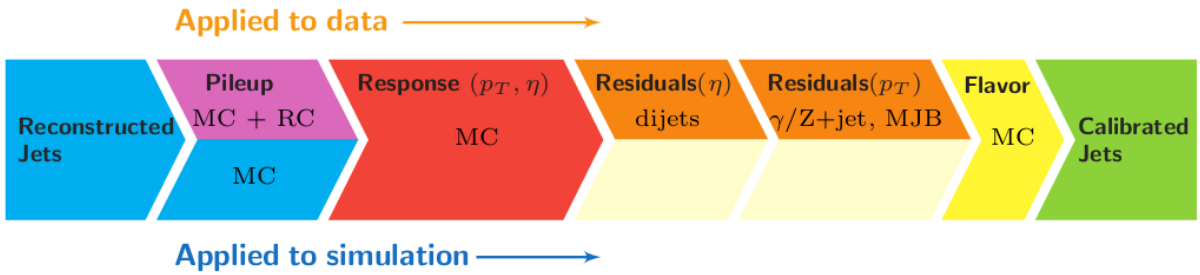

Figure 12: Jet Energy Corrections in CMS.[6]

In more detail, the first step are the pileup offset corrections (L1), described in Chapter 4.3, which are determined from simulation and aim to subtract the average extra energy that comes from additional to the main hard scatter pp collisions in the same or different bunch crossing, and is deposited inside the cone of the jet. Consequently, the relative (L2) and absolute (L3) corrections (described in Chapter 4.4), also derived from simulation, are applied in order to equate the reconstructed jet energy to the respective true one.

[6]Taken from Chapter 1 of [16] (Figure 2)

Afterwards, the residual corrections, obtained from data, are applied in order to account for any residual differences between data and simulation. More specifically, the residual relative ones (L2Res) aim to correct the response of jets over a wide range of p_T, relative to the response of jets with $|\eta| < 1.3$. For this purpose a data sample with dijet events is used, where the "tag" jet is in the barrel region $|\eta| < 1.3$ and the "probe" jet, whose response needs to be corrected, has an unconstrained pseudo rapidity. For the absolute residual corrections (L3Res), the absolute jet energy scale at $|\eta| < 1.3$ is determined with $Z(\rightarrow \mu^+\mu^-) + \text{jet}$, $Z(\rightarrow e^+e^-) + \text{jet}$ and $\gamma + \text{jet}$ events for $30\ GeV < p_T < 800\ GeV$ by comparing the reconstructed p_T of the jet to that of a precisely measured object, namely the Z boson or the photon. For $p_T > 800\ GeV$ the response is corrected using multijet events (MJB). Finally, there are also the optional jet flavor corrections, which are derived from simulation and account for the parton flavor (gluon or quark jets).

The consecutive steps of the factorized jet energy corrections are illustrated in Fig. 12. Summarizing the entire procedure, the overall correction factor that leads to the calibrated jets is given by:

$$\mathcal{C} = C_{offset}(p_T^{raw}, \rho, \eta, A_j) \cdot C_{L2L3}(p_T^{'}, \eta) \cdot C_{L2Res}(\eta) \cdot C_{L3Res}(p_T^{''}) \tag{82}$$

where ρ is the offset p_T density per event area, A_j is the jet area, $p_T^{'}$ is the transverse momentum of the jet after applying the offset correction and $p_T^{''}$ is the jet momentum after applying all previous corrections.

More details regarding the residual and flavor corrections can be found in [16], [17]. The main focus of this thesis will be the baseline jet energy corrections derived from simulation, namely the offset (L1) corrections and the relative and absolute ones (L2L3).

4.2 Ultra Legacy Simulations

The baseline jet energy corrections in this thesis will be derived using the Ultra Legacy (UL) Monte Carlo samples of years 2016, 2017 and 2018. The UL reconstruction was introduced as a counterpart to the previously used reconstruction where some significant differences between data and simulation were noticed in the distributions of the neutral electromagnetic (NEMF) and neutral hadron energy fraction (NHF).

More specifically, taking into consideration dijet events that contain back-to-back jets in the final state with $\Delta\phi = |\phi_{j_1} - \phi_{j_2}| > 2.7$ and using the PFJet200 trigger for data along with a $p_T > 200\ GeV$ cut, comparisons between data and simulation were made for the 2017 samples.

In Fig. 13 the neutral electromagnetic fraction for pseudo rapidity regions $|\eta| < 0.5$

(top) and $2.7 < |\eta| < 3.0$ (bottom) is examined for the older reconstruction (left-hand side) and the UL reconstruction (right-hand side). Although in the central region the agreement between data and simulation is very good for both reconstructions, in the $2.7 < |\eta| < 3.0$ bin large discrepancies are observed for the older reconstruction, which are not present in the UL one.

Similarly, in Fig. 14 the distribution of neutral hadron fraction can be found. Since NEMF and NHF quantities are correlated, their behavior with respect to the two reconstructions is expected to resemble with each other. Indeed, the compatibility between data and simulation is greatly improved in the UL reconstruction in the $2.7 < |\eta| < 3.0$ region, while maintaining at the same time a good agreement in the central η region.

Since the UL reconstruction is henceforth established to be used in analyses, the baseline jet energy corrections that will be described in the following chapters will use the UL simulations for years 2016, 2017 and 2018. In detail, there are 4 different MC to be examined, since for year 2016 there are two MC available; one with the APV issue simulated (APV UL 2016) and one without it (Non - APV UL 2016). The APV issue is related to readout chips for the silicon microstrips in the CMS tracker being destroyed while the 2016 data were being collected (Chapter 2.2.2). Therefore, the performance of the corrections for the APV UL 2016 simulation is expected to be worse due to the presence of this problem. All the simulation samples that were used in this thesis can be found in Appendix A.

Finally, another issue that needs to be addressed before proceeding to the study of jet energy corrections are the "veto" regions of UL 2017 and 2018. Due to various detector issues (e.g. broken pixel detectors) there are some problematic regions in the $\eta - \phi$ plane that cause inefficiencies and should not be taken into consideration while extracting the corrections. For that reason, the regions that are mentioned below are excluded from the derivation of both offset and relative and absolute corrections:

× **UL 2017** :

- ✓ $1.31 < \eta < 2.96$ & $-0.8727 < \phi < -0.5236$
- ✓ $0 < \eta < 1.4835$ & $2.705 < \phi < 3.1416$

× **UL 2018** :

- ✓ $-2.96 < \eta < -1.31$ & $-1.5708 < \phi < -0.8727$
- ✓ $0 < \eta < 1.31$ & $0.4363 < \phi < 0.7854$

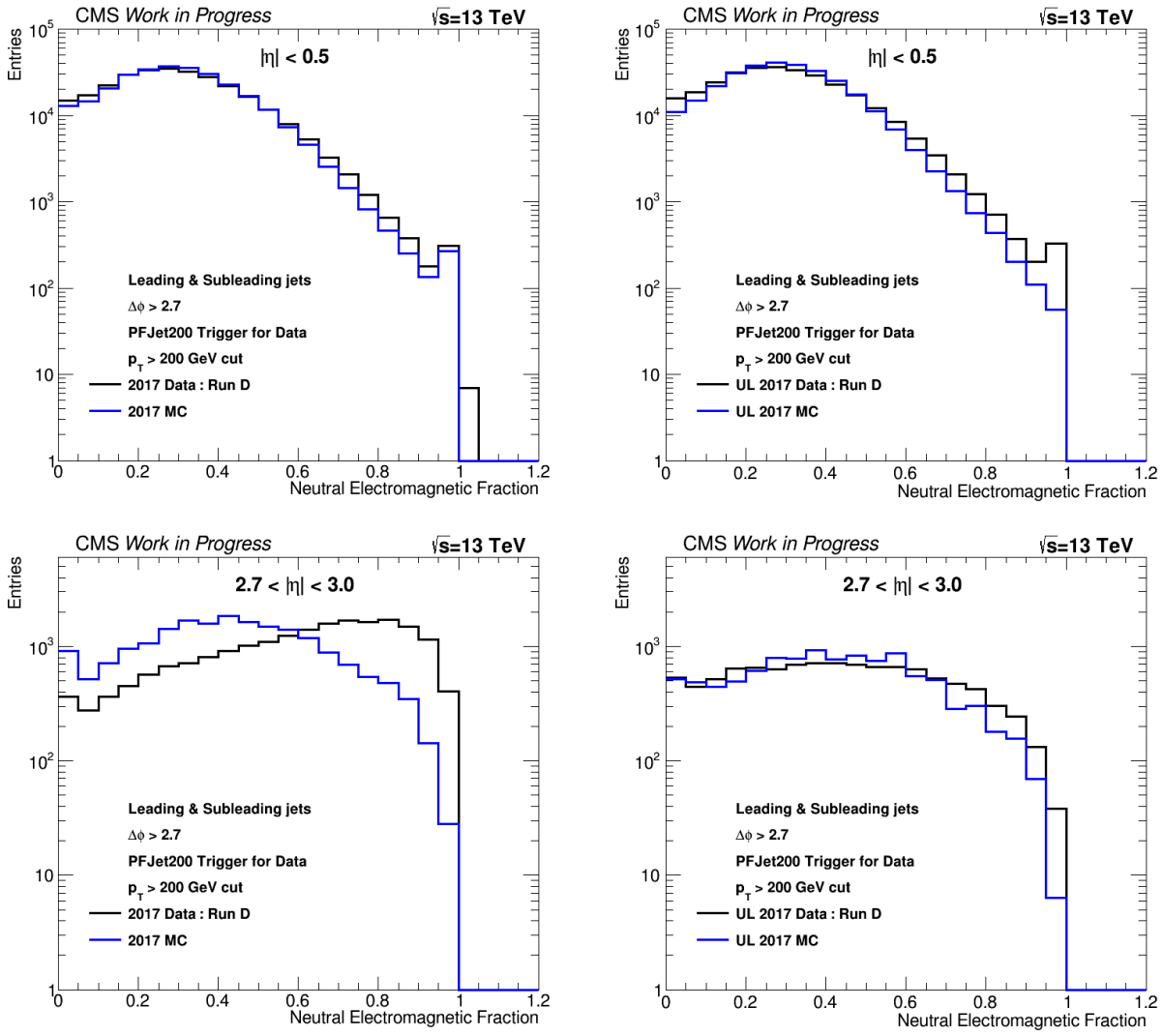

Figure 13: Left-hand side: Neutral Electromagnetic Fraction (NEMF) for 2017 Data from Run D (black) and 2017 MC (blue). Right-hand side: Neutral Electromagnetic Fraction (NEMF) for UL 2017 Data from Run D (black) and UL 2017 MC (blue).

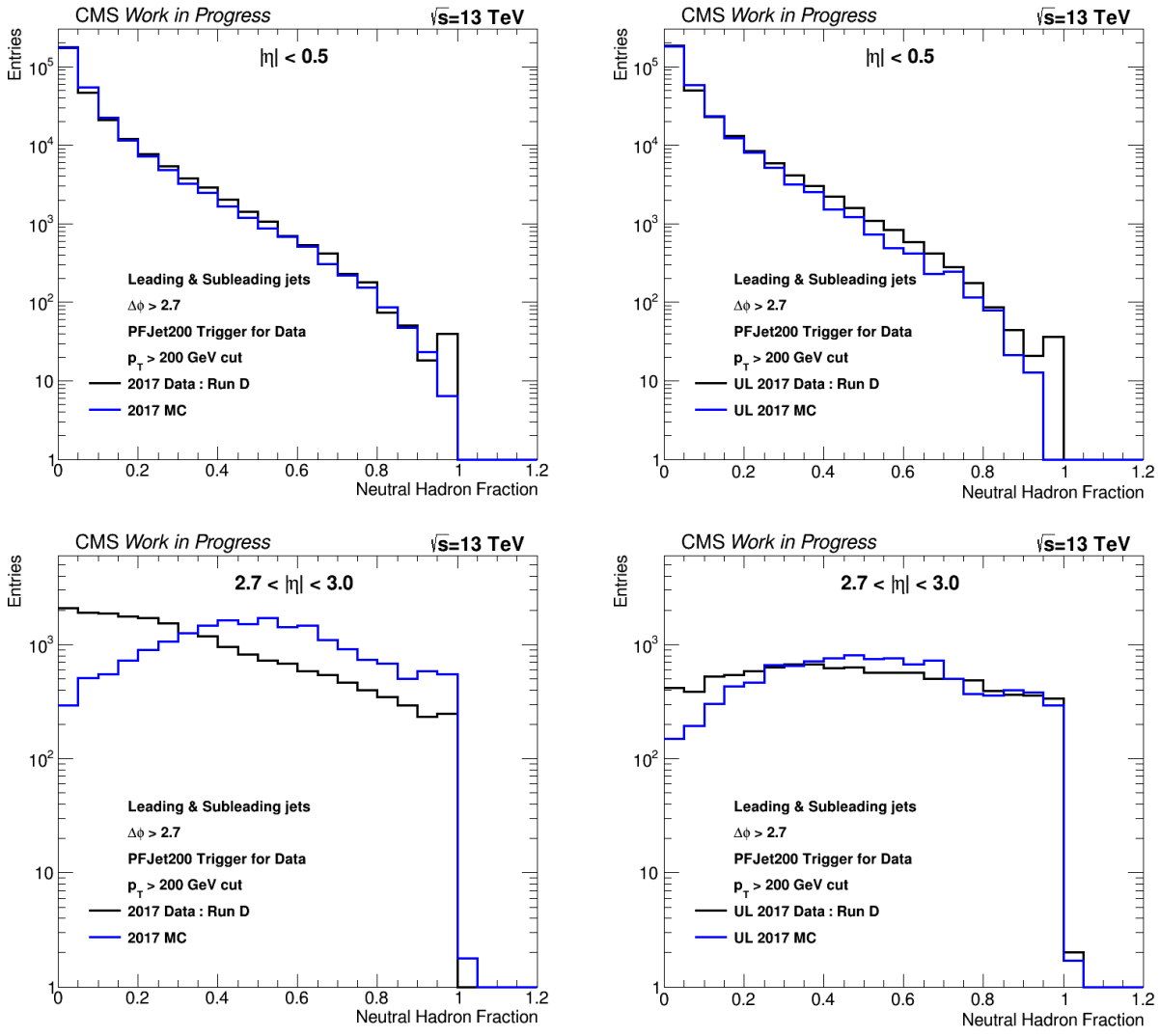

Figure 14: Left-hand side: Neutral Hadron Fraction (NHF) for 2017 Data from Run D (black) and 2017 MC (blue). Right-hand side: Neutral Hadron Fraction (NEMF) for UL 2017 Data from Run D (black) and UL 2017 MC (blue).

4.3 Pileup Offset Corrections

As it was briefly discussed in the introductory chapter 4.1, the pileup offset corrections (L1) are the first step in the chain of the baseline jet energy corrections. The aim of these corrections is to estimate and subtract the energy that is not associated with the high p_T hard - scatter collisions.

Due to the high instantaneous luminosity of LHC, a lot of additional proton - proton collisions take place within a bunch crossing, resulting in multiple tracks in the tracker and energy deposits in the calorimeters that do not come from the main, hard interaction that we are interested of. This contribution is called in - time pileup (IT PU). Moreover, since the calorimeters can not be perfect, it is common that the previous and subsequent beam crossings also contribute to the calorimetric energy in the same time window as the primary interaction. This kind of contribution is called out - of - time pileup (OOT PU).

As a result, it is evident that pileup, namely particles that are not part of the primary interaction, create an excess in the measured energy that needs to be calculated and removed.

As far as the PUPPI jets are concerned, as it was described in chapter 3.3, they are already taking into account the pileup by rescaling the particle four - momentum such that ideally a particle coming from a pileup interaction would get a zero weight and a particle coming from the hard scatter would get a weight of one. For this reason, the offset corrections are not needed for PUPPI jets and will not be derived for this jet collection.

Besides PUPPI jets, the other jet collection that will be examined are the Charged Hadron Subtracted (CHS) jets. Through the charge hadron subtraction method, the IN PU from charged particles is reduced by identifying the vertex that the pileup particle originates from and then removing that vertex before the jet clustering occurs. With this method approximately only half of the IN PU within the tracker coverage is removed, while the OOT PU is still present. Therefore, it is of the utmost importance for the offset corrections to be applied to CHS jets, both AK4 and AK8, since pileup can affect significantly the measured energy and momentum, not only for low p_T jets, where the pileup is mostly present, but also for higher p_T jets, as there is a non negligible probability of two or more soft jets overlapping, creating jets of higher p_T (tens of GeV), called pileup jets.

4.3.1 Investigation of Offset

The strategy used in order to derive the offset corrections requires two different Monte Carlo (MC) samples; one that has been created without the presence of pileup (No PU sample) or with a negligible amount of pileup added (Epsilon PU sample) and one that

has been created with the pileup conditions simulated (PU sample), by artificially adding soft four - vectors in the events and clustering them with the true jet components. These MC samples can be found in Table 3 of Appendix A. All the events of the PU sample are being matched with the respective events of the No PU (or Epsilon PU) and consequently, all the jets in each event are being matched between the two simulation samples with a $\Delta R = \sqrt{(\eta_{PUjet} - \eta_{NoPUjet})^2 + (\phi_{PUjet} - \phi_{NoPUjet})^2} < 0.2$ criterion. Due to studies that have concluded that the efficiency of this matching is worse for the No PU samples compared to the Epsilon PU samples, leading to even negative values of the offset energy, the latter samples will be preferred for the derivation of the corrections for the UL simulations.

In order to study the excess energy that is caused from the presence of pileup the offset quantity is introduced as the momentum of a jet from the PU sample minus the momentum of its matched jet from the Epsilon PU sample:

$$Offset = p_T^{PUjet} - p_T^{EpsilonPUjet} \tag{83}$$

Before describing the details of the offset corrections' derivation one should primarily examine the behavior of this offset quantity, as a function of the geometrical characteristics of the jets, their momenta and also the number of vertices that have been measured for each event that is studied.

Therefore, it is very convenient to split the detector in four areas in terms of the pseudo rapidity; the barrel region ($|\eta| < 1.3$), the inner endcap ($1.3 < |\eta| < 2.5$), the outter endcap ($2.5 < |\eta| < 3.0$) and the forward region ($3.0 < |\eta| < 5.0$). Due to the conditions of the pp collisions it is expected that pileup will be more present in the forward region, where most soft, in terms of p_T, jets can be found.

Furthermore, the offset energy can be studied for various number of pileup interactions per event, either at the reconstructed level or the generated. It is worth mentioning that the PU Monte Carlo samples that are used are flat samples, namely the distribution of true number of pileup interactions (μ) is flat from 0 up to 70, and the respective reconstructed number (N_{PU}) is also flat with a turn down at higher values, as shown in Fig. 15. Then, with the application of weights the distributions can be manipulated in order to match the respective conditions in data. However, PU reweighting does not affect the derivation of the pileup offset corrections.

In order to examine if the reconstructed number of interactions correspond to the proper values of the true number of interactions, two dimensional plots of N_{PU} (reconstructed) vs μ (generated) are shown in Fig. 32 of Appendix B, where it indeed seems that the majority of the events lie in the diagonal for all four UL MC samples.

For each one of the four pseudo rapidity areas, and for a specific bin of reconstructed

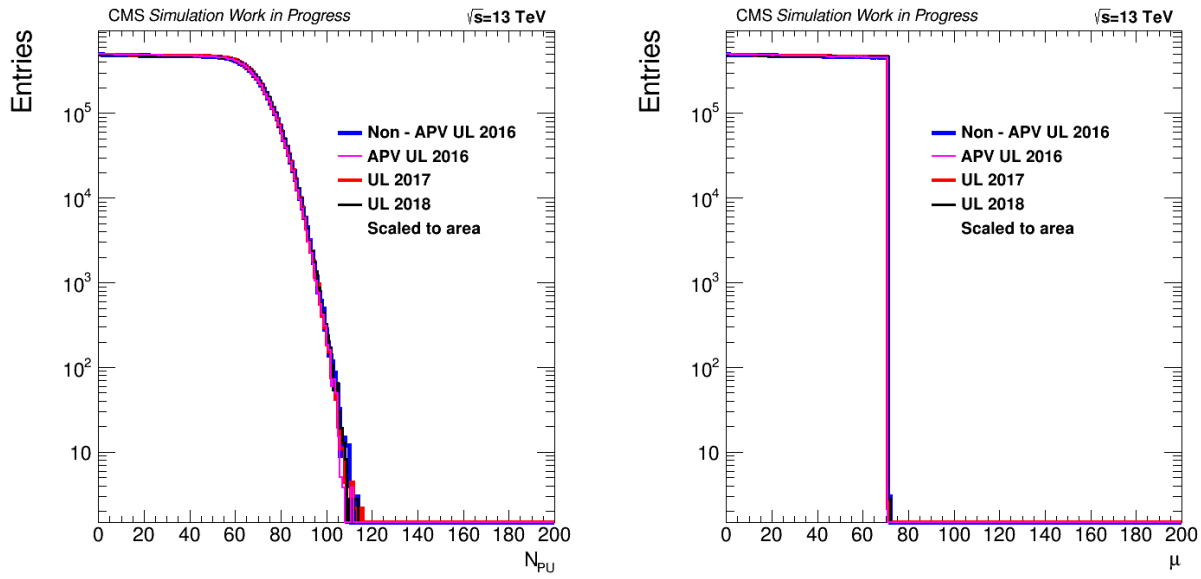

Figure 15: Distribution of reconstructed (left) and true (right) number of pileup interactions for all UL MC samples.

or generated number of pileup interactions, a two dimensional histogram of the offset as a function of the generated p_T can be studied. These histograms, that scan all detector areas and events with different number of vertices, are filled while the Monte Carlo events are reweighted. An example of such plot, when examining AK4 CHS jets of the UL 2017 MC in the outer endcap region ($2.5 < |\eta| < 3.0$) and for events that have $30 \leq \mu < 40$ can be viewed on the top side of Fig. 16.

Since it can become quite uncomfortable to examine so many of such two dimensional histograms for all μ bins, for visualization purposes one could study the mean value of the offset peak for each generated p_T bin. In order to do this, the peak of the offset distribution for each generated p_T bin can be fitted with the gaussian function and the mean and its error can be derived. This procedure can be found as an example on the bottom side of Fig. 16, where for the generated p_T bin of $150 - 200\ GeV$ the offset distribution is shown.

By following this approach for all the generated p_T bins, number of pileup interactions bins and η regions, the mean offset can be extracted and afterwards studied.

As far as AK4 CHS jets are concerned, in Fig. 17, the mean offset can be seen as a function of the generated p_T for all true pileup interactions (μ) and detector regions in terms of η, for the UL 2017 MC. Similarly, in Fig. 33 - 35 of Appendix B the offset is examined for all the other MC samples as well. As it was described before, all the points in these offset plots have been derived from gaussian fits, similar to the one depicted in Fig. 16. It is also worth clarifying that since the MC samples simulate events with pileup

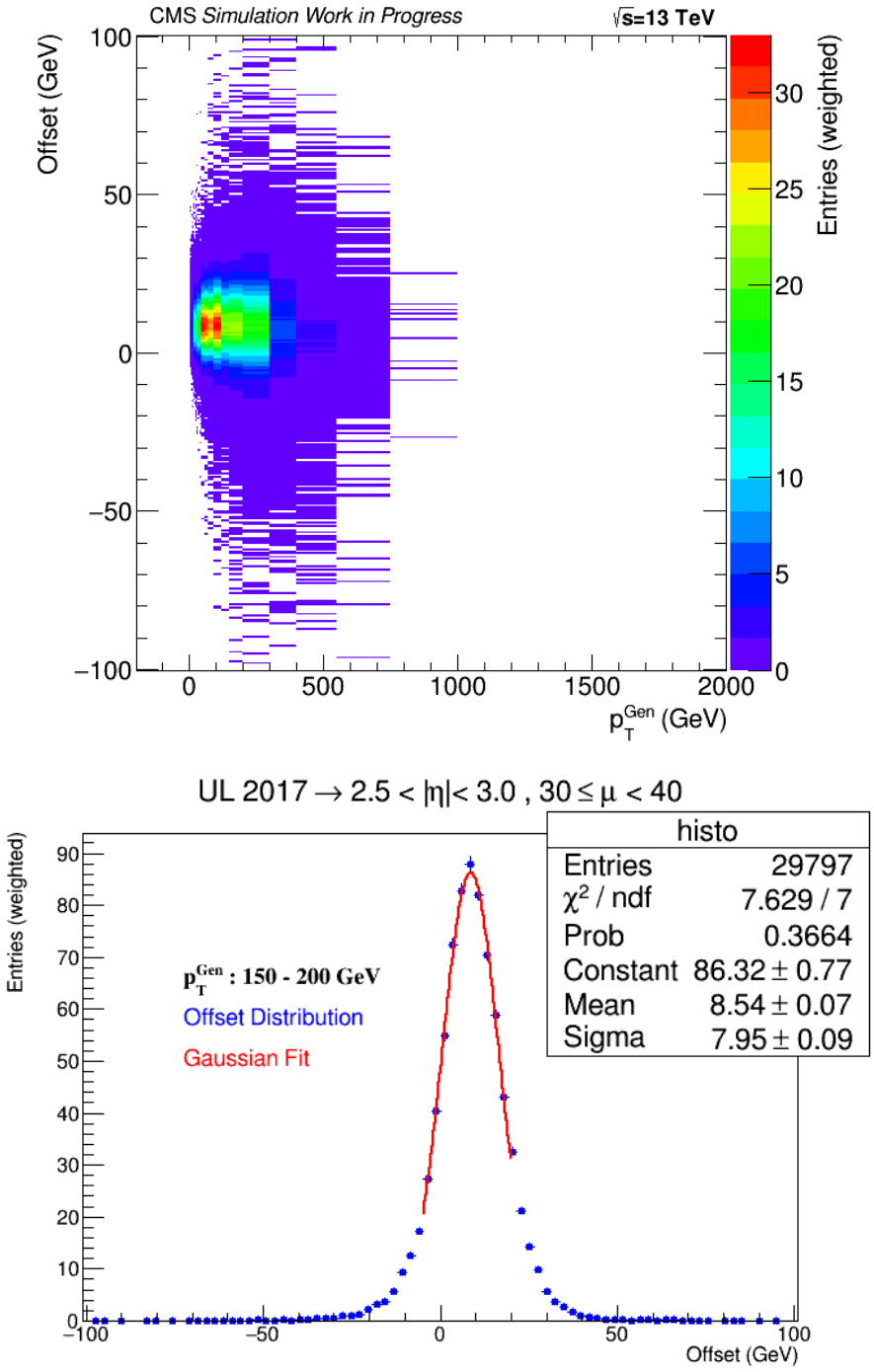

Figure 16: Top image: Weighted distribution of offset as a function of generated p_T for AK4 CHS jets of the UL 2017 MC with $2.5 < |\eta| < 3.0$ and $30 \leq \mu < 40$. Bottom image: Projection of the top image for generated p_T: $150 - 200\ GeV$.

interactions in the generated level up to 70, the bin noted as $70 \leq \mu < 80$ in the offset plots actually contain only the events that have $\mu = 70$. With the exception of the APV UL 2016 MC in the $|\eta| < 2.5$ region, the mean offset seems to have a very similar behavior

among the Ultra Legacy samples in every part of the phase space, which is reasonable since the physics procedures should not depend on the year the data were collected and consequently the Monte Carlo samples used to simulate these datasets should not differ significantly, with the exception of deviations in the reconstructed level due to different detector conditions. Moreover, it is concluded that the offset increases as the number of pileup interactions increase, which is something that is obviously expected, as more vertices in an event imply more pileup presence. In addition, the offset seems to be flat as a function of the generated p_T in most detector areas, which can be easily explained by the fact that the pileup particles that do not come from the main interaction are clustered randomly along with the true components of jets with different energy, and as a result the offset energy should not strongly depend on the jet momenta.

