MEASUREMENT OF MULTIJET CROSS-SECTION RATIOS IN PROTON-PROTON COLLISIONS WITH THE CMS DETECTOR AT THE LHC

A THESIS

Submitted to the

FACULTY OF SCIENCE

PANJAB UNIVERSITY, CHANDIGARH

for the degree of

DOCTOR OF PHILOSOPHY

2017

Anterpreet Kaur

DEPARTMENT OF PHYSICS

CENTRE OF ADVANCED STUDY IN PHYSICS

PANJAB UNIVERSITY, CHANDIGARH

INDIA

Dedicated to

my Grand-Parents

&

Parents

Contents

List of Figures

List of Tables

Chapter 1

Measurement of the Differential Inclusive Multijet Cross Sections and their Ratio

The inclusive *n-jet* event samples include the events with number of jets $\geq n$, where $n = 2$ and 3 in the current study. The inclusive multijet event yields are transformed into a differential cross section which is defined as :

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\big(H_{\mathrm{T},2}/2\big)} = \frac{1}{\epsilon\ \mathcal{L}_{\mathrm{int,eff}}} \frac{N_{\mathrm{event}}}{\Delta\big(H_{\mathrm{T},2}/2\big)} \tag{1.1}$$

where N_{event} is the number of inclusive 2- or 3-jet events counted in an $H_{\mathrm{T},2}/2$ bin, ϵ is the product of the trigger and jet selection efficiencies, which are greater than 99%, $\mathcal{L}_{\mathrm{int,eff}}$ is the effective integrated luminosity, and $\Delta\big(H_{\mathrm{T},2}/2\big)$ are the bin widths. The measurements are reported in units of (pb/GeV).

The differential inclusive multijet cross sections are measured as a function of the average transverse momentum, $H_{\mathrm{T},2}/2 = \frac{1}{2}(p_{\mathrm{T},1}+ p_{\mathrm{T},2})$, where $p_{\mathrm{T},1}$ and $p_{\mathrm{T},2}$ denote the transverse momenta of the two leading jets. The cross section ratio R_{32}, defined in Eq. 1.2 is obtained by dividing the differential cross sections of inclusive

3-jet events to that of inclusive 2-jet one, for each bin in $H_{\mathrm{T},2}/2$.

$$R_{32} = \frac{\frac{\mathrm{d}\sigma_{3\text{-}jet}}{\mathrm{d}\left(H_{\mathrm{T},2}/2\right)}}{\frac{\mathrm{d}\sigma_{2\text{-}jet}}{\mathrm{d}\left(H_{\mathrm{T},2}/2\right)}} \tag{1.2}$$

For inclusive 2-jet events ($\mathrm{n_j} \geq 2$) sufficient data are available up to $H_{\mathrm{T},2}/2 = 2\,\mathrm{TeV}$, while for inclusive 3-jet events ($\mathrm{n_j} \geq 3$) and the ratio R_{32}, the accessible range in $H_{\mathrm{T},2}/2$ is limited to $H_{\mathrm{T},2}/2 < 1.68\,\mathrm{TeV}$.

1.1 Data Samples

This measurement uses the data collected at the center of mass energy of 8 TeV by CMS experiment in the 2012 run period of the LHC. The 2012 data is taken in four periods A, B, C, D and the data sets are divided into samples according to the run period. Further each sample is grouped into subsets based on the trigger decision. For run B-D, the `JetMon` stream datasets contain prescaled low trigger threshold paths (HLTPFJet40, 80, 140, 200 and 260) while the `JetHT` stream datasets contain unprescaled high threshold trigger paths (HLT PFJet320 and 400). For run A, the `Jet` stream contains all the above mentioned trigger paths. The data to be used in physics analysis must satisfy a certain criteria which include proper performance of all detector subsystems as well as the passing of data quality monitoring (DQM) steps during the validation process. CMS uses JSON (Java Script Object Notation) format files to store the range of good lumi sections within a run. In the current analysis, the applied certification file[1] is based on the final event reconstruction of the 2012 CMS data sets. The datasets used in the current study are mentioned in the Table 1.1 along with the luminosity of each dataset.

[1] Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON

Table 1.1: Four data sets collected in run periods A, B,C and D during 2012, along with the corresponding run numbers and luminosity.

Run	**Run range**	**Data set**	**Luminosity** fb^{-1}
A	190456-193621	/Jet/Run2012A-22Jan2013-v1/AOD	0.88
B	193834-196531	/Jet[Mon,HT]/Run2012B-22Jan2013-v1/AOD	4.41
C	198022-203742	/Jet[Mon,HT]/Run2012C-22Jan2013-v1/AOD	7.06
D	203777-208686	/Jet[Mon,HT]/Run2012D-22Jan2013-v1/AOD	7.37

The data sets have the LHC luminosity increasing with period, full data sample of 2012 corresponds to an integrated luminosity of 19.71 fb^{-1}.

1.1.1 Monte