All in all, the offset seems to be healthy apart from a problematic behavior in the APV UL 2016 MC within the tracker region. In more detail, in the top plots of Fig. 34 a trend towards negative offset can be observed. This is unphysical, as the offset distribution by definition (Eq. 83) should always peak at positive values, since $p_T^{PUjet} > p_T^{EpsilonPUjet}$ due to an excess of pileup energy in the momentum of jets in the PU sample. This problem can be explained if one takes into account that this particular MC sample has the APV issue simulated. Since this issue is related to "holes" in the tracker, which in 2016 extended to pseudo rapidity regions of $|\eta| < 2.4$, it can be deduced that the offset energy, albeit present, was not detected and measured with a sufficient efficiency. Therefore, as the number of pileup interactions and p_T increases, the offset energy is not accounted for as intended, leading to the offset distribution peaking at lower and even negative values, which most likely is an artefact of this exact issue in the detector.

Similar plots as the ones that were just discussed are produced and studied with respect to AK8 CHS jets as well. In Fig. 18 the offset as a function of p_T^{Gen}, μ and η can be viewed for this jet collection and for the UL 2017 simulation. As a comparison, in Fig. 36 - 38 of Appendix B the respective plots for all the other UL MC samples can be found.

The first thing that is noticed is that the overall behavior of the offset for this jet collection is very much comparable to the one examined before, which is expected since the only difference between AK4 and AK8 CHS jets is their cone radius and not the strategy followed in order to produce them. As a result, the main difference that one could expect to see among these jets is the value of the mean offset. More specifically, AK8 jets have twice the cone radius as AK4 jets and consequently four times larger jet area. Therefore, AK8 jets are expected to have approximately four times more pileup particles that are clustered along with the true components of jets, leading to four times larger offset. Comparing Fig. 17 with Fig. 18 confirms that this is indeed the case.

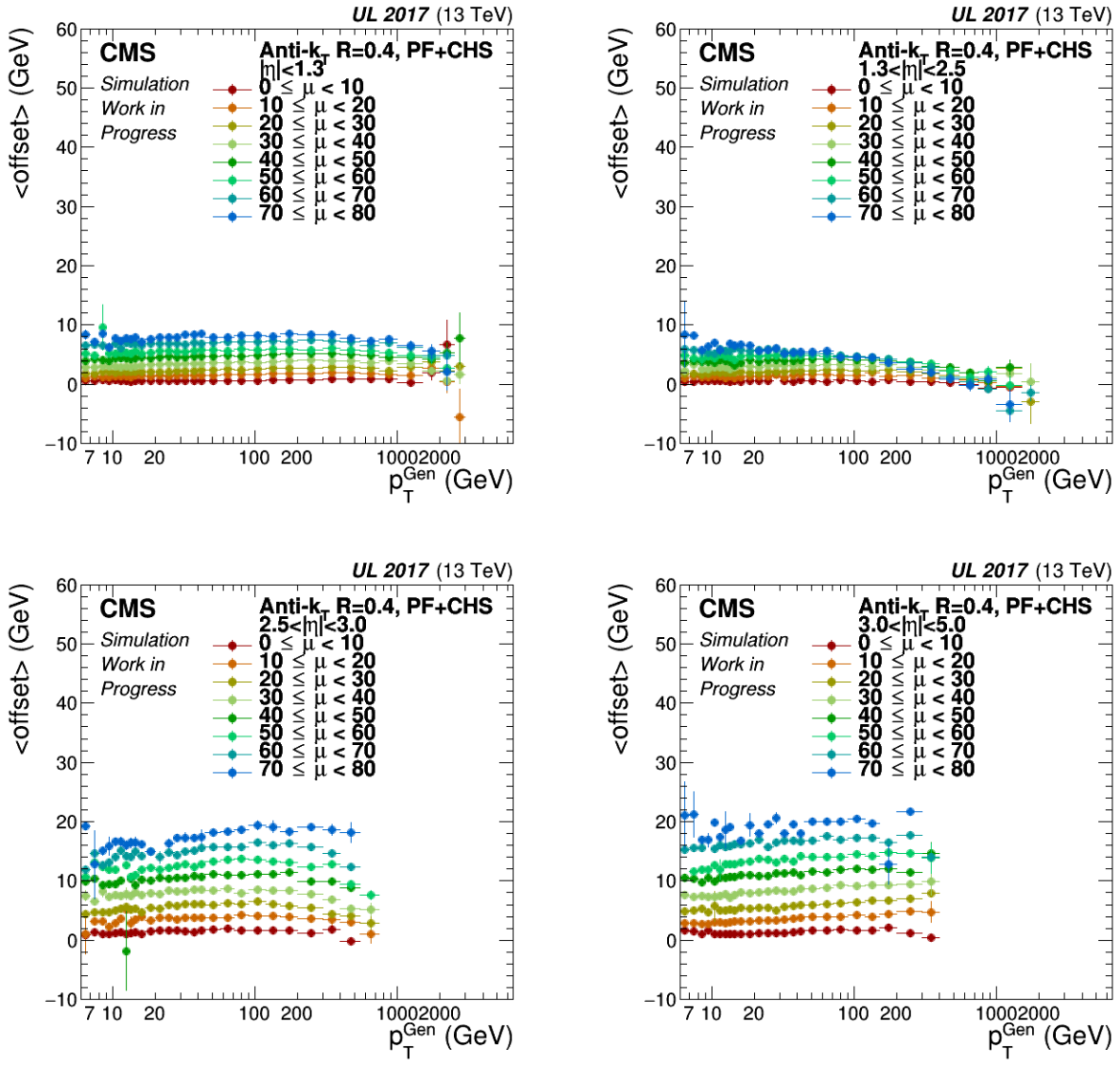

Figure 17: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2017 MC.

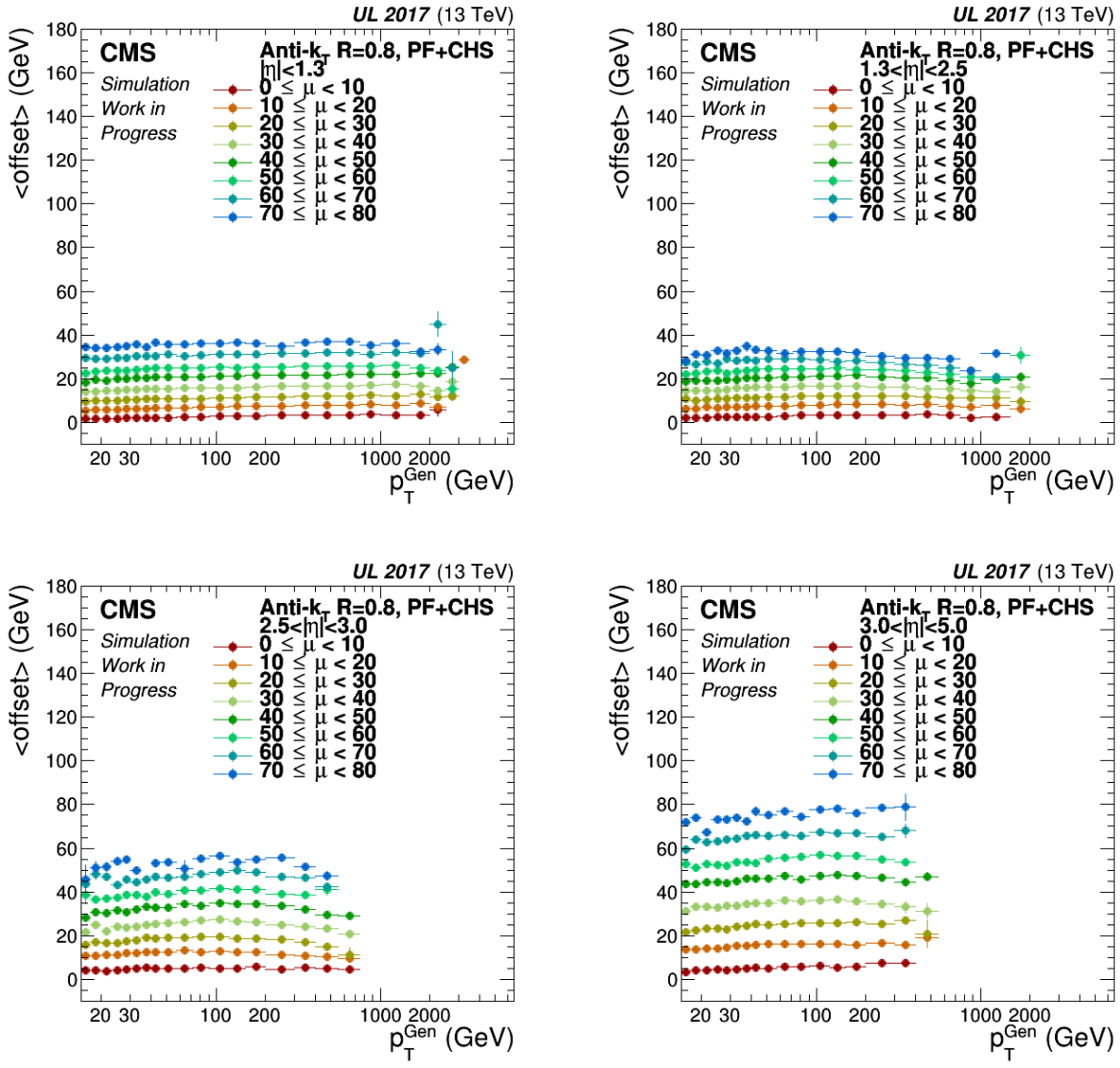

Figure 18: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the UL 2017 MC.

4.3.2 Derivation of the Offset Corrections

Since the offset behavior has been thoroughly studied, the next logical step would be to construct a methodology in order to remove it so that it would ideally be as close to zero as possible in the whole range of the phase space. This strategy will be described thereupon.

If the raw (i.e. uncorrected) transverse momentum that is measured in the reconstructed level is denoted as p_T^{raw}, then the output of the offset corrections will be a multiplicative factor (C_{offset}) that should rescale the measured momentum to a value that has all the pileup energy removed:

$$p_T^{corrected} = C_{offset} \cdot p_T^{raw} \tag{84}$$

Taking into consideration that the measured p_T^{raw} can be considered as the reconstructed p_T from the PU sample, while the corrected momentum as the reconstructed p_T from the Epsilon PU sample, it is obvious that the offset quantity of Eq. 83 can be rewritten as $Offset = p_T^{raw} - p_T^{corrected}$. Therefore, the offset correction is given by:

$$C_{offset} = 1 - \frac{Offset}{p_T^{raw}} \equiv 1 - \frac{A_j}{p_T^{raw}} \cdot (Offset\ /\ A_j) \tag{85}$$

where $A_j \approx \pi R^2$ is the jet area, which depends only on the cone radius of the jets that are studied ($R = 0.4$ for AK4 jets and $R = 0.8$ for AK8 jets).

It is evident that in order to calculate the offset correction, the pileup energy, namely the offset per jet area needs to be somehow parametrized and estimated, since both A_j and p_T^{raw} are known quantities. It is clear that this energy will strongly depend on the per - event offset p_T density per event area (ρ), but also needs to have an η - dependence so as to correct for the detector non - uniformity. Studies have also shown that this quantity also depends on the reconstructed transverse momentum p_T, i.e the uncorrected p_T^{raw}. More specifically regarding the ρ quantity, it is an event based variable which is calculated as the median of the distribution of the jet momenta p_{Tj} divided by their area A_j, where j runs over all jets in the event and is not sensitive to the presence of hard jets, since the median is taken into consideration instead of the mean. As a result, ρ characterizes the soft jet activity and is a combination of the underlying event, electronic noise and pileup.

There are three different parametric functions that are most frequently used for describing this pileup energy per jet area:

$$f\left(\rho, p_T^{raw}\right) = p_1\ (\rho - p_0)\left(1 + p_2\ \ln\left(\frac{p_T^{raw}}{15(GeV)}\right)\right) \tag{86}$$

$$g\left(\rho, p_T^{raw}\right) = p_0 + p_1\ (\rho - 20) + p_2\ \ln\left(\frac{p_T^{raw}}{\mathcal{P}}\right) + p_3\ \ln^2\left(\frac{p_T^{raw}}{\mathcal{P}}\right) + p_4\ (\rho - 20)\ \ln\left(\frac{p_T^{raw}}{\mathcal{P}}\right) + \\ + p_5\ (\rho - 20)\ \ln^2\left(\frac{p_T^{raw}}{\mathcal{P}}\right) \tag{87}$$

$$\text{where} \qquad \mathcal{P} = \begin{cases} 30\ GeV & \text{for AK4 jets} \\ 90\ GeV & \text{for AK8 jets} \end{cases}$$

$$h(\rho) = p_1\ (\rho - p_0) \tag{88}$$

The η dependence on all the above functions is applied by splitting the detector in 82 fine η bins from -5.191 to +5.191, that correspond to the calorimeter tower segmentation, and treating each η bin separately. It is also worth mentioning that function (88) does not have a p_T^{raw} dependence, which is later expected to be absorbed when applying the relative and absolute corrections.

Studies that can be summarized on Appendix C have shown that among the three functions, Eq. 86 gives the best closure results. Since, additionally, function 88 is too naive with respect to not having a p_T dependence, and function 87 can become too complex for uncertainty analysis, the first one (86) has been chosen for the derivation of the corrections for the Ultra Legacy samples and will be preferred to be used in this thesis.

The mean offset over jet area $< Offset/A_j >$, along with mean $< \rho >$ and $< p_T^{raw} >$ are calculated for each 3D cell of variables (η, μ, p_T^{Gen}). Afterwards, for each one of the 82 η bins this offset over jet area is plotted as a function of pileup density ρ and raw jet momentum p_T^{raw}, or simply denoted as p_T^{Reco} since it is the measured momentum at the reconstructed level. An example of such plot for AK4 CHS jets of the UL 2018 MC and the range $3.139 < \eta < 3.314$ can be seen on the left hand side of Fig. 19. In this figure, although not explicitly stated, it is implied, as mentioned previously, that all quantities have been calculated as an average value. Since PU reweighting is a constant number in each μ bin, there is no difference if it is applied or not in the derivation of these corrections. Afterwards, the offset over jet area is fitted with function $f(\rho, p_T^{Reco})$ of Eq. 86 in all the pseudo rapidity bins, setting the ranges of the two dimensional fits in all those bins as:

- ✓ $8\ GeV < p_T^{Reco} < 6500\ GeV$
- ✓ $0\ GeV < \rho < 70\ GeV$

On the right hand side of Fig. 19 the fitted function for this particular η bin and for the specified ranges of ρ and p_T^{Reco} can be viewed.

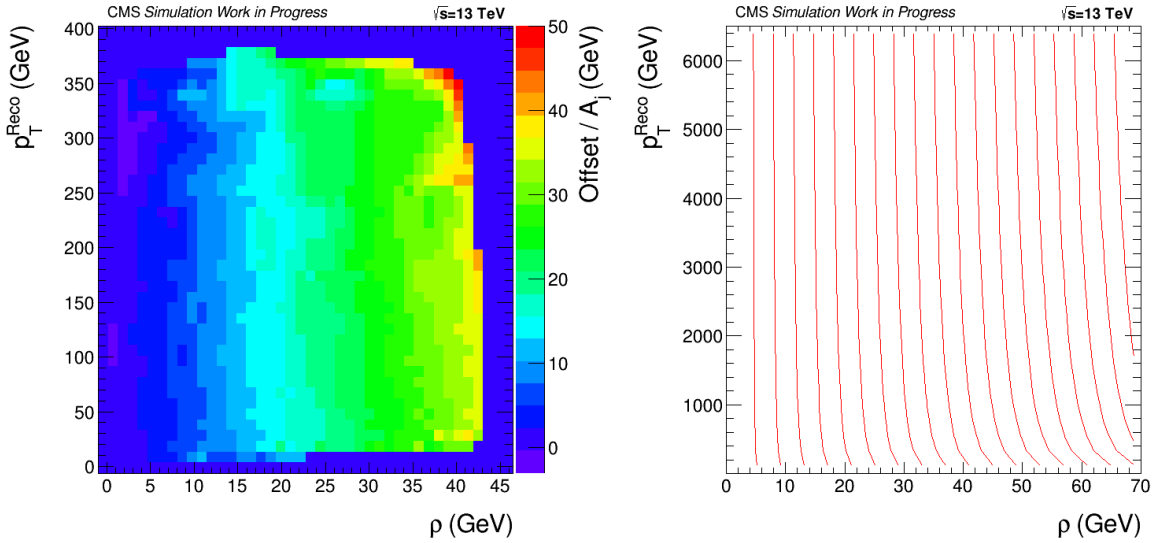

Figure 19: On the left: Offset over jet area as a function of p_T^{Reco} and ρ for the UL 2018 MC and AK4 CHS jets with $3.139 < \eta < 3.314$. On the right: The fitting function $f(\rho, p_T^{Reco})$ drawn for this specified η bin.

After having performed this fitting procedure, the parameters p_0, p_1 and p_2 of Eq. 86 are extracted for each one of the 82 η bins and the correction factor C_{offset} (Eq. 85) is calculated. If C_{offset} becomes less than 0.0001 when calculated from Eq. 85 then it is set by hand to be equal to 0.0001 so as to avoid negative or zero values that could cause problems.

As an example, in Fig. 20 this correction factor is drawn for AK4 CHS jets, as a function of ρ, for two different η values and fixed transverse momentum of $p_T^{Reco} = 100\ GeV$. It can be noticed that the correction factor has a linear behavior as a function of ρ, as Eq. 85 and 86 dictate. More specifically the correction factor has a negative slope because the more pileup is present the lower C_{offset} should be in order to downsize the jet momentum and remove the offset energy properly. As a result, the offset correction is also $C_{offset} \leq 1$ because the raw energy needs to be rescaled downwards in order to subtract the extra pileup energy. However, for the APV UL 2016 simulation and central η regions, as the $\eta = 0.8$ value that is seen in the left-hand side plot of Fig. 20, an opposite behavior is observed. This can be explained by the fact that the APV issue in the tracker has caused even negative mean offset, as already discussed in the previous section, thus forcing the corrections to scale the energy upwards, instead of downwards, in order to bring the offset energy to zero.

In order to cross check that the correction factors are indeed calculated properly the offset corrections are applied on the same pileup enriched MC that was used to produce them and the offset behavior is studied once more. This examination will be performed

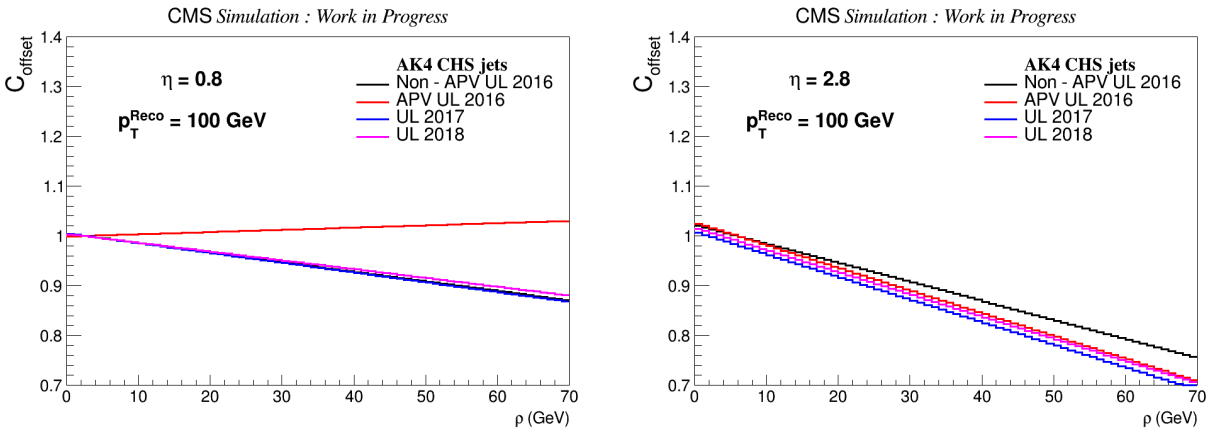

Figure 20: L1 correction factor (C_{offset}) for AK4 CHS jets, as a function of ρ for $\eta = 0.8$ (left) and $\eta = 2.8$ (right) and fixed p_T^{Reco} at 100 GeV.

with the same way that the plots of Fig. 17 and 18 were produced. However, this time the offset is expected to be around zero, if the corrections have indeed removed the pileup energy, as intended. The offset is examined in bins of true number of pileup interactions (μ), as before, in order to avoid taking into consideration fake vertices that are measured at the reconstructed level.

In Fig. 21 the offset is examined for AK4 CHS jets from the UL 2017 simulation, while in Fig. 22 the respective closure plots are presented for the AK8 CHS jets. In order to compare with all the other UL simulations at hand, in Fig. 39 - 41 and 42 - 44 of Appendix B the corrected offset is studied for AK4 and AK8 CHS jets respectively, for the Non - APV UL 2016, APV UL 2016 and UL 2018 MC.

As far as CHS jets with cone radius of 0.4 are concerned, it can be concluded that the offset corrections have successfully removed the pileup energy in the UL 2017 and 2018 simulations (Fig. 21 and 41), as the offset is flat and zero as a function of the generated jet momentum for all μ bins and detector regions. Regarding the 2016 simulations, in the forward region where the pileup is most present the offset is, as expected, zero, while for the rest of η regions small deviations around zero can be observed in both MC samples. However, a significant non closure is noticed in the APV MC, at the very high p_T spectrum and for $|\eta| < 2.5$. There, the corrections seem to not be able to fully correct the pileup energy making the offset zero, since the offset before the application of the corrections is very problematic and has a trend towards negative values, as already discussed and seen in Fig. 34.

For AK8 CHS jets, similar conclusions can be derived by examining the behavior of the corrected offset, but due to larger cone and consequently larger offset energy, the deviations from zero are more intense. In more detail, for the UL 2017 and 2018 MC the

offset is zero in most cases, except for the presence of $\simeq 2-3$ GeV pileup for the highest μ bins, and also a non closure at the higher momenta in the inner endcap, that does not exceed the energy of $\simeq 5$ GeV offset in jets with total energy of $\simeq 1-2$ TeV. For jets from the 2016 simulations on the other hand, the performance of offset corrections seem to degrade, with the sample that has the APV issue simulated giving the worst closure among all four MC samples, mostly at higher μ bins where a lot of soft interactions take place.

Since the offset corrections comprise the first step of the baseline jet energy corrections and the relative and absolute corrections will be applied on top of them, it is expected that any remaining non closures in the offset will be later absorbed and the overall closure after the application of the L2 and L3 corrections will be the desired. For that reason it is evident to proceed with the examination and extraction of the relative and absolute corrections.

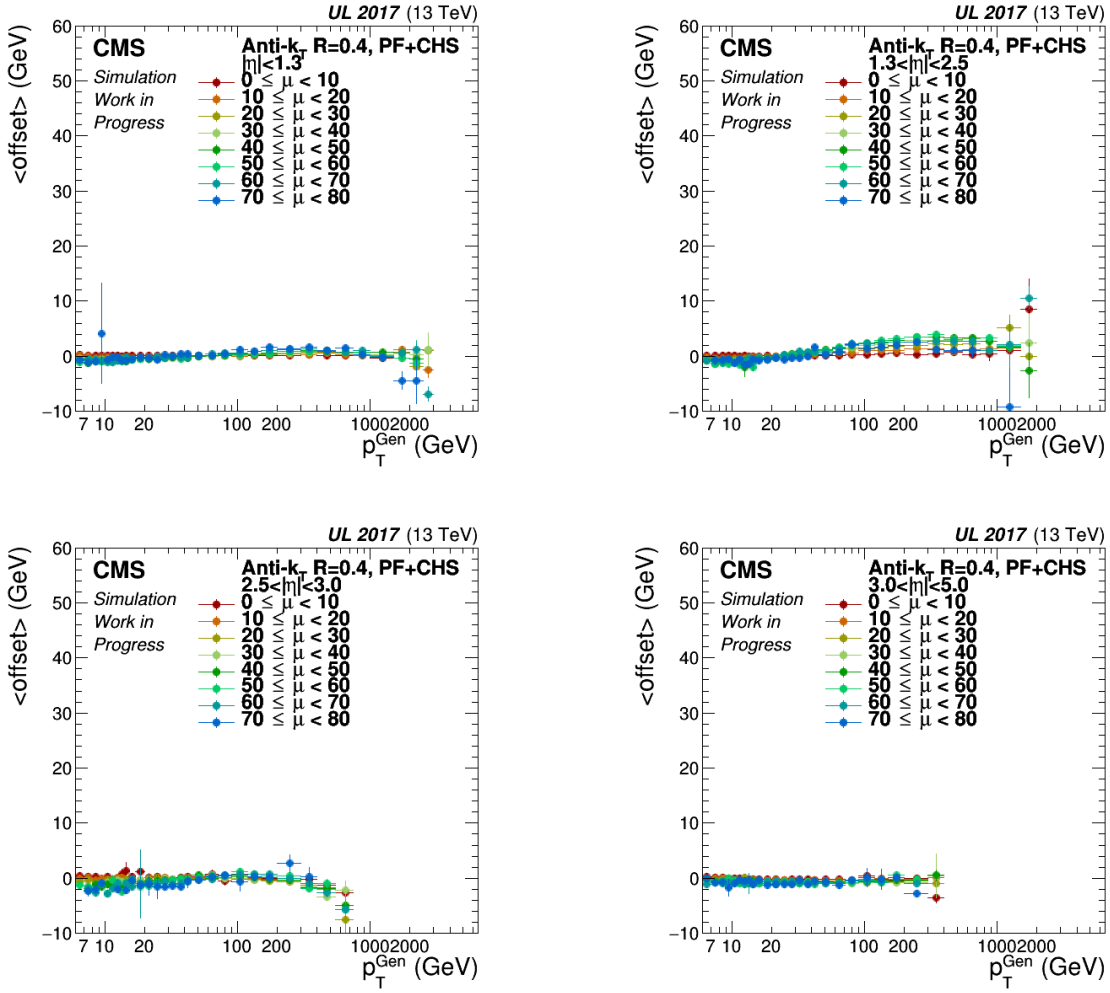

Figure 21: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2017 MC.

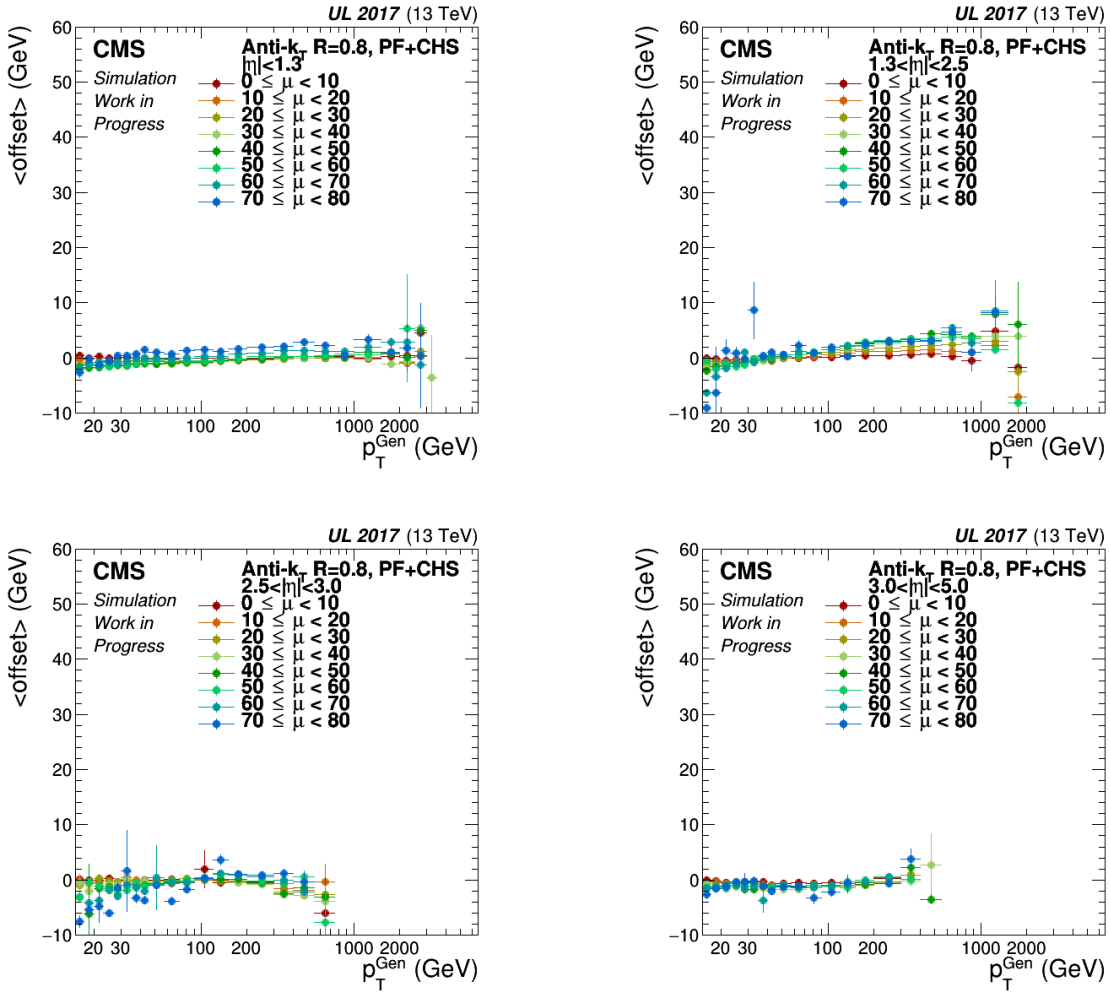

Figure 22: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the UL 2017 MC.

4.4 Relative and Absolute Corrections

The relative corrections (L2) as a function of jet p_T and η are obtained to equalize the response from different sub detector systems, while the absolute corrections (L3) aim to correct the reconstructed jet energy to the true (generated) one; the latter being the simulated energy of jets from theory without taking into account the detector system, namely the hypothetically measured energy if all detector conditions were perfect. Both of these corrections (L2 & L3) are derived from simulation and applied at the same time. For the PUPPI jets they comprise the baseline jet energy corrections, since the offset ones are not calculated for this jet collection. On the other hand, as far as the CHS jet are concerned, the relative and absolute corrections are applied on top of the offset ones, namely they correct the jet $p_T^{'}$, where $p_T^{'} = C_{offset}\, p_T^{raw}$.

4.4.1 Derivation of the Relative and Absolute Corrections

For these corrections a Monte Carlo file with pileup conditions simulated is used, different, however, than the one used for the offset corrections. In Table 4 of Appendix A all simulation samples that were used can be found. Since the goal of the absolute corrections is to equate the reconstructed momentum (p_T^{Reco}) with the true one at the generated level (p_T^{Gen}), an efficient matching between reconstructed and generated jets needs to be made. For this reason, a $\Delta R = \sqrt{(\eta_{Recojet} - \eta_{Genjet})^2 + (\phi_{Recojet} - \phi_{Genjet})^2} < 0.2$ criterion will be used for both CHS and PUPPI jets and for different cone sizes (AK4 and AK8).

After the application of the offset corrections, in the case of CHS jets, and the determination of the corrected jet momentum, which from now on will be denoted simply as p_T^{Reco}, the response is calculated and its distribution is examined. This quantity is defined as follows:

$$Response = \frac{p_T^{Reco}}{p_T^{Gen}} \tag{89}$$

The response distribution is greatly dependent on both the detector area where the examined jets have been located, and their energy. As a result, the detector is split in terms of the pseudo rapidity in 82 fine bins (the same ones that were taken into consideration in the offset corrections) and additionally a fine binning in reconstructed and generated momentum is considered. It is obvious that the main goal of the relative and absolute corrections is to correct the response so that it will become as close to unity as possible for every η and p_T region.

The main strategy that is being followed for the derivation of such corrections starts with the production of the response distribution for all the determined η and p_T^{Gen} bins.

An example of such distribution for AK4 CHS jets with $0.879 < \eta < 0.957$ and $30\ GeV < p_T^{Gen} < 35\ GeV$ from the Non - APV UL 2016 MC sample can be viewed in Fig. 23. Afterwards, the median of this histogram is calculated and this procedure is repeated for all the p_T^{Gen} bins in the specified η range. As a result, the median of the response as a function of the generated momentum can be studied for the specific η bin, and consequently for all the other η ranges as well. The graph for the $0.879 < \eta < 0.957$ range that was discussed previously can be found in Fig. 24.

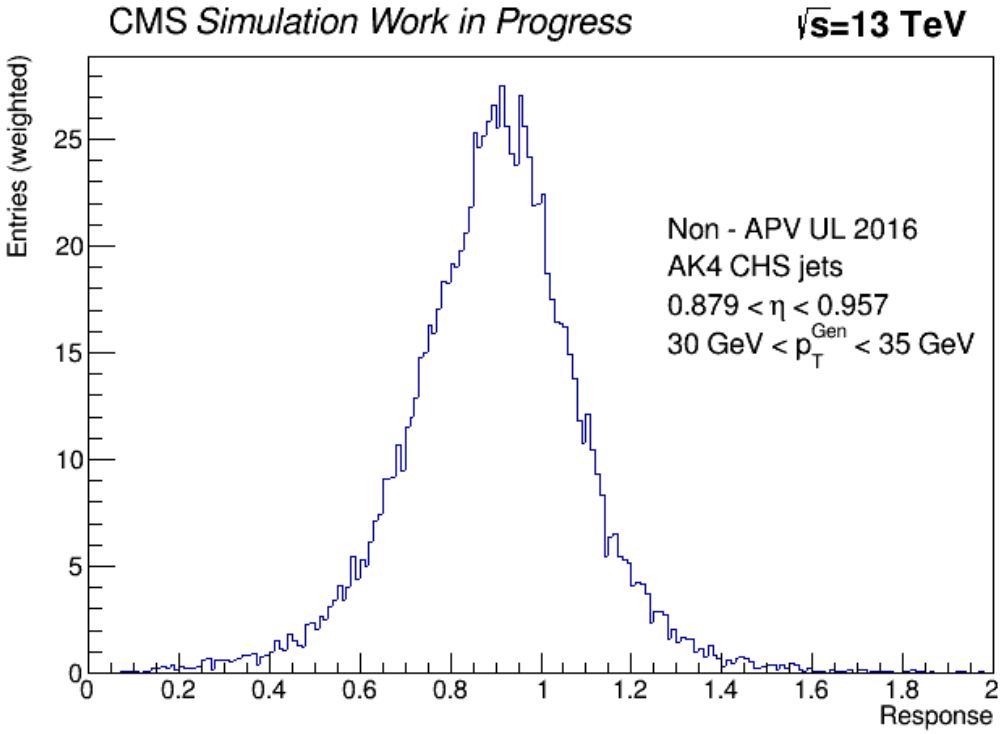

Figure 23: Response distribution for AK4 CHS jets from the Non - APV UL 2016 MC file with $0.879 < \eta < 0.957$ and $30\ GeV < p_T^{Gen} < 35\ GeV$.

As it was the case for the offset corrections, the relative and absolute (L2L3) ones are also a multiplicative factor C_{L2L3}, which corrects the jet energy. Since $p_T^{Gen} = C_{L2L3}\, p_T^{Reco}$, it is obvious that the correction factor that needs to be calculated is given by the following formula:

$$C_{L2L3} = \frac{1}{\text{median}(Response)} \tag{90}$$

As a result, the correction factor C_{L2L3} can be calculated via the median response, as a function of the reconstructed, i.e. uncorrected, jet p_T. In Fig. 25 the correction factor (with black points) can be examined for the bin $0.879 < \eta < 0.957$. In order to derive the correction factor for the continuous spectrum of the jet momentum and not just the

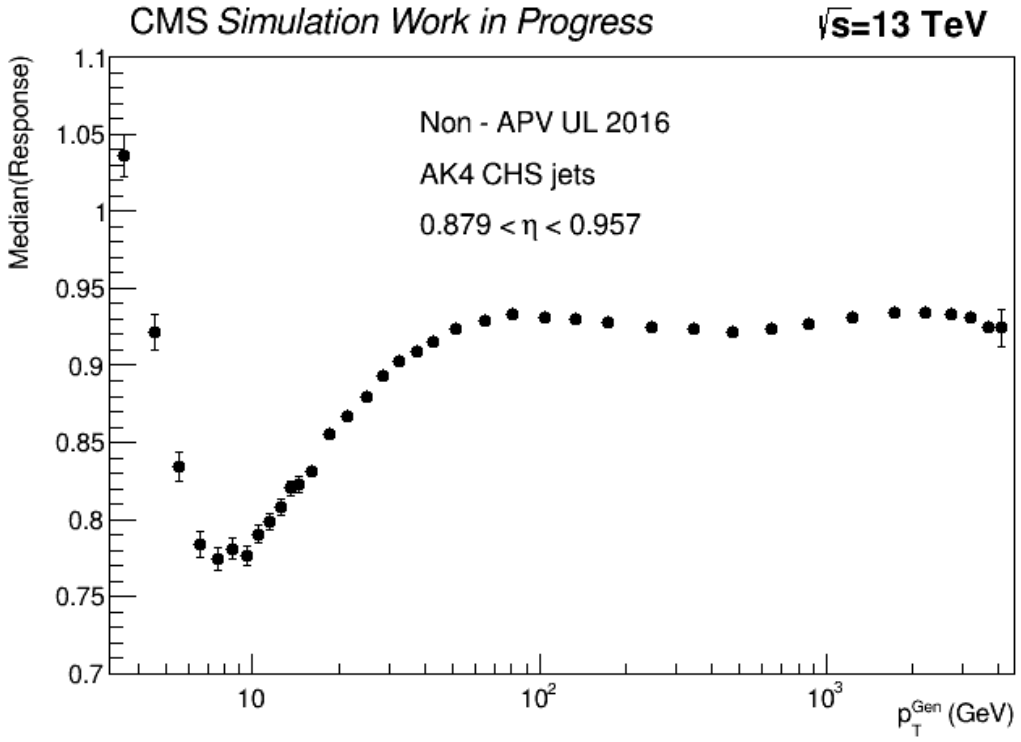

Figure 24: Median response as a function of p_T^{Gen} for AK4 CHS jets from the Non - APV UL 2016 MC file with $0.879 < \eta < 0.957$.

p_T^{Reco} bins where it is calculated, a one dimensional fit is performed using a 9 parameter function, called "standard+Gaussian". Its formula can be found on Eq. 91.

$$f\left(p_T^{Reco}\right) = p_0 + \frac{p_1}{log_{10}^2\left(p_T^{Reco}\right) + p_2} + p_3\, e^{-p_4\left(log_{10}\left(p_T^{Reco}\right) - p_5\right)^2} + p_6\, e^{-p_7\left(log_{10}\left(p_T^{Reco}\right) - p_8\right)^2} \tag{91}$$

This procedure is repeated for all the η ranges, so that the corrections can be derived for the whole detector system. The fitting procedure with Eq. 91 is performed for all jet collections and jet cones, with the exception of PUPPI jets in the most forward η regions where parameters p_3 and p_6 are fixed to zero in order to improve the fit probability.

An important issue that needs to be addressed regarding the fitting of the correction factors is that for very low reconstructed jet momenta an unphysical behavior of the correction is observed. In more detail, as it can be seen in Fig. 25, there is a turn on at the lowest p_T^{Reco} values where the correction starts to decrease. This is problematic since the lower the jet momentum the more information in the measurement of the energy is lost, leading to a lower response and consequently to a larger correction factor. For that reason, the correction factor for the lowest momenta cannot be trusted.

Therefore, in order to avoid taking into consideration such problematic behavior when

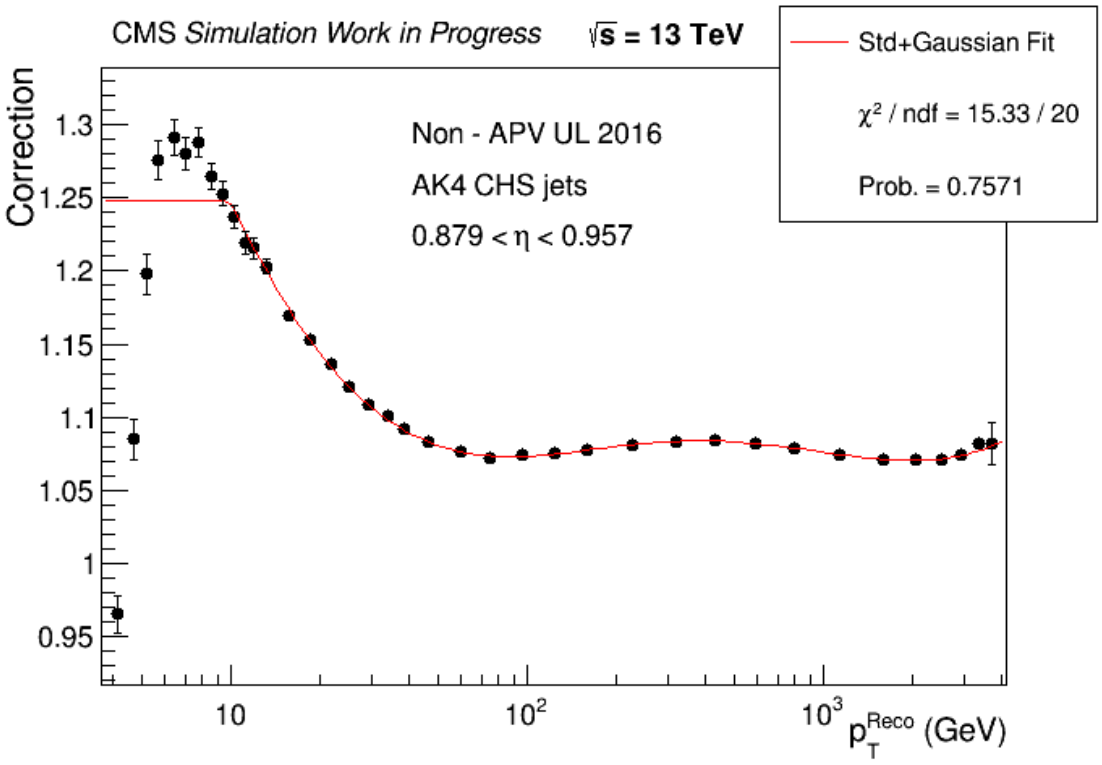

Figure 25: Correction factor (black points) as a function of p_T^{Reco} for AK4 CHS jets from the Non - APV UL 2016 MC file with $0.879 < \eta < 0.957$. Fit with standard+Gaussian function (red line) is also shown.

deriving the absolute corrections, the pt - clipping technique is introduced. With this technique the correction factor for $p_T^{Reco} < X\ GeV$ is set by hand to be equal to the correction factor of jets with $p_T^{Reco} = X\ GeV$. In this thesis, the pt - clipping technique at $X = 10\ GeV$ has been chosen to be applied when fitting for all the detector regions, since it has been found that jets with momenta greater than 10 GeV have a healthy correction factor, regardless of their η coordinate. Instead of using a fit function to describe the correction factor in each η area and jet energy, previous iterations of jet energy corrections were using splines. Splines are actually straight lines that connect each point in the Correction vs p_T^{Reco} plane with its neighbor ones. Along with this method the pt - clipping technique, which was described earlier, was being used.

Corrections that have been derived with splines give comparable results to the "standard+Gaussian" fitting function with respect to the closure tests. However, the method with the fitting function is henceforth preferred and has been used in this thesis. The reason behind this is that splines generally have higher local statistical fluctuations than fit functions, and they occasionally generate localised bumps or extrapolate badly at high and low momenta. The bumps in particular can create problems with respect to inclusive

jet analyses, as they inflate the statistical uncertainties in data and make robust PDF fits difficult.

After fitting all 82 η bins with the method mentioned above, the parameters $p_0,...,p_9$ of Eq. 91 are calculated and saved in a text file, thus providing all the information needed for the correction factors to be derived. As it was also the case for the offset corrections, if the factor C_{L2L3} becomes less than 0.0001 for some η or p_T bins, then it is set by hand to be equal to 0.0001.

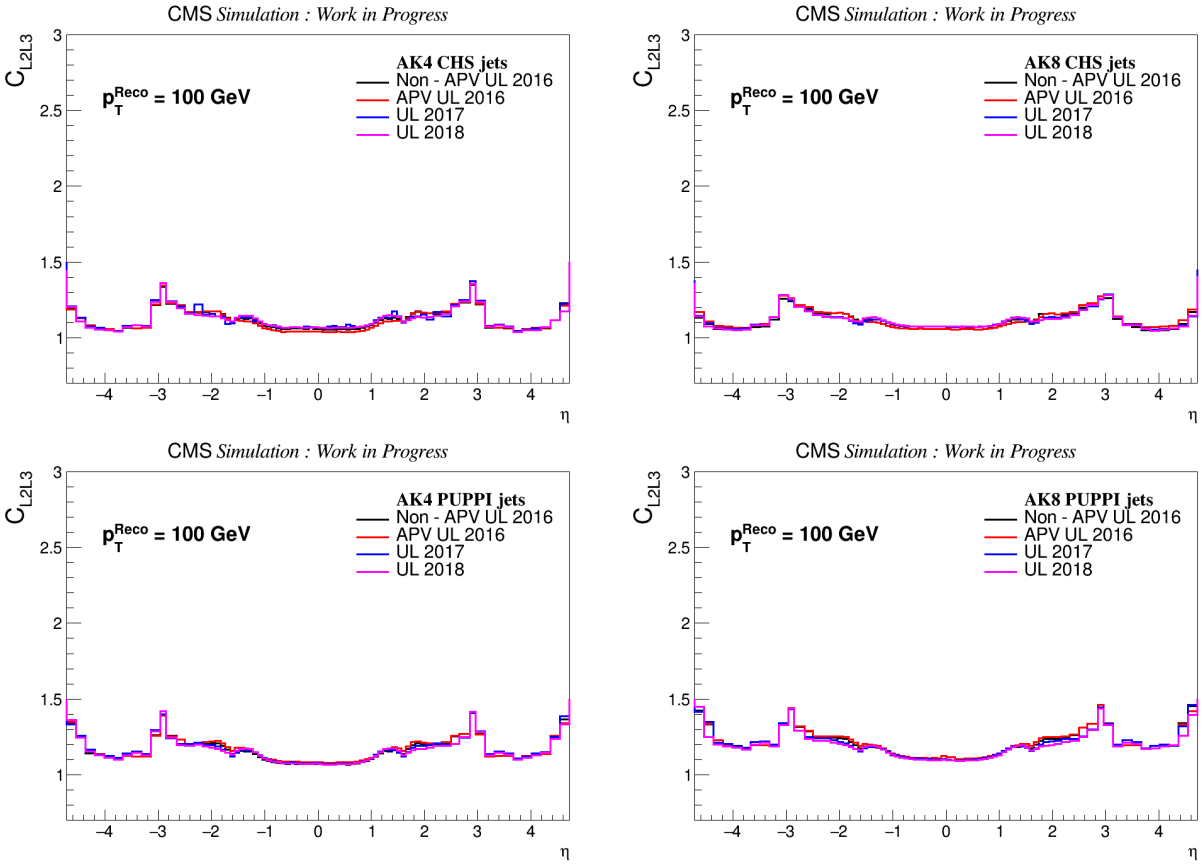

Figure 26: L2L3 correction factor (C_{L2L3}) as a function of η for jets with $p_T^{Reco} = 100\ GeV$, for all four jet collections and UL simulations superimposed.

In Fig. 26 the correction factor C_{L2L3}, for all UL simulations superimposed, is drawn as a function of jet η for jets with $p_T^{Reco} = 100\ GeV$, and for all four jet collections. From these figures it can be observed that the correction factor is in a very good agreement among all simulations. Additionally, its distribution is symmetric between positive and negative η values, which is an observation that proves that there are no issues regarding the derivation of corrections. Another important remark is the fact that in the regions $\eta \approx \pm 3$ the correction factor seems to peak for all the UL simulations, opposed to the rest of the $-4.5 < \eta < 4.5$ phase space where it is relatively more flat. This can be explained by

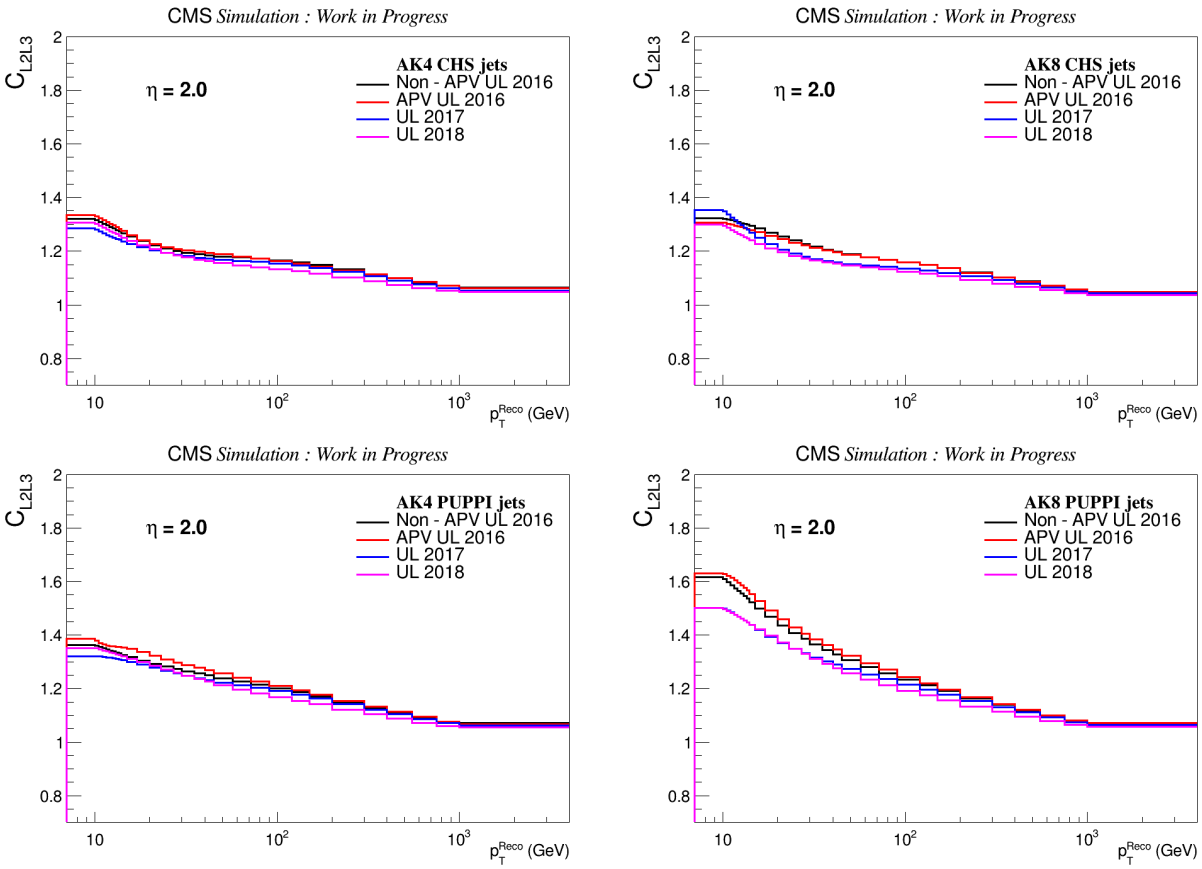

Figure 27: L2L3 correction factor (C_{L2L3}) as a function of p_T^{Reco} for jets with $\eta = 2$, for all four jet collections and UL simulations superimposed.

the fact that this is the interface region between where the hadron calorimeter coverage extends ($|\eta| < 3$) and the forward region, with the iron/quartz-fibre calorimeter installed, begins. As a result, it is a quite unstable region where more information regarding the jet energy is lost, thus affecting the correction factor.

In Fig. 27 the same correction factor is drawn this time as a function of p_T^{Reco} for jets with $\eta = 2$. The distributions seem to be smooth, without any large discontinuities or pathological behaviors. Apart from a small difference at low momenta between AK8 PUPPI jets from the 2016 and 2017+2018 simulations, the correction factors of all MC are in agreement with each other.

All in all, the relative and absolute corrections have a healthy behavior, therefore it is important to proceed with the examination of the closure in order to confirm whether the response is corrected as intended. This will be the main focus of the subsequent chapter. A final remark worth mentioning here is that both as a function of η and p_T the correction factor is always $C_{L2L3} \geq 1$ because the reconstructed, i.e. uncorrected jet momentum is lower than the generated one due to information loss in the measurement. As a result, the

correction factor in this step of the factorized corrections needs to be greater than unity in order to rescale the jet momentum upwards until it is equated with the true one.

4.4.2 Closure and Sanity Checks on the Relative and Absolute Corrections

It is clear that it is crucial to find a way to confirm that the corrections that have been derived for each Monte Carlo file and jet collection with the method that was described in the previous section are working properly. The methodology that is used in order to achieve this is that they are applied to a Monte Carlo file, that can be either the same that was used for the derivation or a different one, and the response (Eq. 89) is examined. In this particular thesis, results when the corrections are applied to the same Pile Up enriched Monte Carlo sample that was used to derive the relative and absolute corrections will be shown. Since the L2L3 corrections are applied on top of the offset ones, it is worth clarifying that the closure tests that will be performed and shown subsequently are the overall closures of the baseline jet energy corrections.

Taking into consideration that the corrections aim to equate the jet energy at the reconstructed level to the respective at the generated level, the response should be equal to unity in all the phase space area after the application of the corrections. Hence, two different closure plots will be studied; the response as a function of the generated momentum (p_T^{Gen}) for various η regions, and the response as a function of jet η for various p_T^{Gen} values. In this section the closure plots for all 4 jets collections from the UL 2017 simulation will be shown. As a reference, similar figures can be found for all the other UL simulations in Appendix D.

Starting with CHS jets that have a cone radius of $R = 0.4$, in the left-hand side plot of Fig. 28 the response as a function of p_T^{Gen} is examined for the UL 2017 MC, while in Fig. 48 - 50 of Appendix D for the rest MC as well. In more detail, the response is studied for jets with $p_T^{Gen} > 15\ GeV$, as corrections start to break down for lower momenta and most importantly no physics analysis uses PF jets with $p_T < 15\ GeV$ anyway, as they are not considered physics jets that come from the hard scattering of protons. In these plots the response is shown superimposed for all four different detector areas in terms of pseudo rapidity. As far as the second kind of closure test that will be examined, in the right-hand side plot in Fig. 28 the response as a function of jet η for $p_T^{Gen} > 15\ GeV$ is studied. It can be observed that for AK4 CHS jets the response after the application of all baseline corrections is equal to unity within one percent error both as a function of p_T^{Gen} and η for all four UL simulations, with the exception of a small deviation of approximately up to 1.5 – 2% in the outter endcap region.

Moving to AK8 CHS jets, which can be found in Fig. 29 for the UL 2017 simulation

and Fig. 51 - 53 for the rest MC, their corrected response seems to deviate from unity as a function of p_T^{Gen} at low momenta ($p_T^{Gen} \lesssim 25\ GeV$) and for the regions where there is no tracker coverage ($|\eta| > 2.5$). More specifically, the response in this area is < 1, namely part of the information regarding the jet energy is lost at the reconstructed level. Since, however, the AK8 jets that are used in physics analyses are of high p_T, usually greater than around 100 GeV, such non closures in the low p_T spectrum can be safely ignored.

As far as AK4 PUPPI jets are concerned, the corrected response for the UL 2017 sample in Fig. 30 seems to be very good, with only some minor deviations of $\approx 2\%$ in the forward region. Similarly, the response for the UL 2018 MC in Fig. 56 has a healthy behavior without any large fluctuations from unity. On the other hand, for the 2016 simulations in Fig. 54 and 55 the closure is worse, leading to deviations of up to 3% in the outter endcap and 7% in the forward region. The origin of this deterioration in the 2016 simulations could be possibly found on a slightly more inefficient PUPPI tune that was used in these particular MC files. Generally, it can be deduced that PUPPI jets have a more unstable behavior regarding the MC truth corrections compared to CHS jets.

Finally, regarding the last jet collection to be examined, namely AK8 PUPPI jets, their response can be found in Fig. 31 for the UL 2017 simulation and Fig. 57 - 59 for the 2016 and 2018 samples. As expected, the behavior is similar to AK4 PUPPI jets, with the deviations from unity being present at a little higher momenta ($p_T^{Gen} \lesssim 30 - 35\ GeV$), which is not alarming since, as stated before, only for high energy AK8 jets analyses are interested of.

Another important thing that is worth mentioning is the fact that as a function of η the responses for all jet collections do not show any deviations from unity within one percent error, as it is the case when the closure is examined as a function of the generated jet momentum. This is due to the fact that for the derivation of these particular response plots all jets with $p_T^{Gen} > 15\ GeV$ are taken into account, thus making the high p_T jets dominating and some small deviations that are noticed in the low energy jets being hidden. For that reason, it would be worth studying the response as a function of η for various low p_T^{Gen} bins in order to check at what values of momentum the closure becomes satisfactory, namely equal to one within one percent error. Such examination can be summarized for all jet collections and UL simulations in Fig. 60 - 75 of Appendix E, where the response is studied in fine bins of p_T^{Gen} from 15 GeV up to 35 GeV.

From these sanity checks it can be concluded that AK4 CHS jets have a very healthy response even at the lowest momenta, while AK8 CHS jets begin to have their response flat and equal to one for $p_T^{Gen} > 23\ GeV$. PUPPI jets, on the other hand, have noticeable deviations from unity for $|\eta| > 3$ that are being corrected for $p_T^{Gen} \gtrsim 23\ GeV$ for AK4 jets

and $p_T^{Gen} \gtrsim 35\ GeV$ for AK8 jets. All in all, most of the remaining instabilities are being observed in the pseudo rapidity region around $\eta \approx 3$, where the calorimeter coverage ends and the forward region begins.

AK4 CHS jets

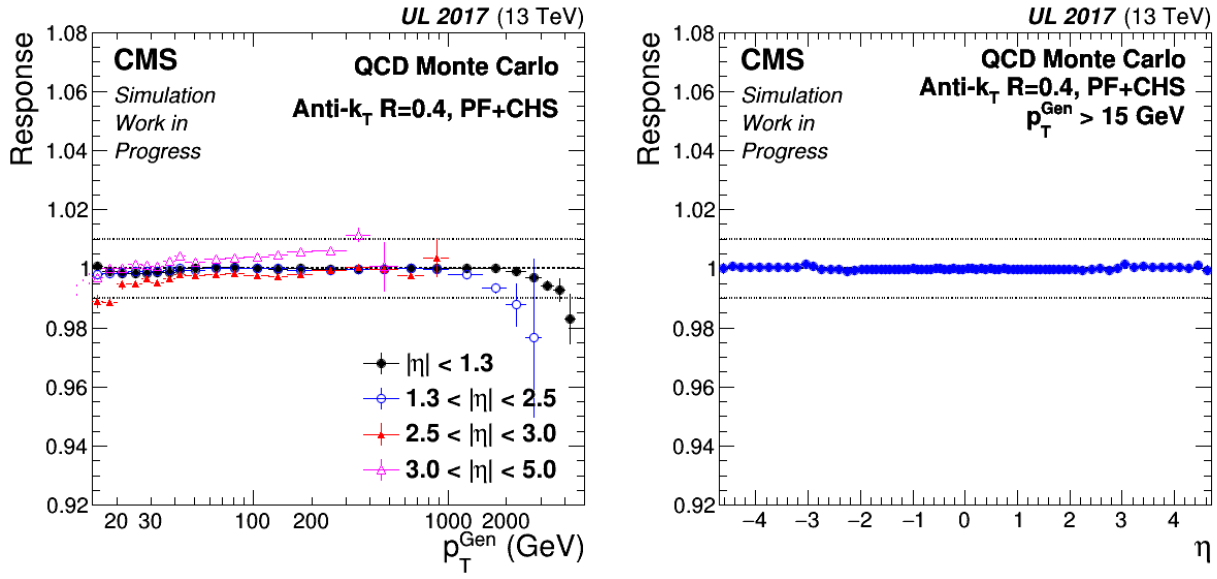

Figure 28: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 CHS from the UL 2017 MC.

AK8 CHS jets

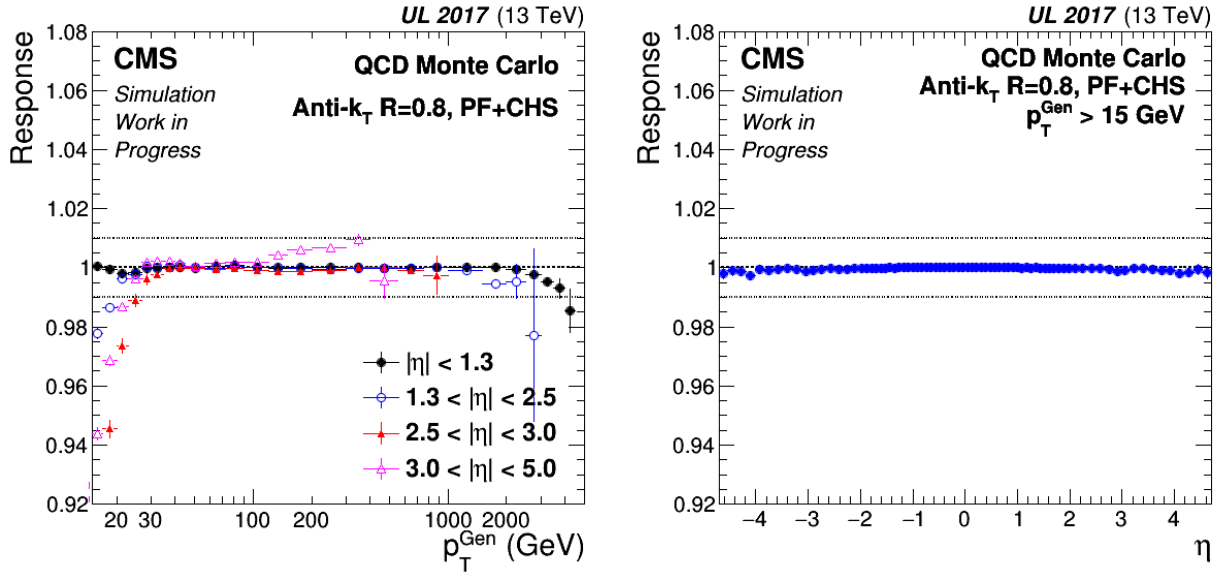

Figure 29: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 CHS from the UL 2017 MC.

AK4 PUPPI jets

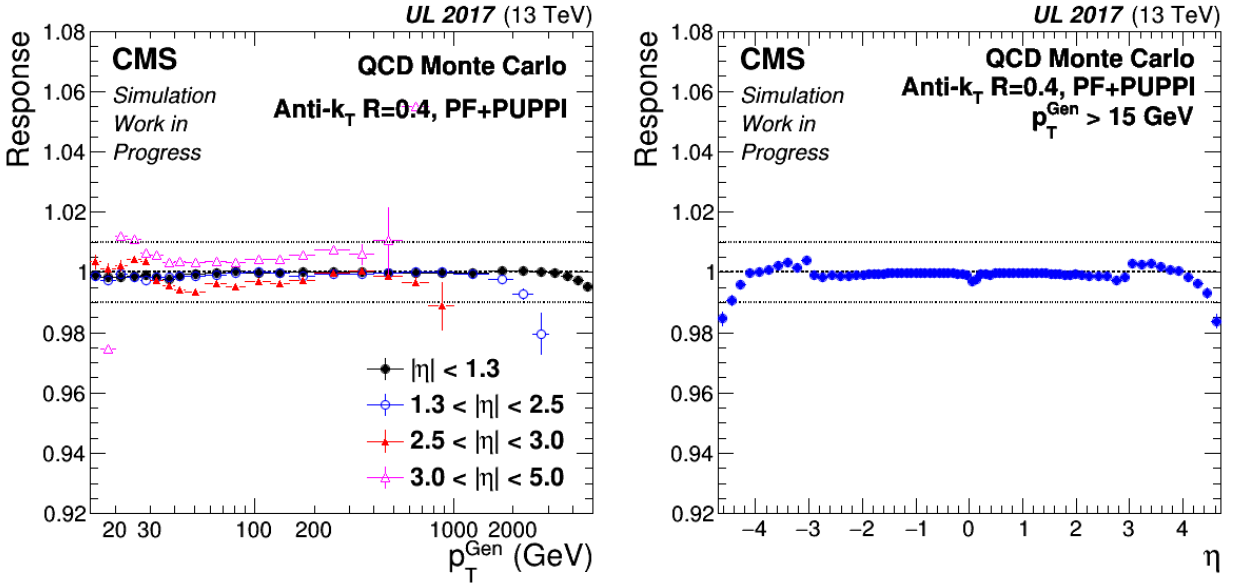

Figure 30: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 PUPPI from the UL 2017 MC.

AK8 PUPPI jets

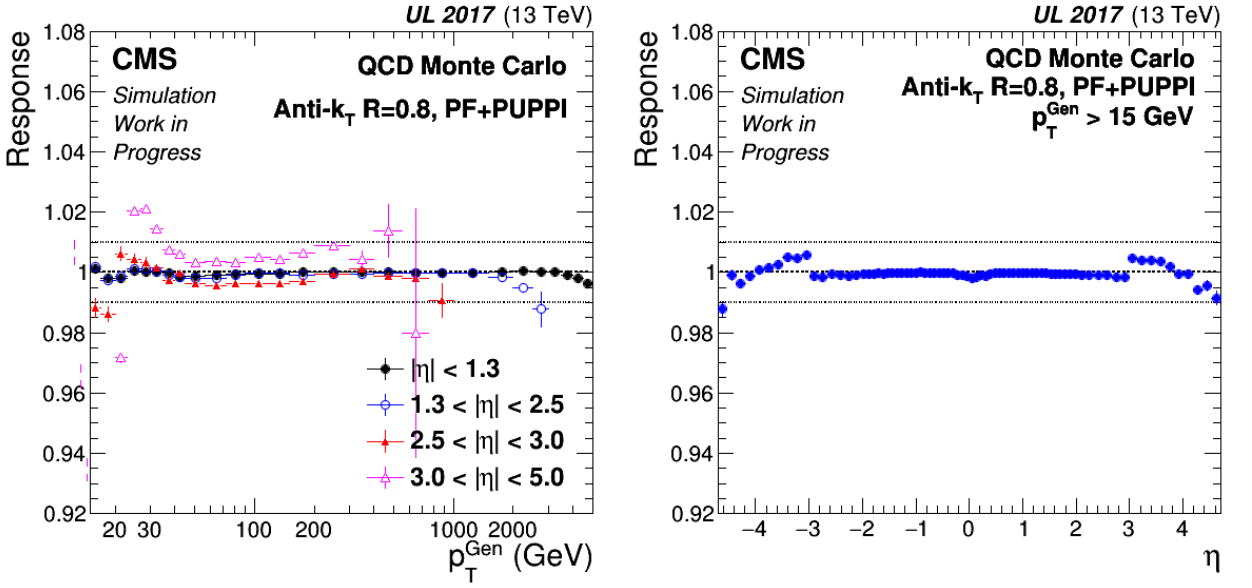

Figure 31: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 PUPPI from the UL 2017 MC.

5 Summary and Outlook

The current methodology that is being used in order to derive the baseline jet energy corrections in the CMS experiment was presented. Jets are the experimental manifestation of high energy quarks and gluons and thus the accurate measurement of their energy is of utmost importance.

Particles in CMS are reconstructed via the Particle Flow (PF) algorithm that combines measurement information from all the sub detector systems, namely the inner tracker, the electromagnetic and hadron calorimeters and the muon system. Afterwards, these particles are clustered into hadronic jets with the anti–k_T algorithm, using either the Charged Hadron Subtracted (CHS) technique in order to remove charged hadrons associated with pileup vertices or the Pile Up Per Particle Identification (PUPPI) algorithm that successfully removes all the excess jet energy casued by particles not originating from the hard scatter.

The baseline jet energy corrections, derived from simulation, comprise the pileup offset corrections that are applied only to CHS jets and the relative and absolute that are applied to all jet collections. Results from the Ultra Legacy (UL) simulation for all years of RunII were presented in Chapter 4 and Appendices B, D, E.

Regarding the first step in the chain of the factorized corrections, the offset energy for CHS jets with cone radius $R = 0.4$ and $R = 0.8$ is successfully removed as a function of momentum p_T, pseudo rapidity η and number of pileup interactions μ for the 2017 and 2018 simulations. For the 2016 simulation remaining pileup offset within the tracker acceptance is observed, that is under investigation and is most likely associated with different detector conditions in the data collection of 2016, and the APV issue in the tracker.

As far as the relative and absolute corrections are concerned, they aim to equate the jet momentum at the reconstructed level to the respective at the generated by correcting the jet response so it is equal to unity. Results from all RunII UL simulations suggest that the application of these corrections indeed rescale the jet response to one within 1% error in the $p_T - \eta$ phase space of interest, absorbing at the same time any non - closure that the pileup offset correction may had left. The multiplicative correction factors, furthermore, have a smooth behavior and are symmetric between positive and negative pseudo rapidity.

As the CMS experiment prepares for the start of RunIII, where the center-of-mass energy will increase to $\sqrt{s} = 14\ TeV$, the need for high accuracy and precision measurements of jet energy will be even more evident. The increased integrated luminosity will be crucial in helping us search for new physics at higher masses and energies, and testing some of the theoretical models that have been proposed, as the one presented in section 1.4.

However, this increase in the number of proton - proton collisions inside the LHC machine will be bring on average even more pileup interactions in each event and hence degrade the jet response. As a result, jet energy corrections will be the key in providing the best possible treatment for energy miscalculations during data collection, that will allow all physics analyses that use jets in the final state to have accurate datasets.

A Monte Carlo Samples

Monte Carlo Samples		
	Pileup Sample	**Epsilon Pileup Sample**
Non - APV UL 2016	/QCD_Pt-15to7000_TuneCP5_ Flat2018_13TeV_pythia8/RunII Summer19UL16MiniAOD-FlatPU0to70_106X_mcRun2_ asymptotic_v13-v2/MINIAODSIM	/QCD_Pt-15to7000_TuneCP5_ Flat2018_13TeV_pythia8/RunII Summer19UL16MiniAOD-EpsilonPU_106X_mcRun2_ asymptotic_v13-v2/MINIAODSIM
APV UL 2016	/QCD_Pt-15to7000_TuneCP5_ Flat2018_13TeV_pythia8/RunII Summer19UL16MiniAODAPV-FlatPU0to70_106X_mcRun2_ asymptotic_preVFP_v8_ext1-v2/MINIAODSIM	/QCD_Pt-15to7000_TuneCP5_ Flat2018_13TeV_pythia8/RunII Summer19UL16MiniAODAPV-EpsilonPU_106X_mcRun2_ asymptotic_preVFP_v8_ext1-v2/MINIAODSIM
UL 2017	/QCD_Pt-15to7000_TuneCP5_ Flat_13TeV_pythia8/RunII Summer19UL17MiniAOD-FlatPU0to70_106X_mc2017_ realistic_v6_ext2-v3/MINIAODSIM	/QCD_Pt-15to7000_TuneCP5_ Flat_13TeV_pythia8/RunII Summer19UL17MiniAOD-EpsilonPU_106X_mc2017_ realistic_v6_ext2-v2/MINIAODSIM
UL 2018	/QCD_Pt-15to7000_TuneCP5_ Flat_13TeV_pythia8/RunII Summer19UL18MiniAOD-FlatPU0to70_106X_upgrade2018_ realistic_v11_L1v1-v2/MINIAODSIM	/QCD_Pt-15to7000_TuneCP5_ Flat_13TeV_pythia8/RunII Summer19UL18MiniAOD-EpsilonPU_pilot_106X_upgrade2018_ realistic_v11_L1v1-v2/MINIAODSIM

Table 3: Monte Carlo samples used for the derivation of offset corrections.

Pileup Monte Carlo Samples	
Non - APV UL 2016	/QCD_Pt-15to7000_TuneCP5_Flat2018_13TeV_pythia8/RunII Summer19UL16MiniAOD-106X_mcRun2_asymptotic_v13-v2/MINIAODSIM
APV UL 2016	/QCD_Pt-15to7000_TuneCP5_Flat2018_13TeV_pythia8/RunII Summer19UL16MiniAODAPV-106X_mcRun2_asymptotic_preVFP_v8_ ext1-v2/MINIAODSIM
UL 2017	/QCD_Pt-15to7000_TuneCP5_Flat_13TeV_pythia8/RunII Summer19UL17MiniAOD-106X_mc2017_realistic_v6_ext2-v2/MINIAODSIM
UL 2018	/QCD_Pt-15to7000_TuneCP5_Flat_13TeV_pythia8/RunII Summer19UL18MiniAOD-pilot_106X_upgrade2018_realistic_v11_ L1v1-v2/MINIAODSIM

Table 4: Monte Carlo samples used for the derivation of relative and absolute corrections.

B Examination of Pileup Offset

In this section the two dimensional distribution of the reconstructed versus true number of pileup interactions is shown in Fig. 32.

Afterwards, the offset quantity is examined for the Non - APV UL 2016, APV UL 2016 and UL 2018 simulations as a function of p_T^{Gen}, η and μ, similar to how it was presented for the UL 2017 simulation in Chapters 4.3.1 and 4.3.2. In Fig. 33 - 38 the uncorrected offset can be found for AK4 and AK8 CHS jets, while in Fig. 39 - 44 the corrected offset is examined for the same jet collections.

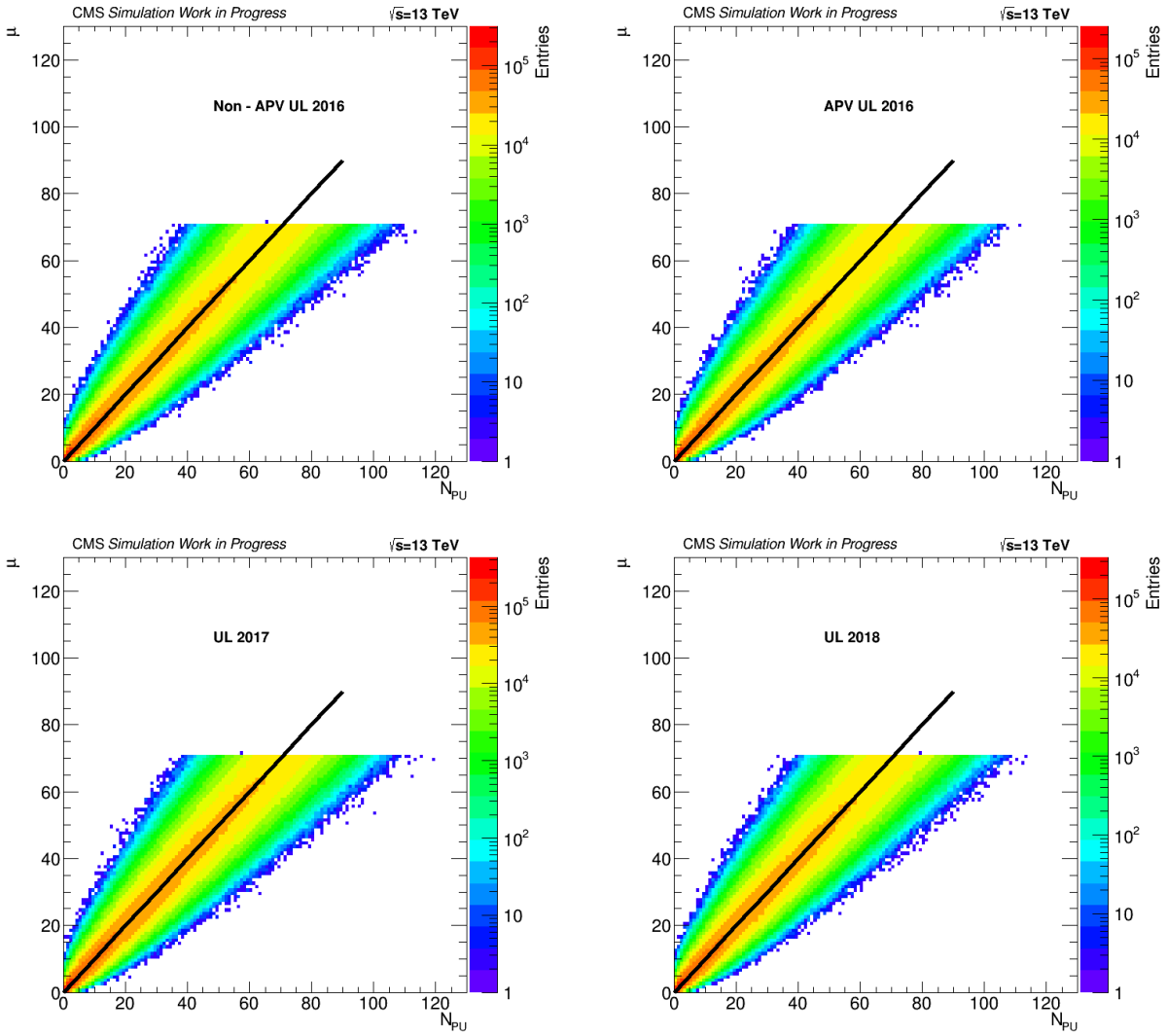

Figure 32: Reconstructed (N_{PU}) versus true (μ) number of pileup interactions for all UL MC samples. The diagonal is illustrated with a black line.

Uncorrected Offset

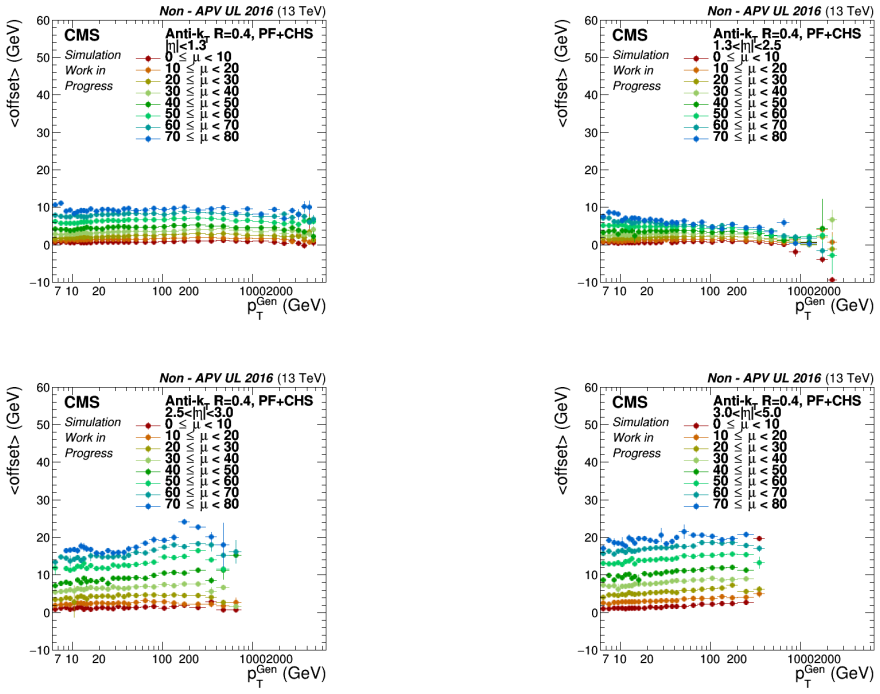

Figure 33: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the Non - APV UL 2016 MC.

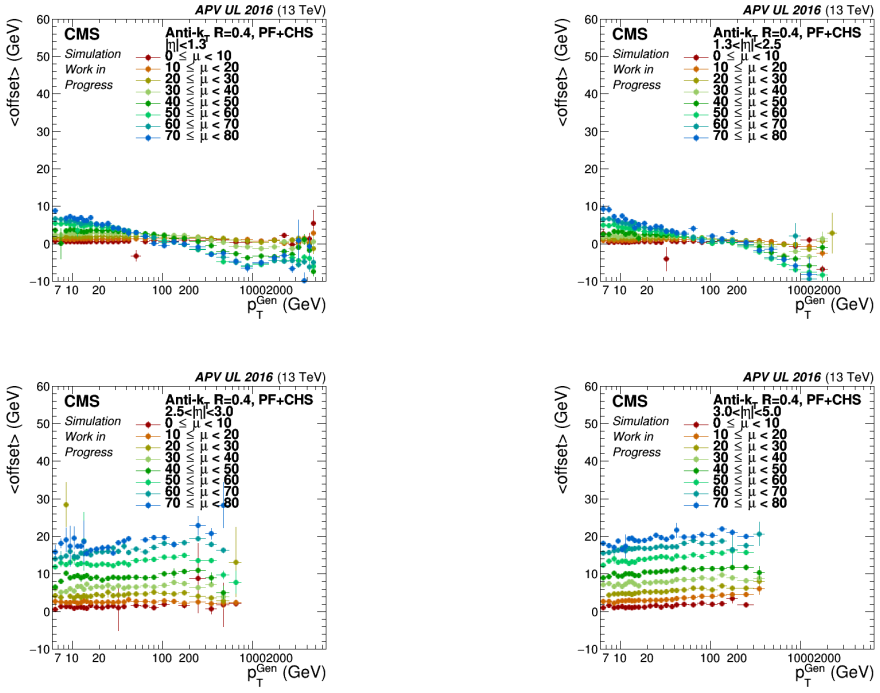

Figure 34: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the APV UL 2016 MC.

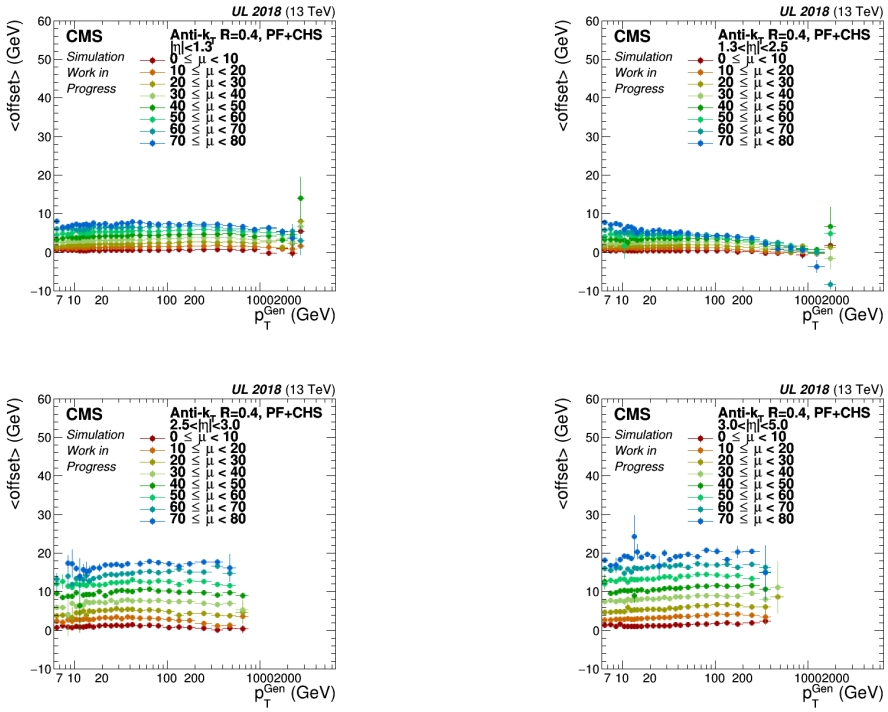

Figure 35: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2018 MC.

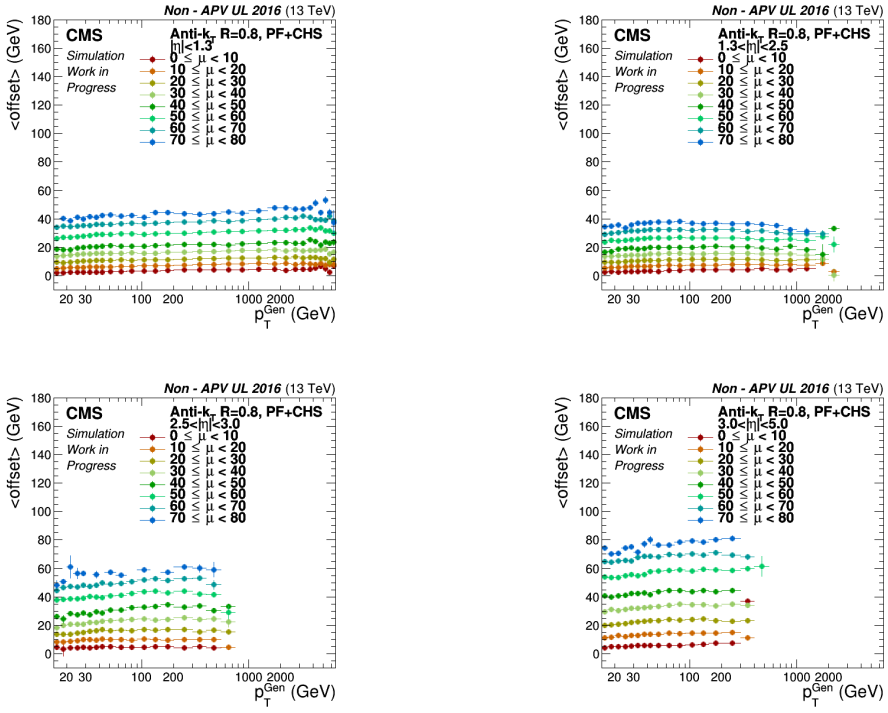

Figure 36: Mean Offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the Non - APV UL 2016 MC.

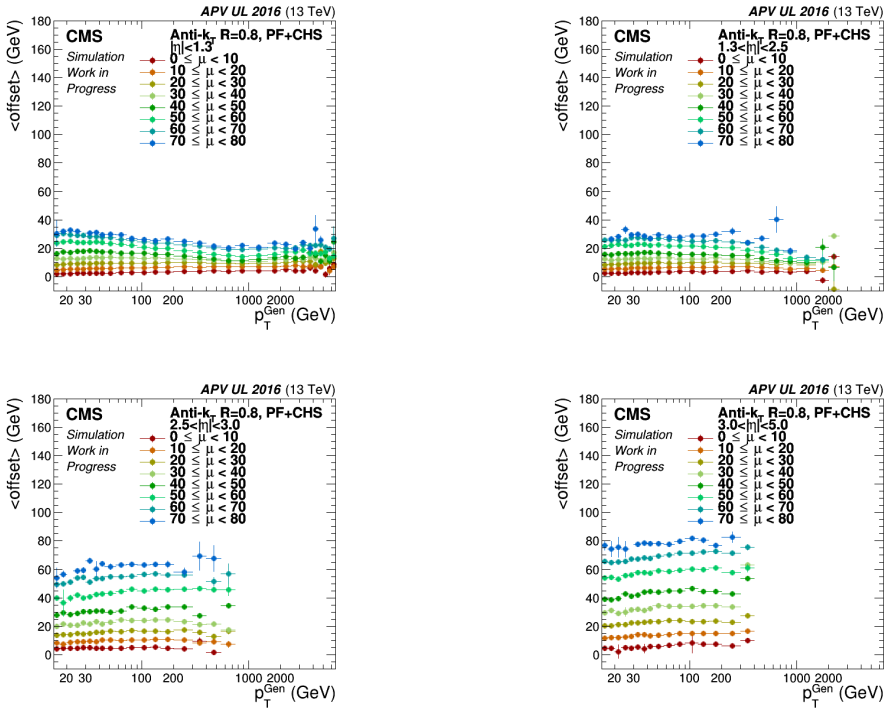

Figure 37: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the APV UL 2016 MC.

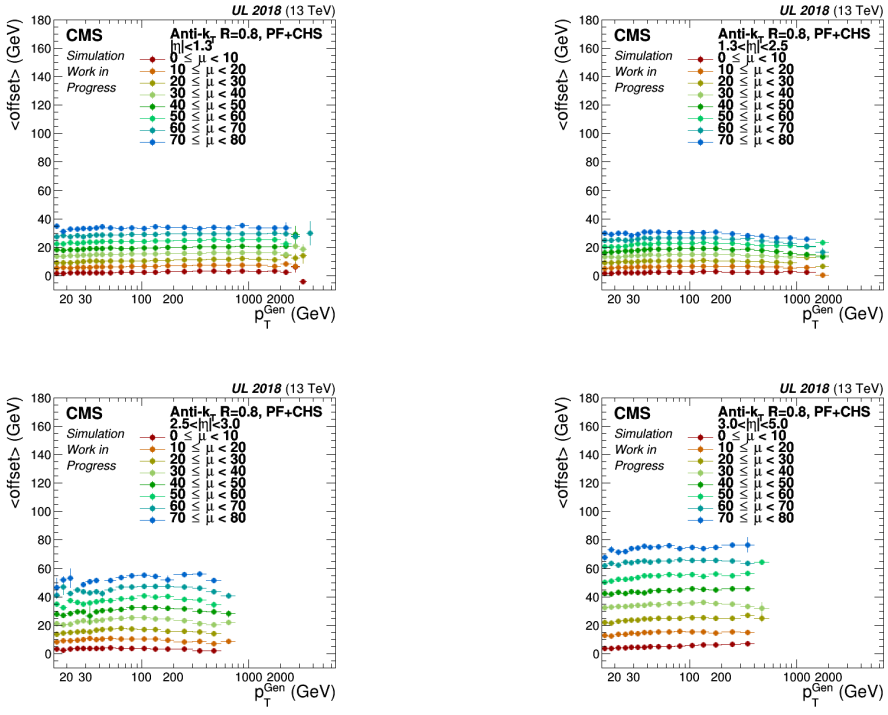

Figure 38: Mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the UL 2018 MC.

Corrected Offset

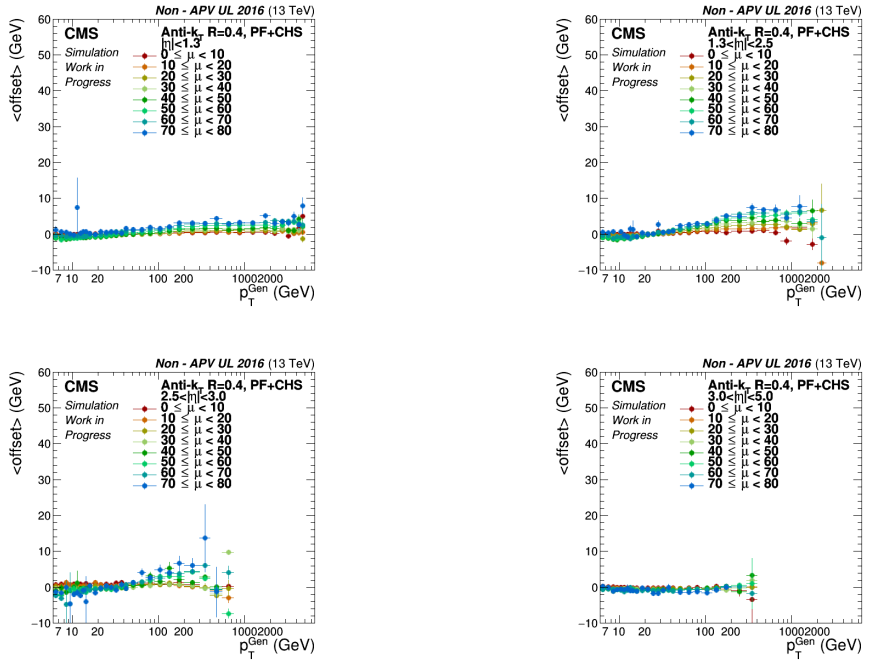

Figure 39: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the Non - APV UL 2016 MC.

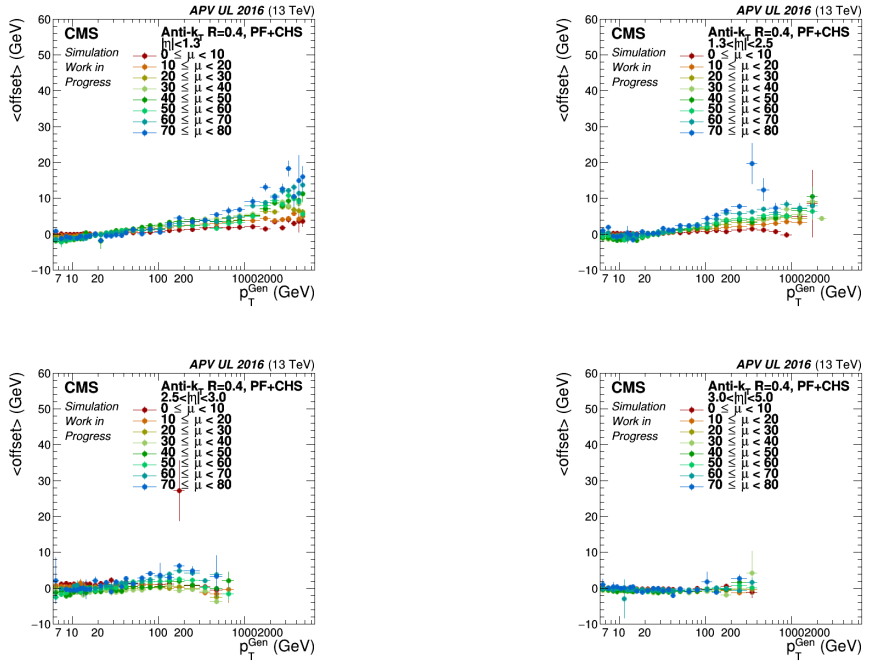

Figure 40: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the APV UL 2016 MC.

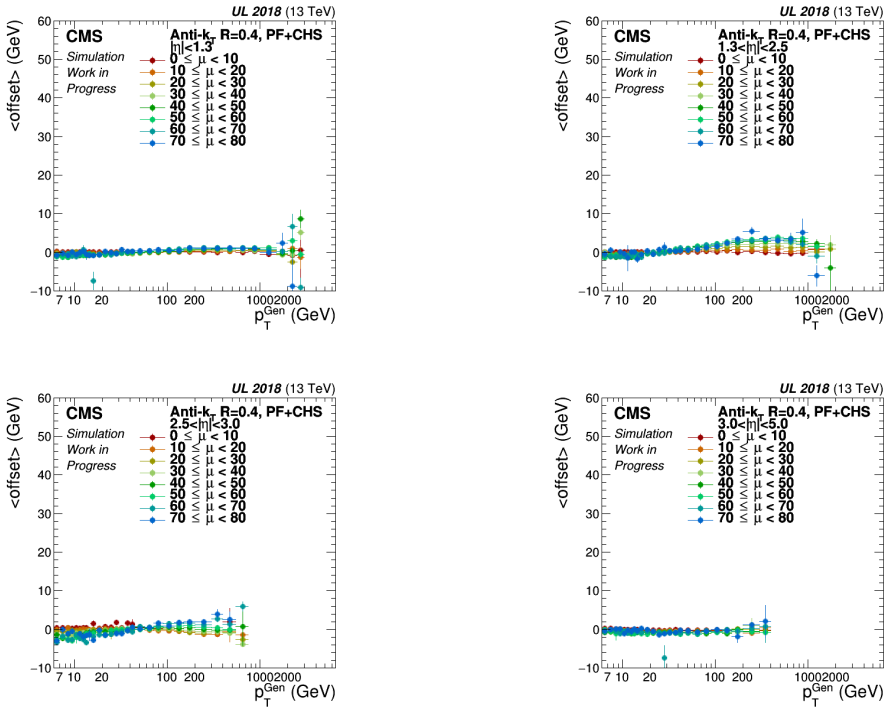

Figure 41: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2018 MC.

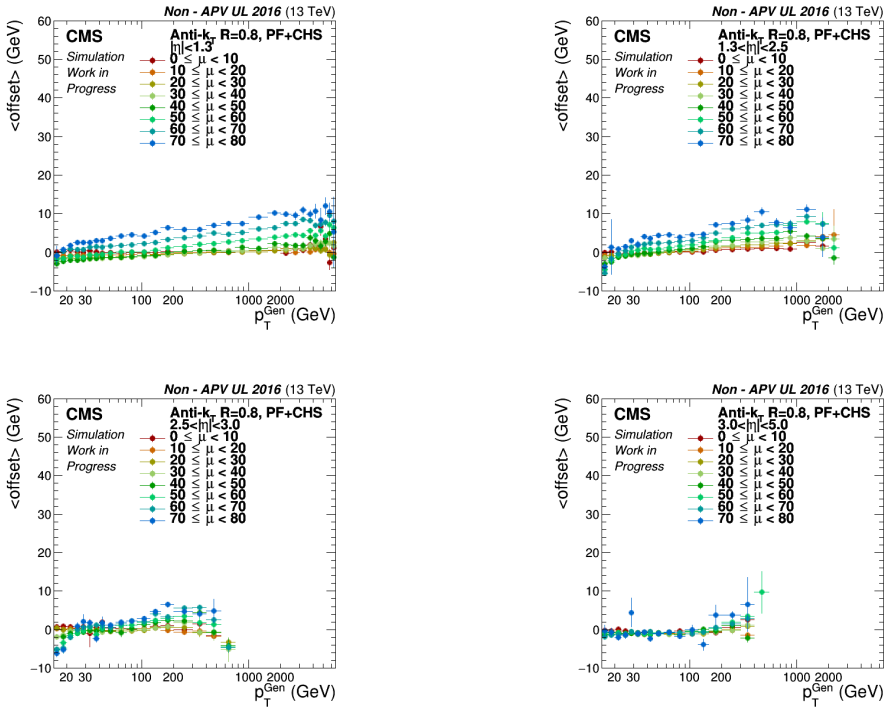

Figure 42: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the Non - APV UL 2016 MC.

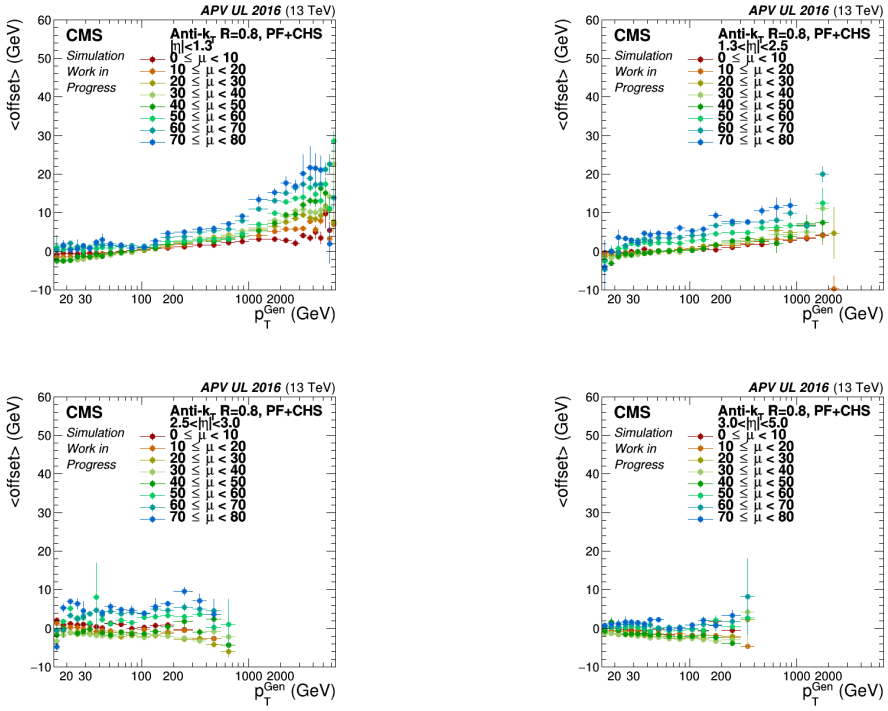

Figure 43: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the APV UL 2016 MC.

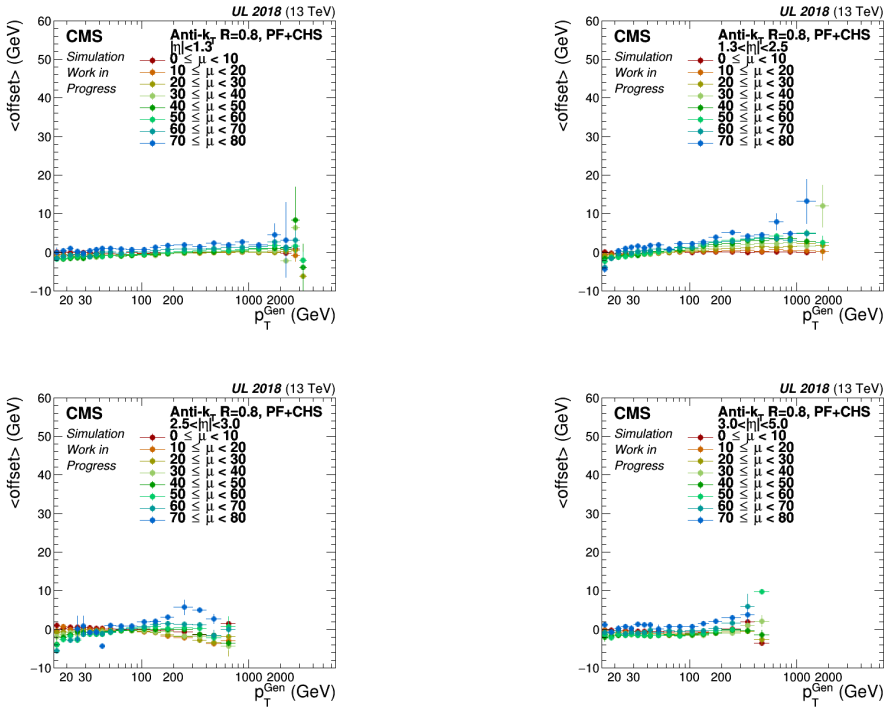

Figure 44: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK8 CHS jets from the UL 2018 MC.

C Comparison among Parametric Functions for the Derivation of Offset Corrections

In order to examine whether the three parametric functions that were presented in Chapter 4.3.2 (Eq. 86, 87, 88) are efficient enough to remove the offset energy caused by pileup, the corrected offset as a function of p_T^{Gen}, μ and η is presented for each case, using AK4 CHS jets from the UL 2017 simulation. For convenience purposes, function g (Eq. 87) will be referred to as "Complex", function h (Eq. 88) as "Simple" and function f which was selected to be used for the final derivation of corrections (Eq. 86) as "Semi - Simple".

It is worth mentioning that for the "Complex" and "Semi - Simple" functions the fitting ranges are:

- ✓ $8\ GeV < p_T^{Reco} < 6500\ GeV$
- ✓ $0\ GeV < \rho < 70\ GeV$

while for the "Simple" function they are constrained to the ranges:

- ✓ $30\ GeV < p_T^{Reco} < 100\ GeV$
- ✓ $0\ GeV < \rho < 70\ GeV$

because this parameterization does not give good results outside the p_T range of 30-100 GeV.

In Fig. 45, 46, 47 the corrected offset for the "Semi - Simple", "Complex" and "Simple" parameterization can be seen respectively. The "Simple" function successfully removes the offset in the region $30\ GeV < p_T^{Gen} < 100\ GeV$, where the reconstructed p_T is fitted, but for lower momenta remaining offset is detected and for higher an overcorrection takes place. As far as the comparison between "Complex" and "Semi - Simple" functions is concerned, the corrected offset in both cases is generally in agreement with each other. However, the "Complex" function leaves a sizeable amount of offset energy remaining at higher generated jet momenta and $|\eta| < 2.5$.

Due to the above reasons, the "Semi - Simple" function is preferred to be used when deriving the offset corrections for the UL simulations.

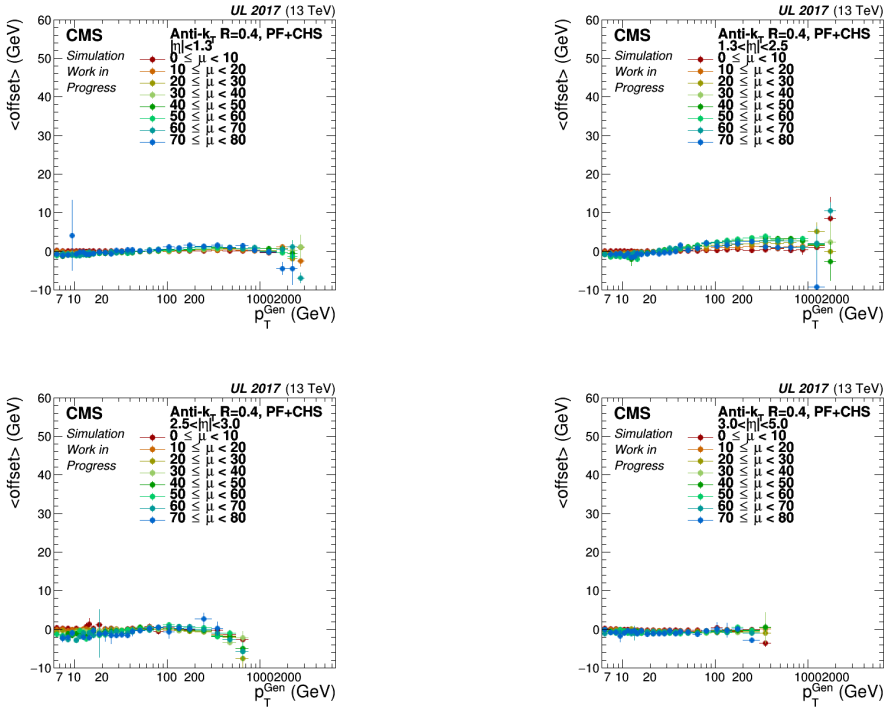

Figure 45: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2017 MC. "Semi - Simple" parametric function is used.

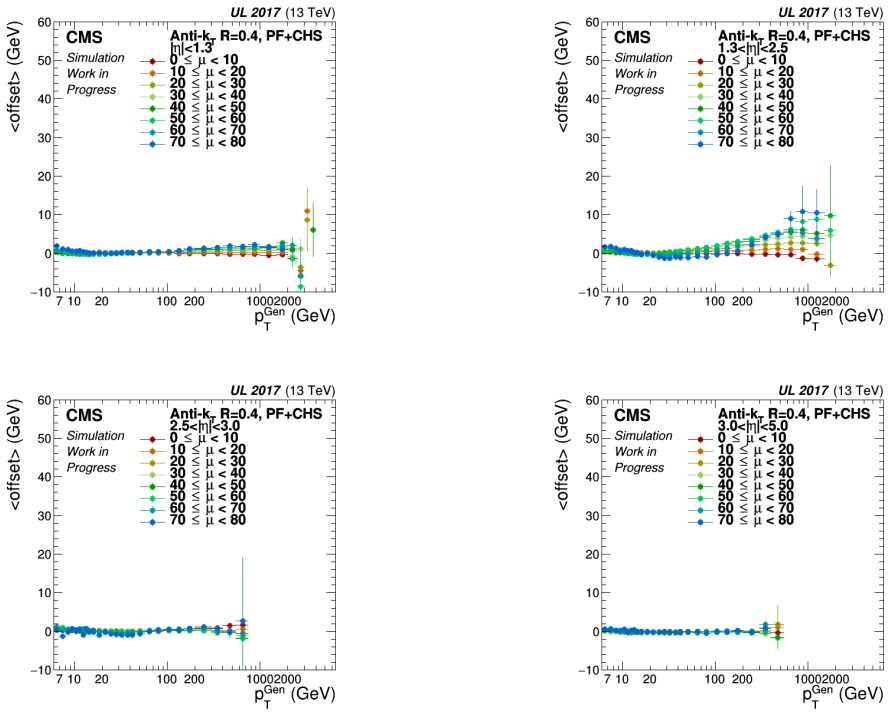

Figure 46: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2017 MC. "Complex" parametric function is used.

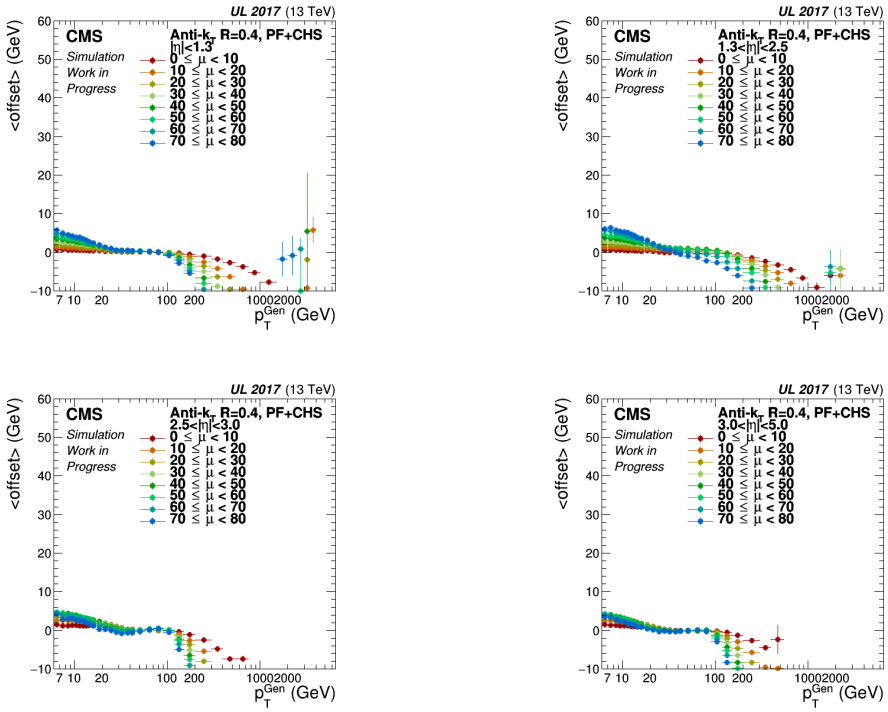

Figure 47: Corrected mean offset as a function of p_T^{Gen} for various true number of pileup interactions (μ) bins and detector areas, for AK4 CHS jets from the UL 2017 MC. "Simple" parametric function is used.

D Examination of Corrected Response

In this section the corrected response as a function of p_T^{Gen} and η will be presented for all jet collections from the Non - APV UL 2016, APV UL 2016 and UL 2018 simulations.

AK4 CHS jets

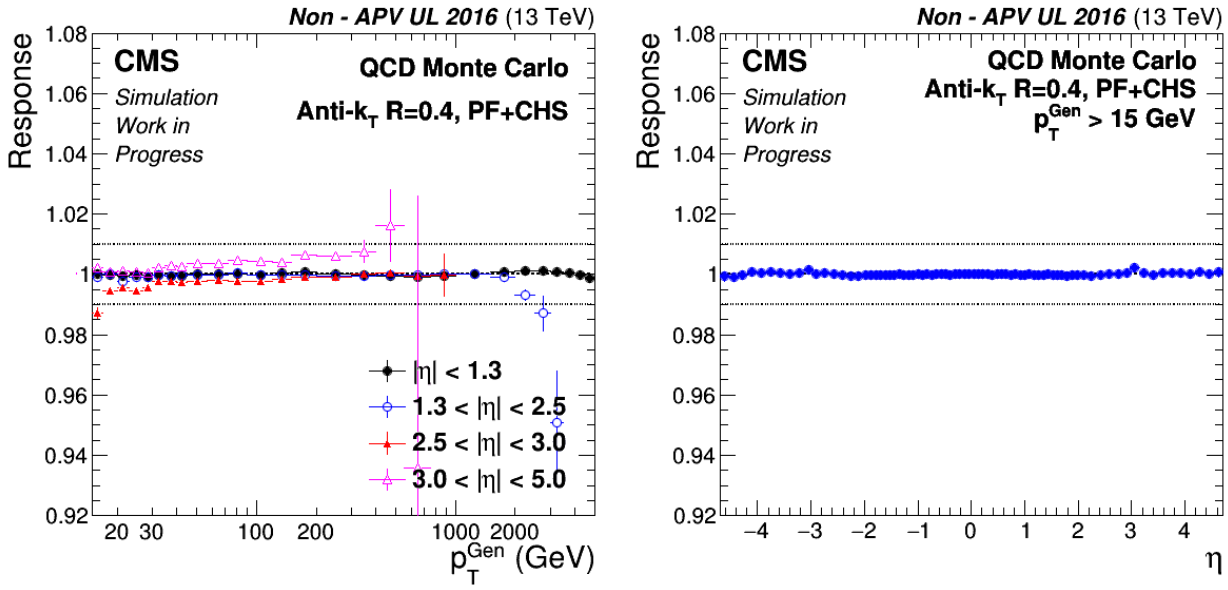

Figure 48: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 CHS from the Non - APV UL 2016 MC.

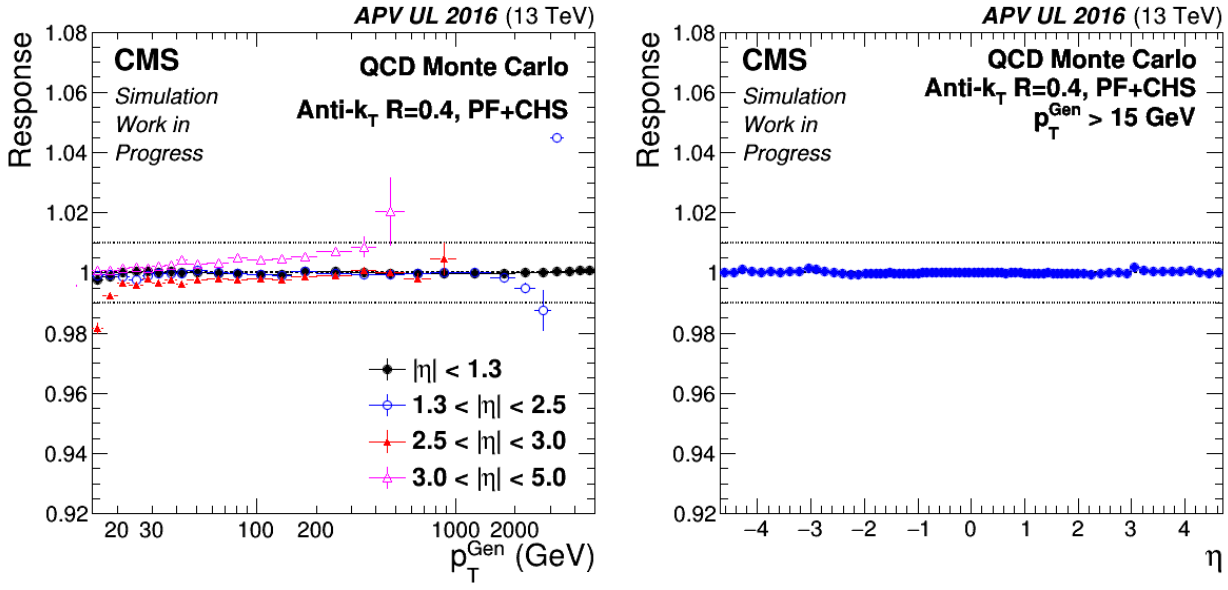

Figure 49: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 CHS from the APV UL 2016 MC.

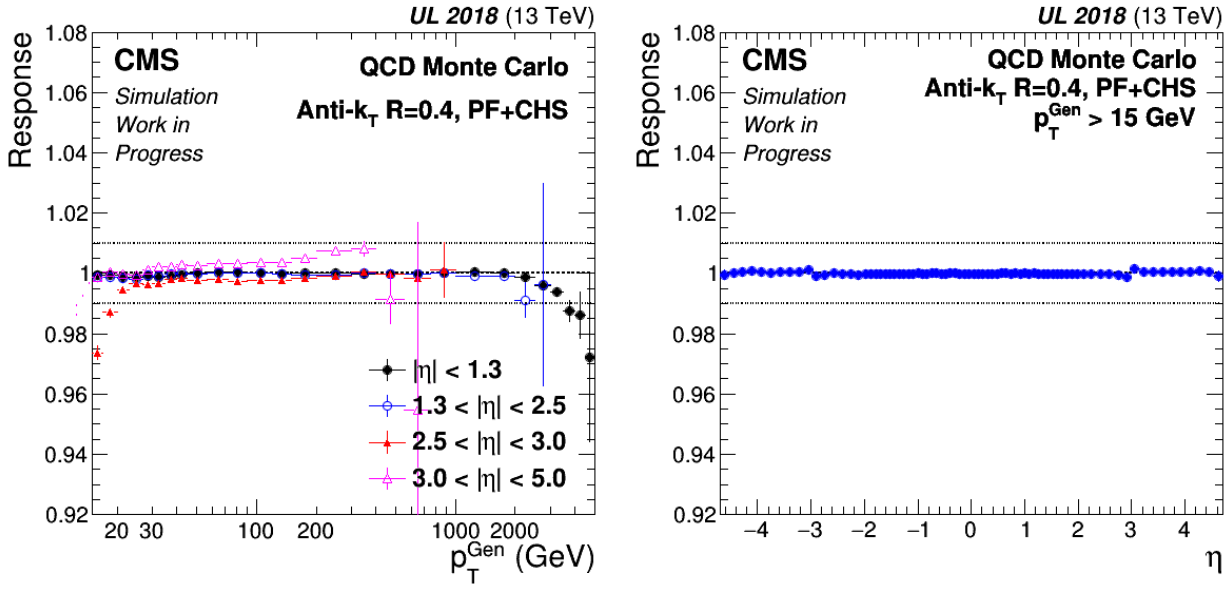

Figure 50: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 CHS from the UL 2018 MC.

AK8 CHS jets

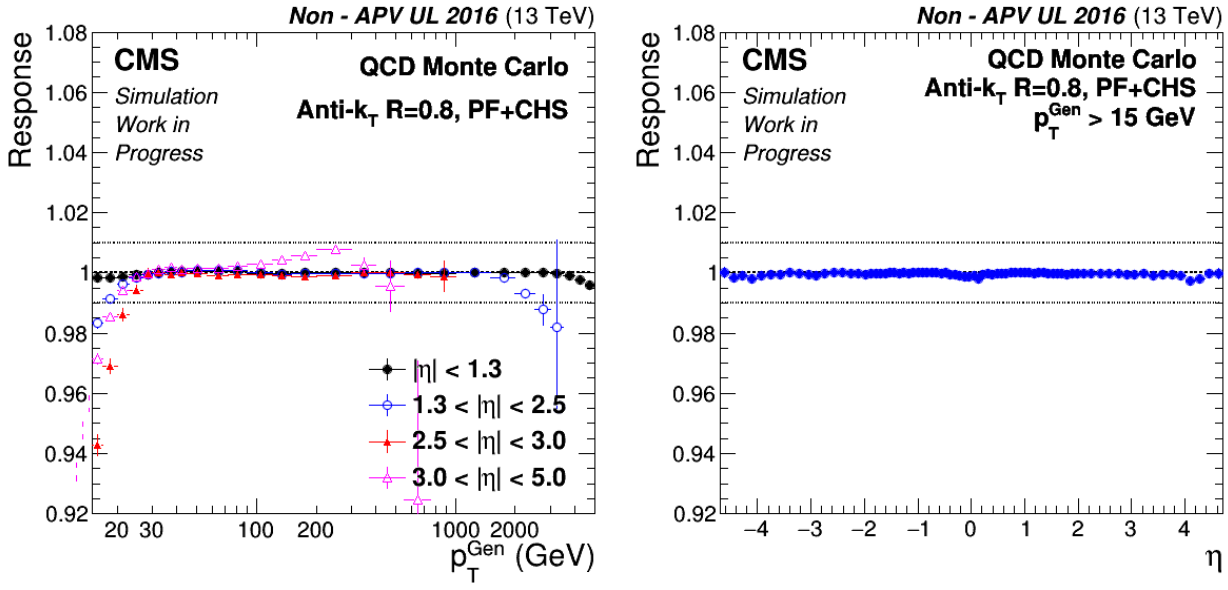

Figure 51: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 CHS from the Non - APV UL 2016 MC.

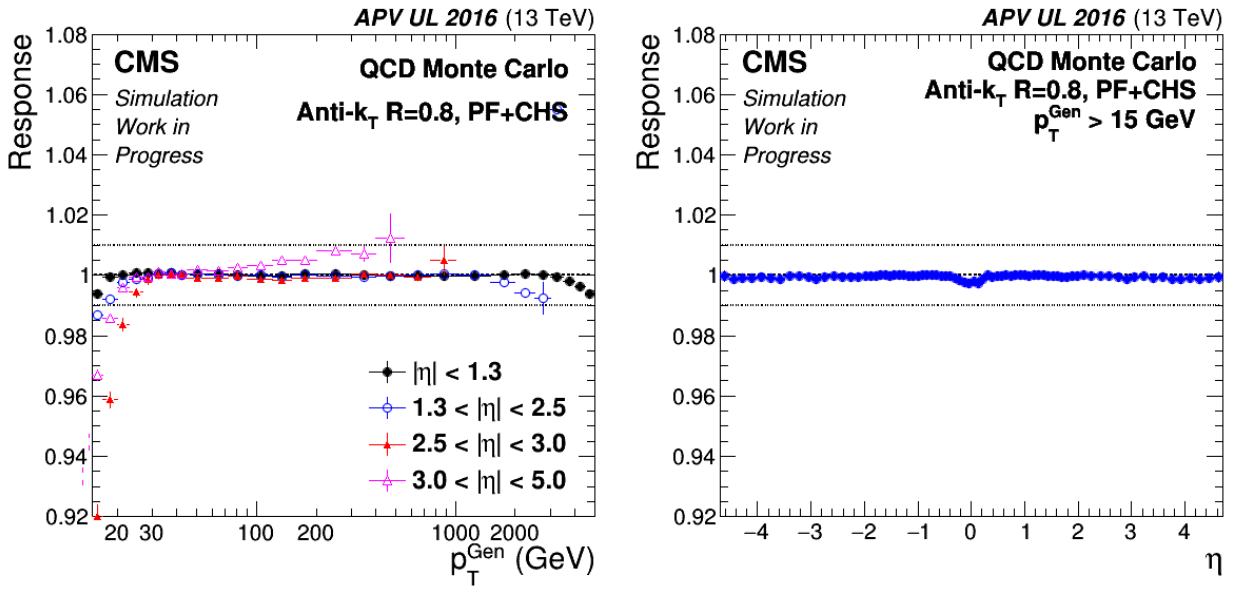

Figure 52: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 CHS from the APV UL 2016 MC.

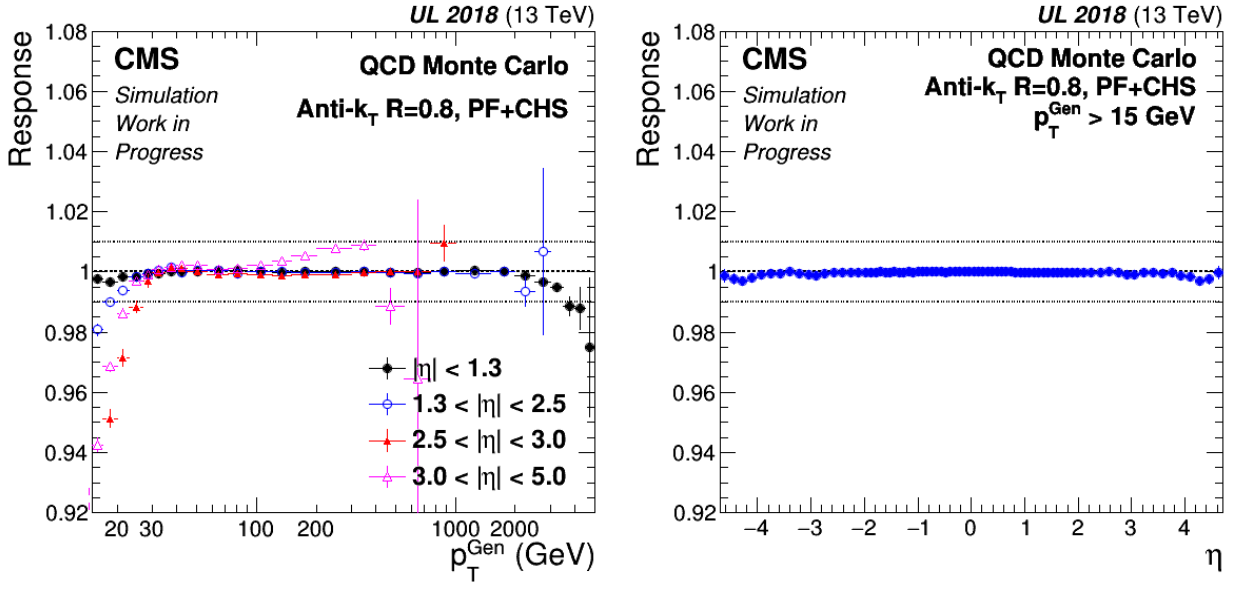

Figure 53: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 CHS from the UL 2018 MC.

AK4 PUPPI jets

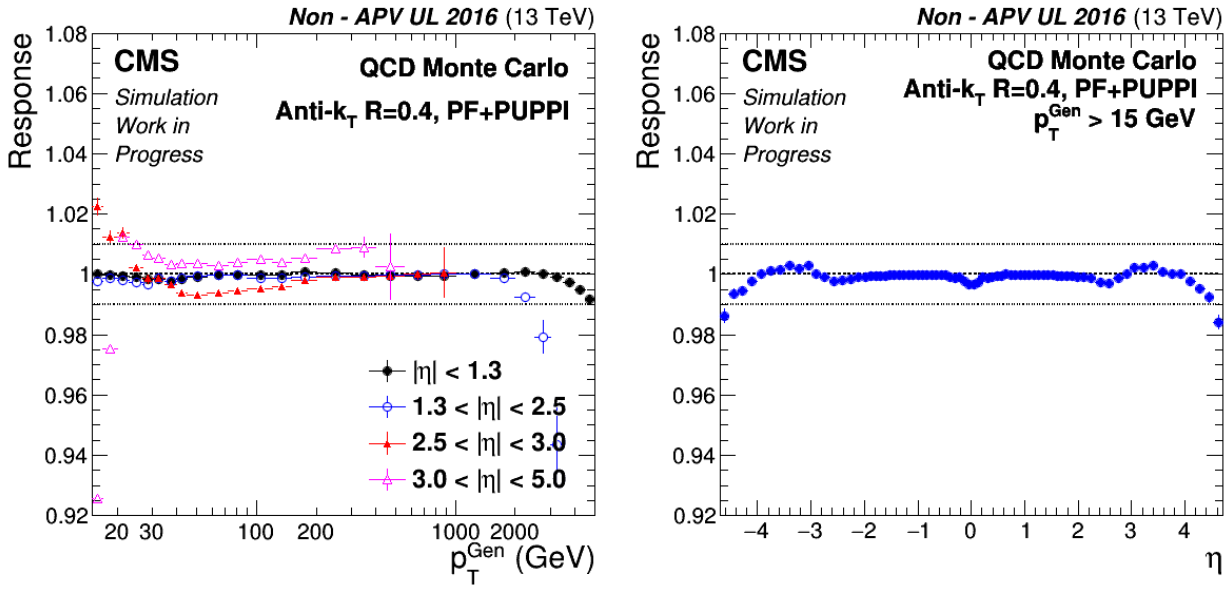

Figure 54: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 PUPPI from the Non - APV UL 2016 MC.

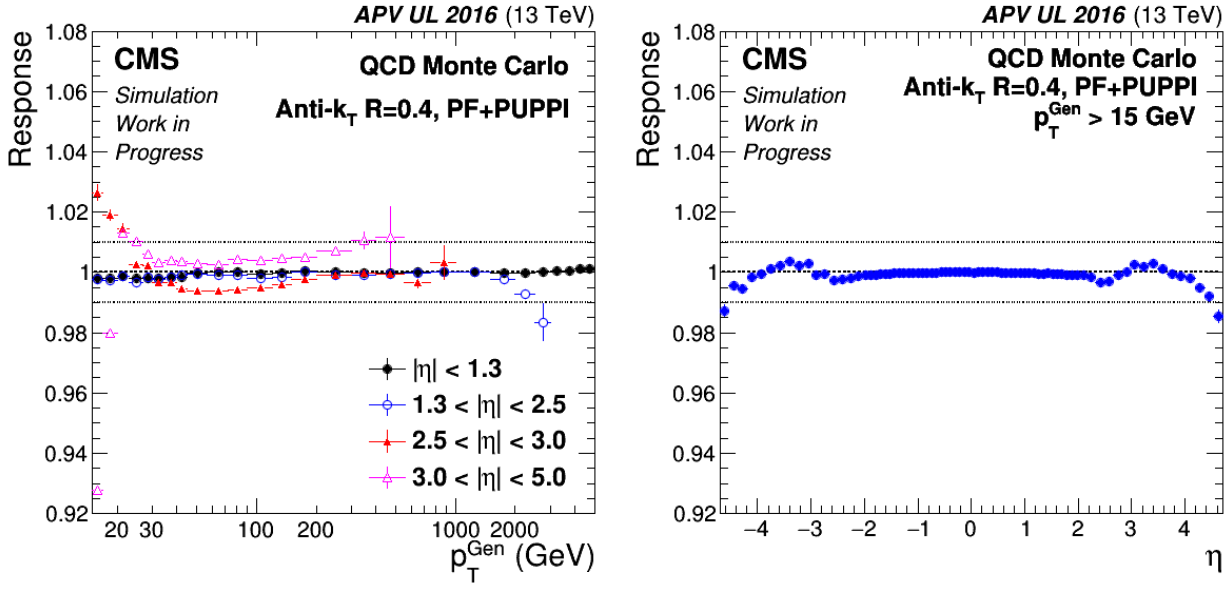

Figure 55: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 PUPPI from the APV UL 2016 MC.

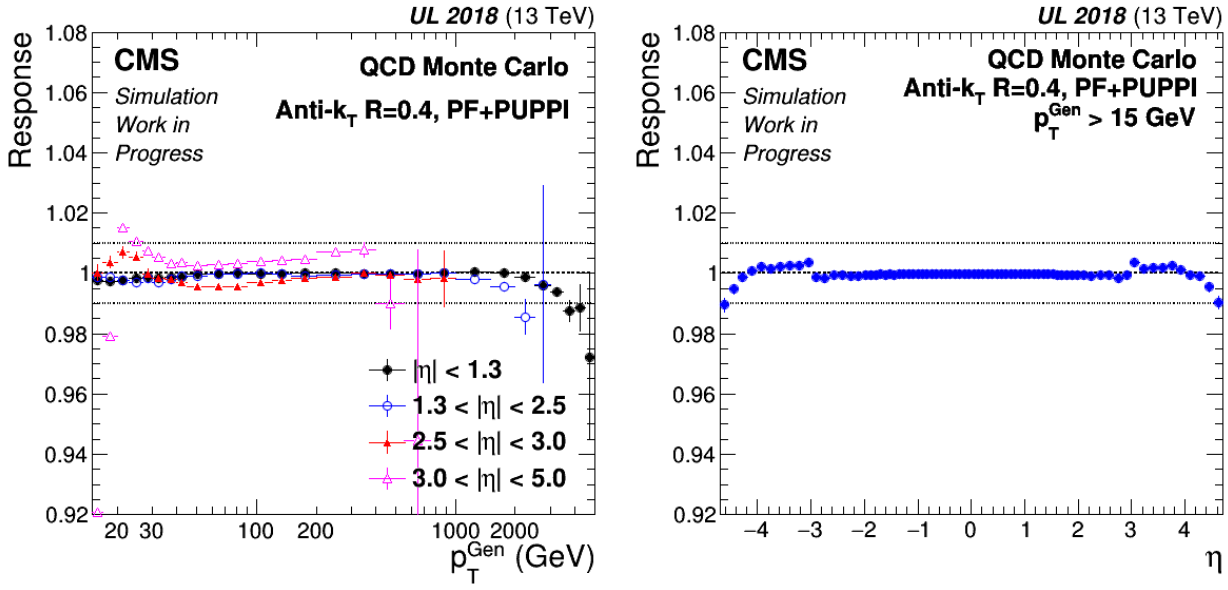

Figure 56: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK4 PUPPI from the UL 2018 MC.

AK8 PUPPI jets

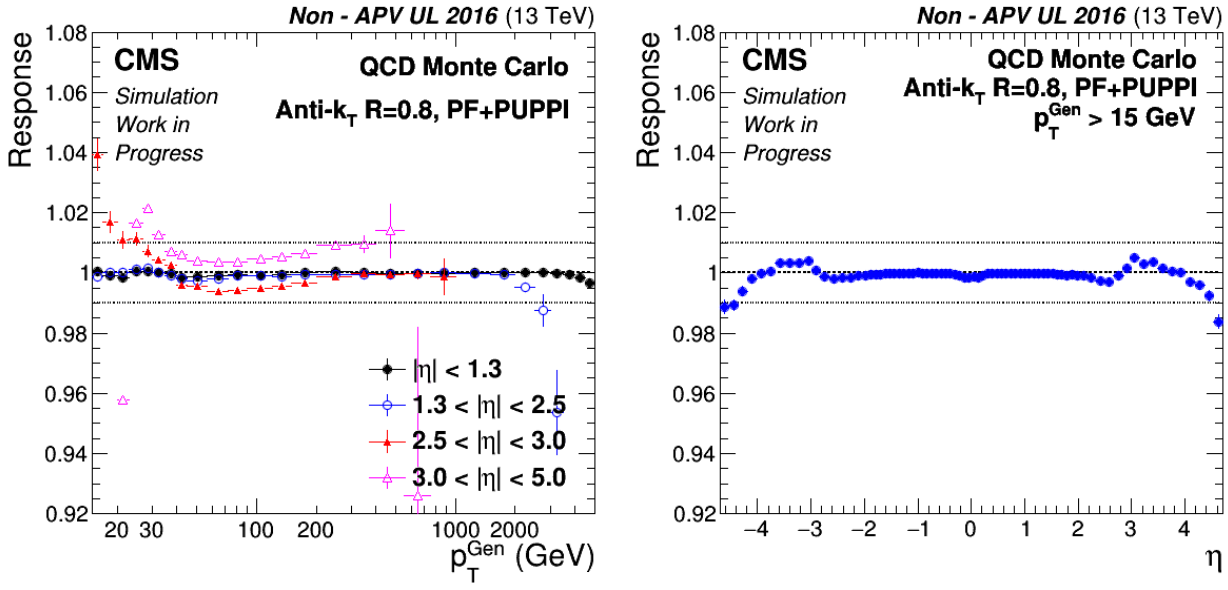

Figure 57: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 PUPPI from the Non - APV UL 2016 MC.

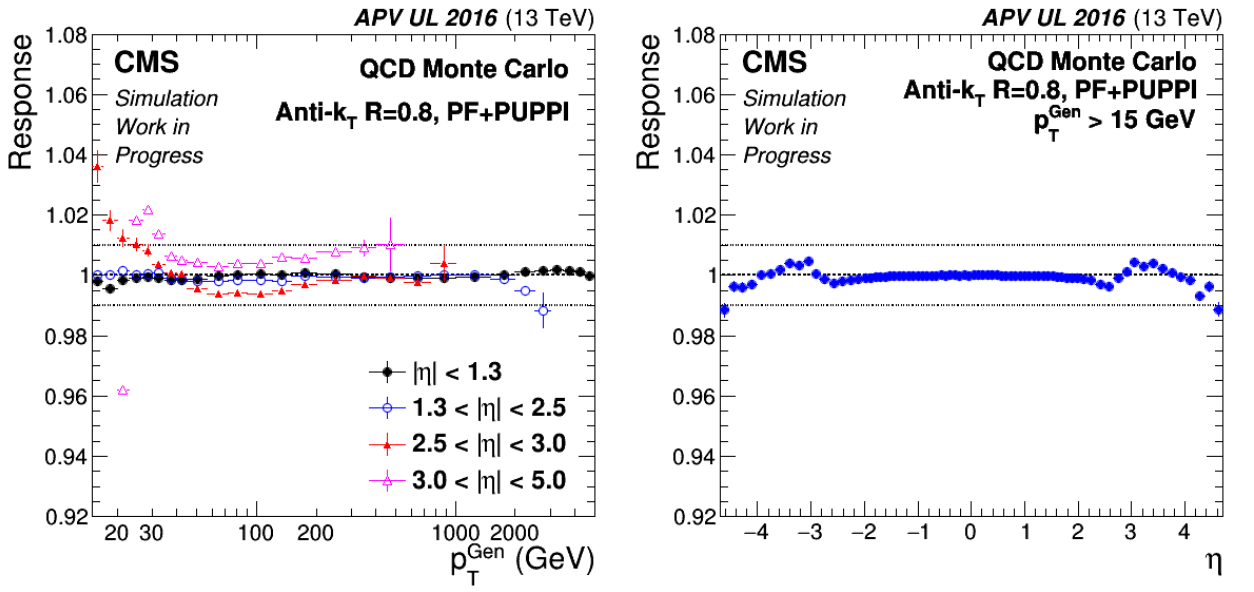

Figure 58: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 PUPPI from the APV UL 2016 MC.

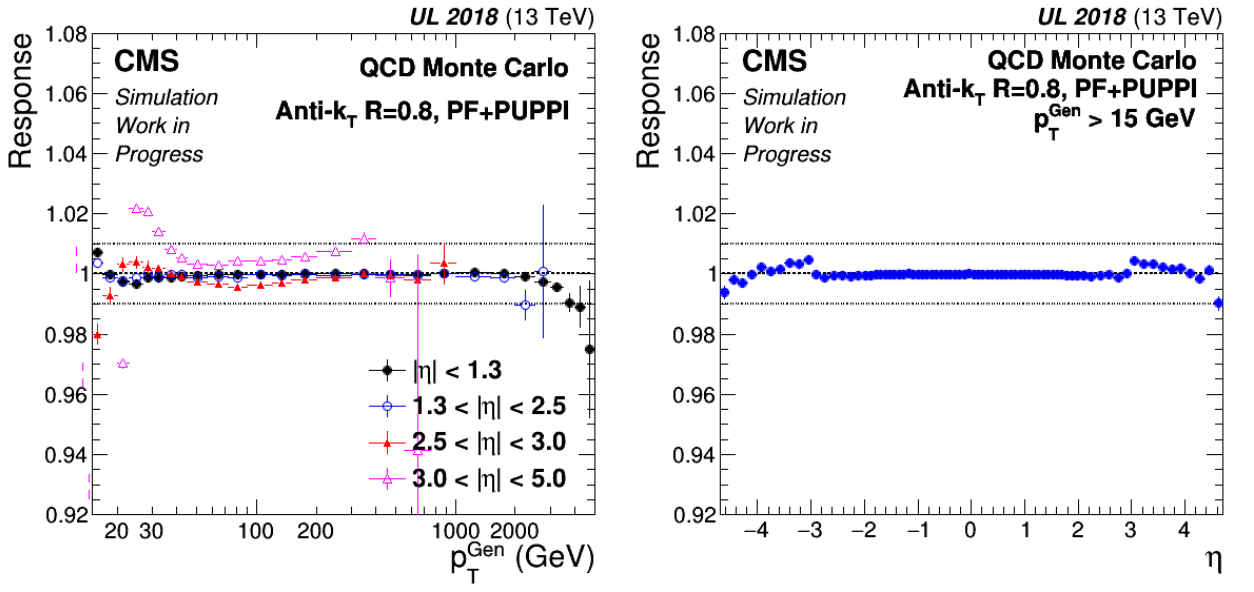

Figure 59: Corrected response as a function of p_T^{Gen} (left) and jet η (right) for AK8 PUPPI from the UL 2018 MC.

E Examination of Corrected Response for detailed p_T^{Gen} bins

AK4 CHS jets

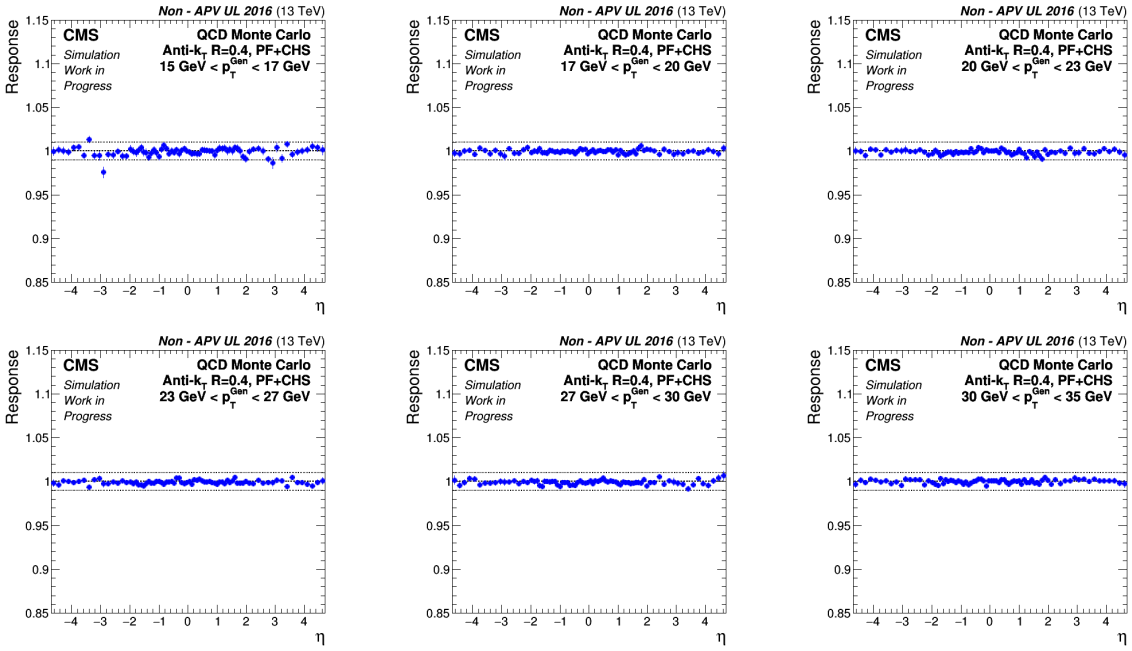

Figure 60: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 CHS jets from the Non - APV UL 2016 MC.

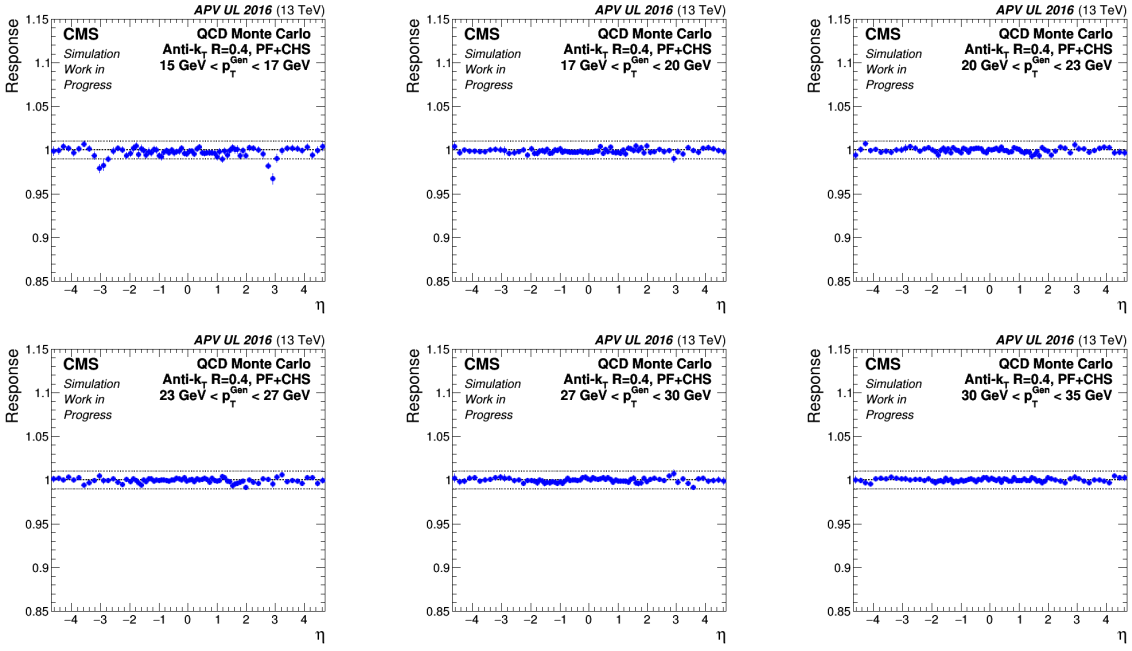

Figure 61: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 CHS jets from the APV UL 2016 MC.

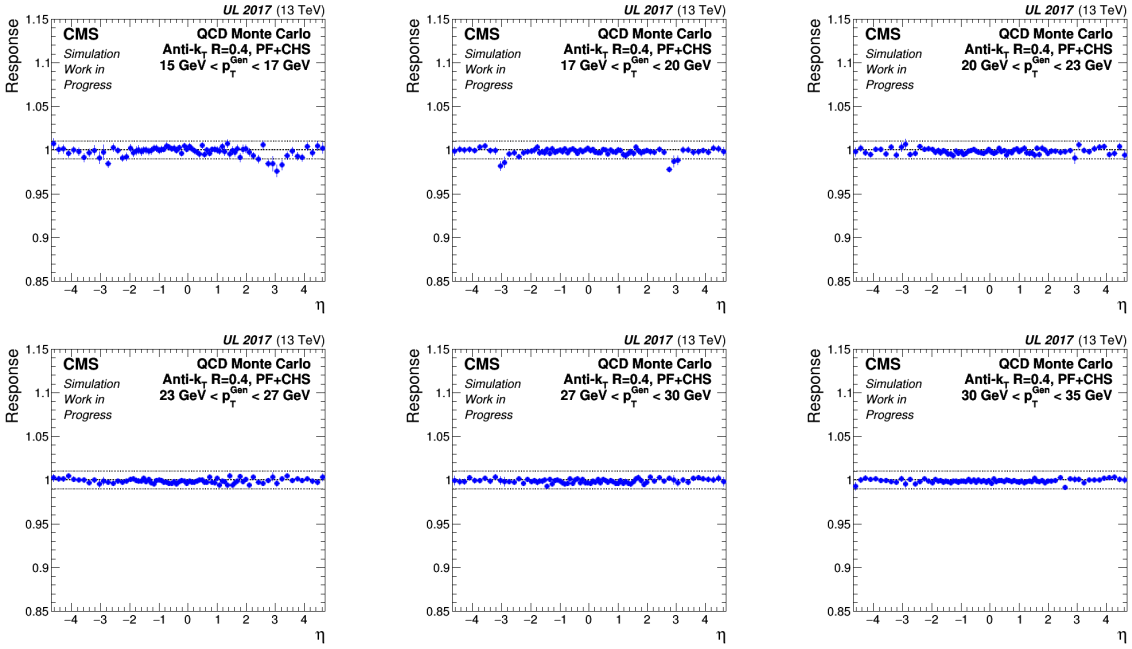

Figure 62: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 CHS jets from the UL 2017 MC.

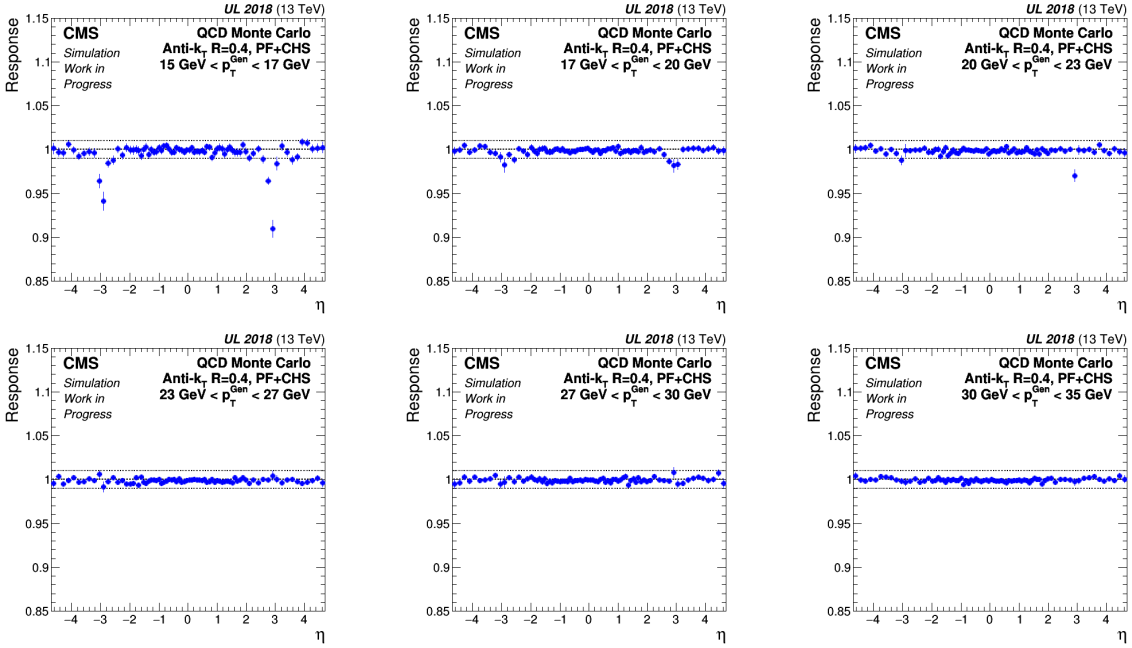

Figure 63: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 CHS jets from the UL 2018 MC.

AK8 CHS jets

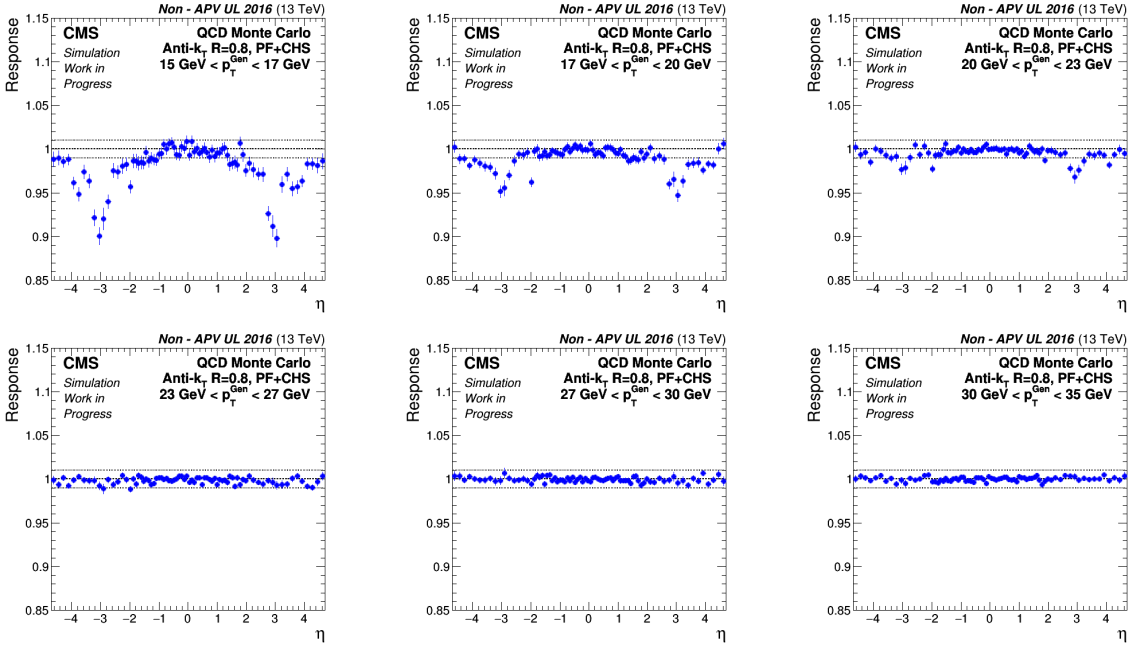

Figure 64: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 CHS jets from the Non - APV UL 2016 MC.

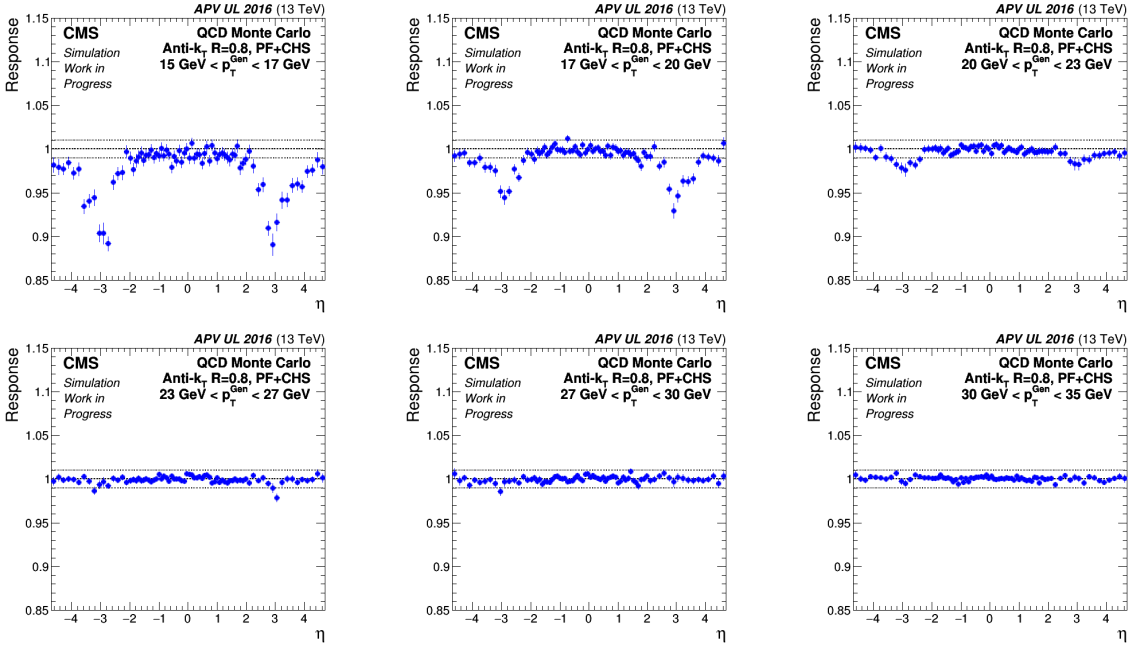

Figure 65: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 CHS jets from the APV UL 2016 MC.

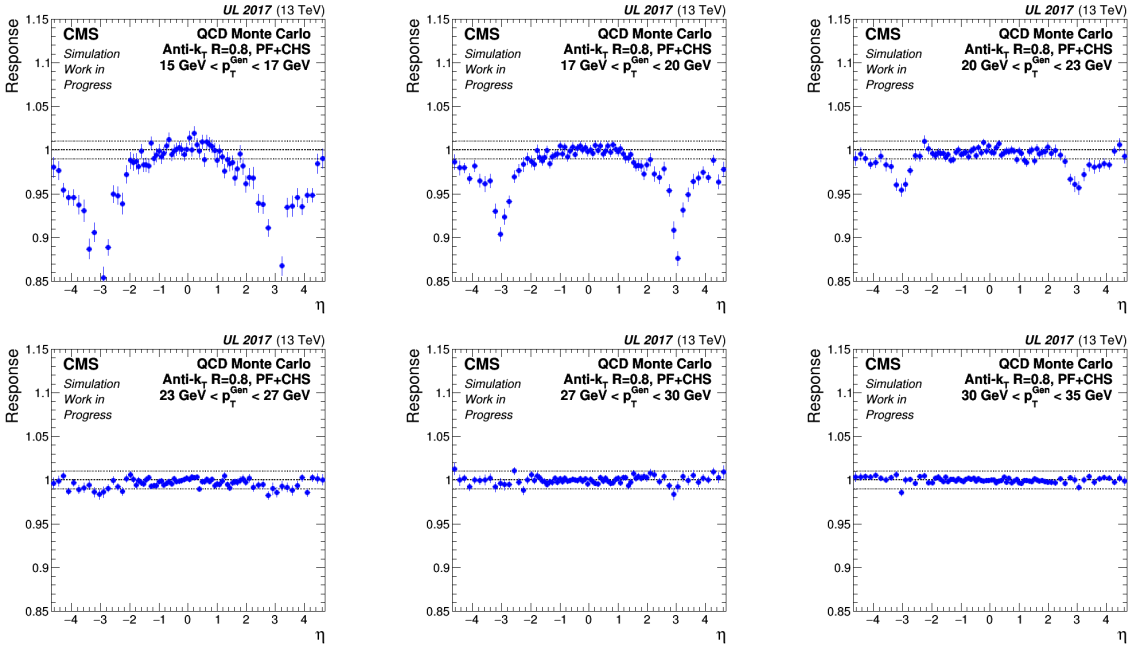

Figure 66: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 CHS jets from the UL 2017 MC.

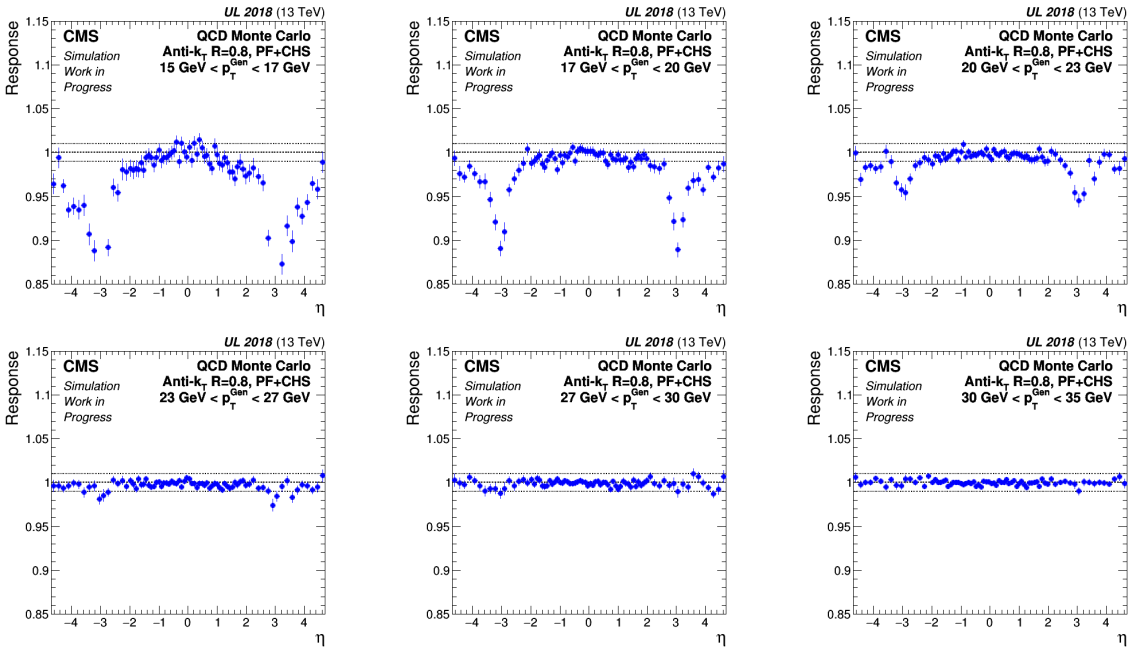

Figure 67: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 CHS jets from the UL 2018 MC.

AK4 PUPPI jets

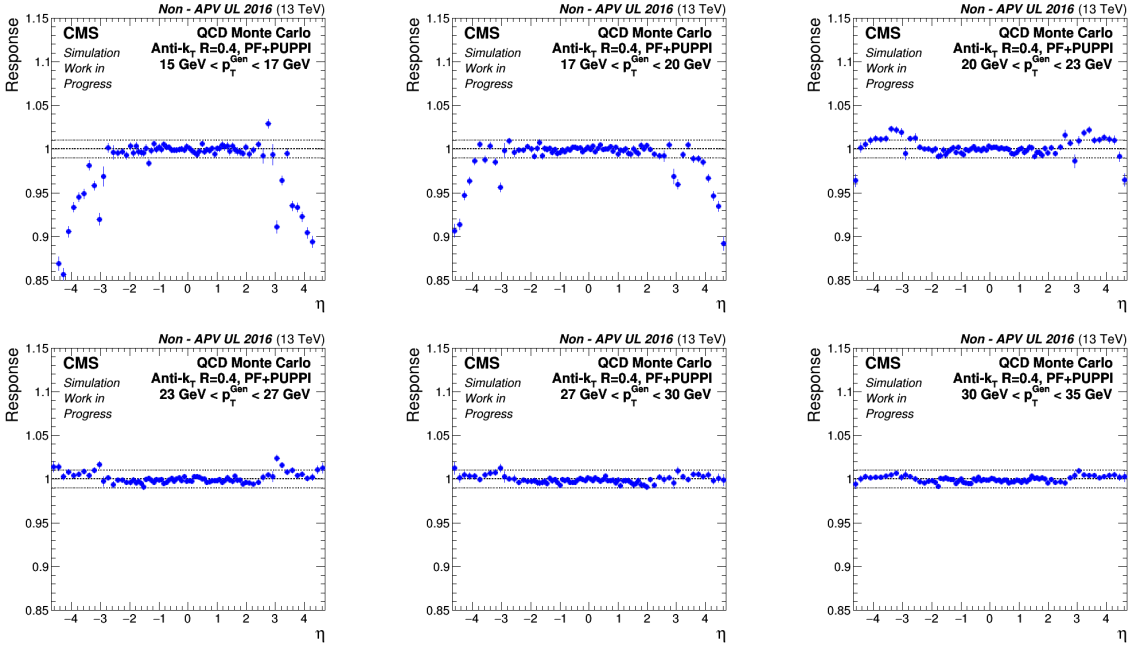

Figure 68: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 PUPPI jets from the Non - APV UL 2016 MC.

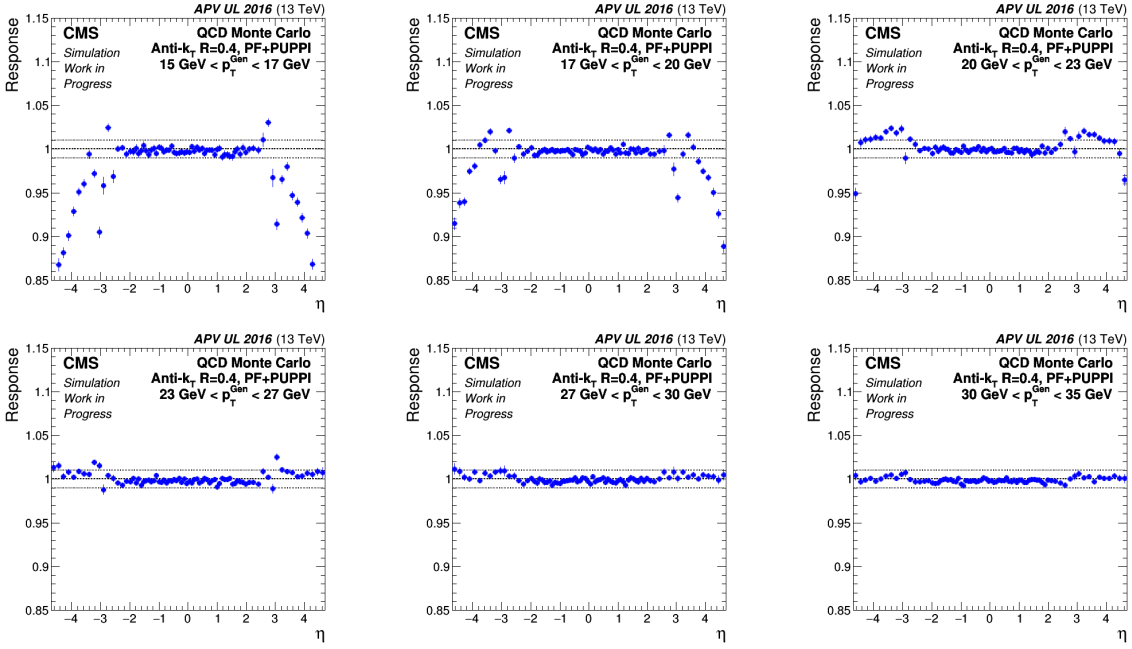

Figure 69: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 PUPPI jets from the APV UL 2016 MC.

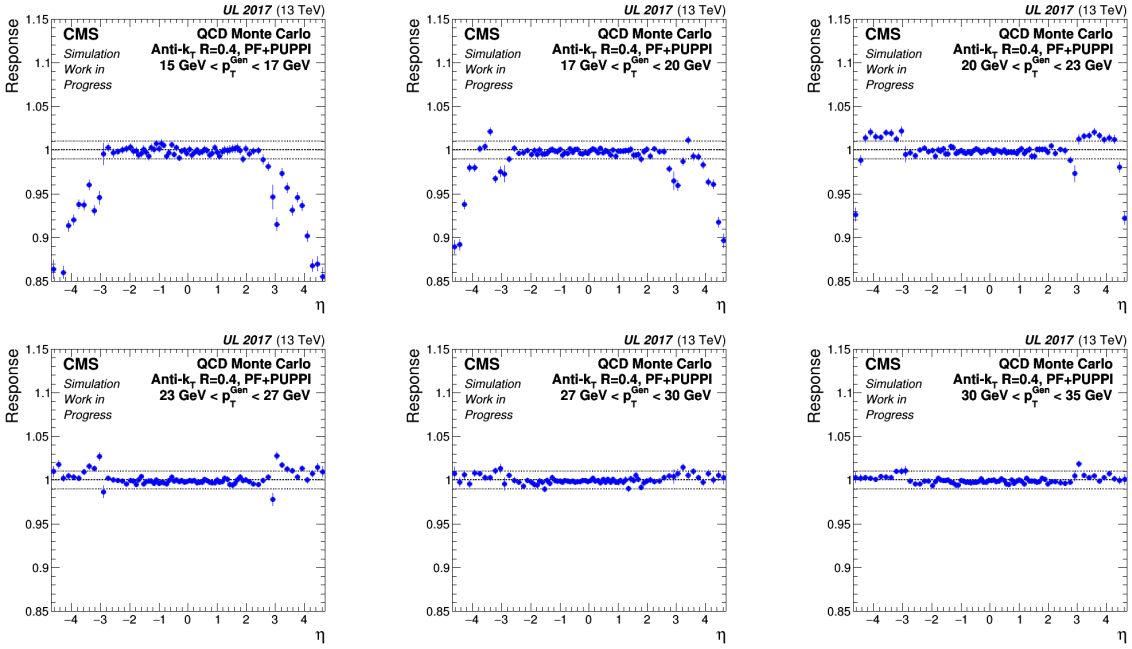

Figure 70: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 PUPPI jets from the UL 2017 MC.

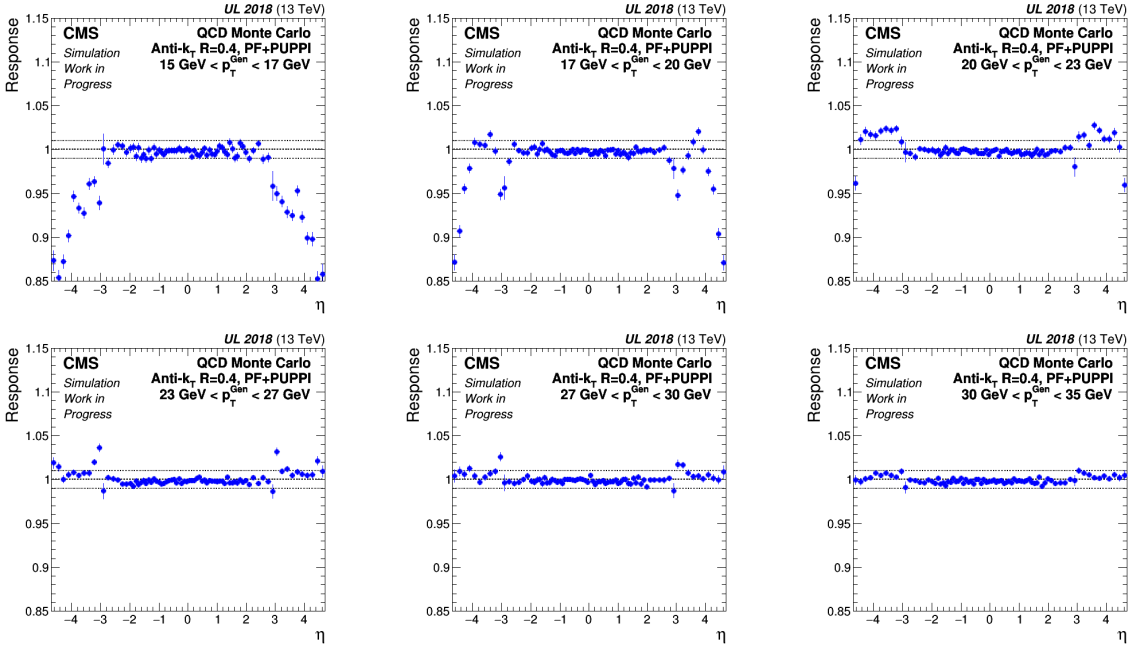

Figure 71: Corrected response as a function of jet η for various p_T^{Gen} bins for AK4 PUPPI jets from the UL 2018 MC.

AK8 PUPPI jets

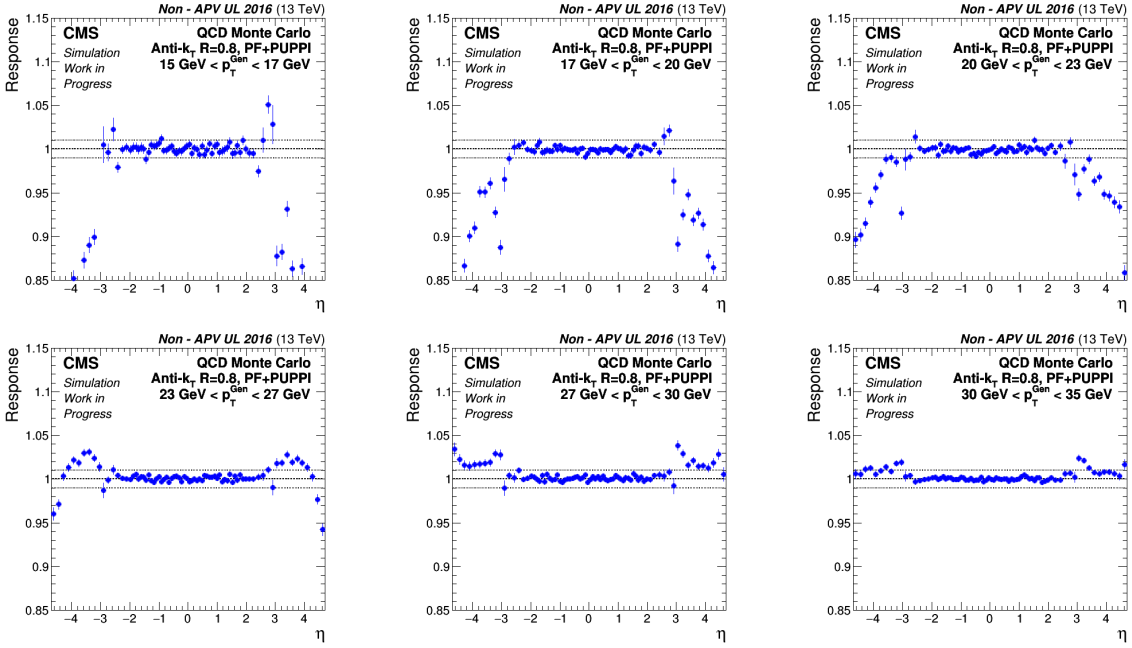

Figure 72: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 PUPPI jets from the Non - APV UL 2016 MC.

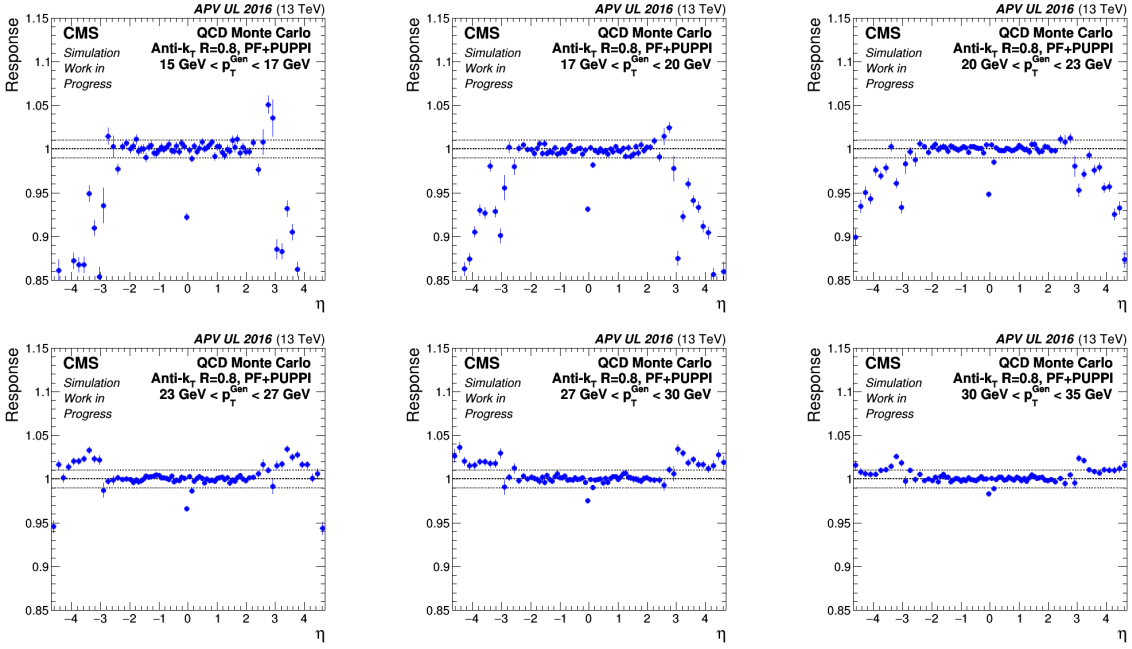

Figure 73: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 PUPPI jets from the APV UL 2016 MC.

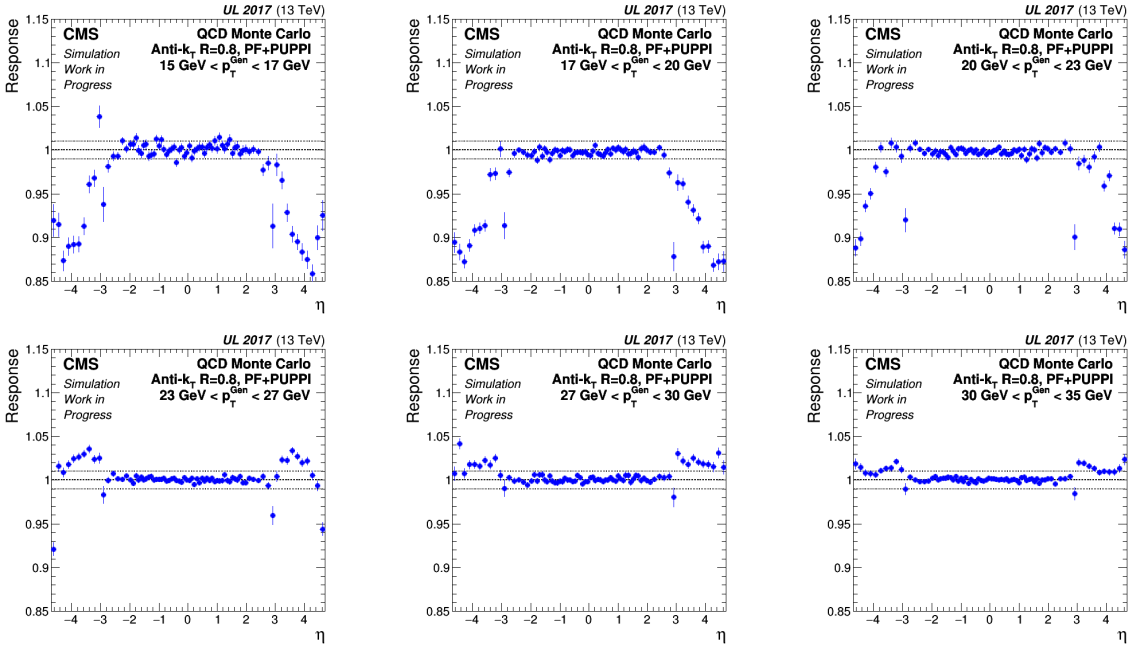

Figure 74: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 PUPPI jets from the UL 2017 MC.

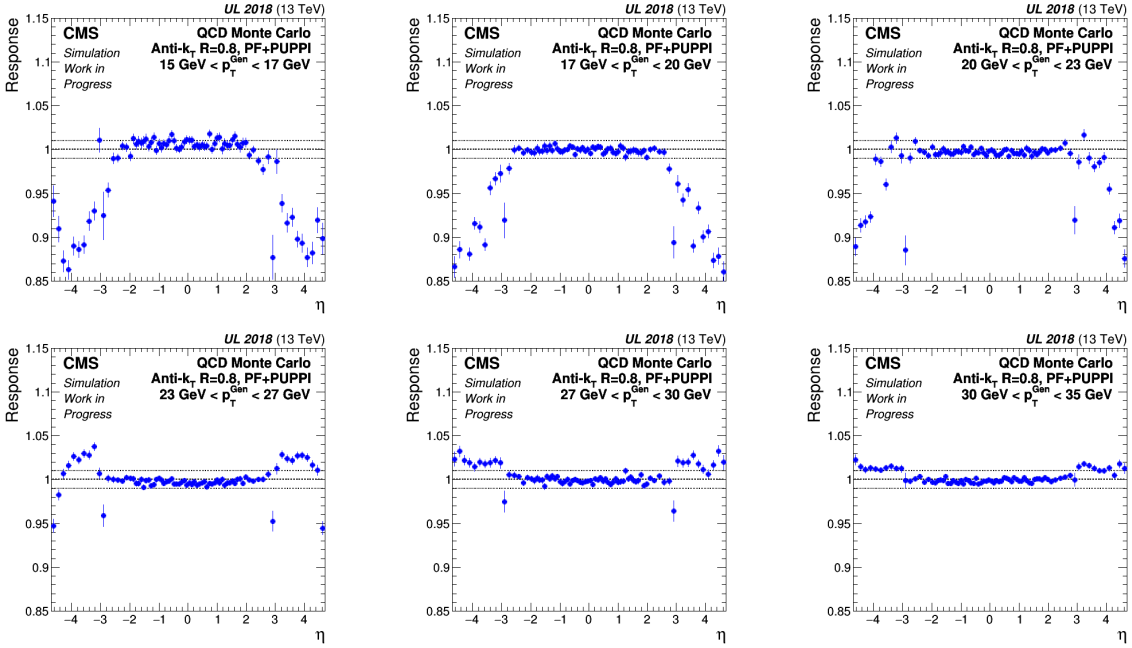

Figure 75: Corrected response as a function of jet η for various p_T^{Gen} bins for AK8 PUPPI jets from the UL 2018 MC.

References

[1] Thomson, Mark, *"Modern particle physics"*, Cambridge University Press, 978-1-107-03426-6, 2013.

[2] Halzen, F. and Martin, Alan D. , *"QUARKS AND LEPTONS: An Introductory Course in Modern Particle Physics"*, 978-0-471-88741-6, 1984.

[3] Harris, Robert M. and Kousouris, Konstantinos, *Searches For Dijet Resonances At Hadron Colliders*, International Journal of Modern Physics A, World Scientific Pub Co Pte Lt, Dec 2011. http://dx.doi.org/10.1142/S0217751X11054905

[4] Lahiri, Amitabha and Pal, Palash B. , *"A First Book of Quantum Field Theory"*, 2005.

[5] Bogdan A. Dobrescu and Robert M. Harris and Joshua Isaacson, *"Ultraheavy resonances at the LHC: beyond the QCD background"*, 2018. https://arxiv.org/abs/1810.09429

[6] CMS Collaboration, *"Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at* $\sqrt{s} = 13$ *TeV"*, Journal of High Energy Physics, Springer Science and Business Media LLC, May 2020. http://dx.doi.org/10.1007/JHEP05(2020)033

[7] CMS Collaboration, *"CMS Physics: Technical Design Report Volume 1: Detector Performance and Software"*, Technical Design Report CMS, CERN-LHCC-2006-001, CERN, Geneva, 2006. https://cds.cern.ch/record/922757

[8] Francesca Cavallari, *"Performance of calorimeters at the LHC"*, Journal of Physics: Conference Series, IOP Publishing, Apr 2011. https://doi.org/10.1088/1742-6596/293/1/012001

[9] Fabjan, Christian Wolfgang and Gianotti, F. , *Calorimetry for Particle Physics*, Rev. Mod. Phys., CERN-EP-2003-075, Oct 2003. https://cds.cern.ch/record/692252

[10] Venkatesh S Kaushik, *"Electromagnetic Showers and Shower Detectors"*, Dept. of High Energy Physics, University of Texas at Arlington, Aug 2002.

[11] CMS Collaboration, *"Particle-flow reconstruction and global event description with the CMS detector"*, Journal of Instrumentation, IOP Publishing, Oct 2017. http://dx.doi.org/10.1088/1748-0221/12/10/P10003

[12] Thomson, M.A., *"Particle flow calorimetry and the PandoraPFA algorithm"*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,

Detectors and Associated Equipment, Elsevier BV, Nov 2009. http://dx.doi.org/10.1016/j.nima.2009.09.009

[13] Cacciari, Matteo and Salam, Gavin P and Soyez, Gregory, *"The anti–k_t jet clustering algorithm"*, Journal of High Energy Physics, Springer Science and Business Media LLC, Apr 2008. http://dx.doi.org/10.1088/1126-6708/2008/04/063

[14] CMS Collaboration, *"Pileup Removal Algorithms"*, CMS-PAS-JME-14-001, CERN, Geneva, 2014. https://cds.cern.ch/record/1751454

[15] Bertolini, D., Harris, P., Low, M. et al. *"Pileup per particle identification."* J. High Energ. Phys. 2014, 59 (2014). https://doi.org/10.1007/JHEP10(2014)059

[16] CMS Collaboration, *"Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV"*, Journal of Instrumentation, IOP Publishing, Feb 2017. https://doi.org/10.1088/1748-0221/12/02/p02014

[17] CMS Collaboration, *"Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS"*, Journal of Instrumentation, IOP Publishing, Nov 2011. https://doi.org/10.1088/1748-0221/6/11/p11002

[18] Cacciari, Matteo and Salam, Gavin P., *"Pileup subtraction using jet areas"*, Physics Letters B, Elsevier BV, Jan 2008. http://dx.doi.org/10.1016/j.physletb.2007.09.077

[19] Ellis, Joshua P., *"TikZ-Feynman: Feynman diagrams with TikZ"*, Computer Physics Communications, Elsevier BV, Jan 2017. http://dx.doi.org/10.1016/j.cpc.2016.08.019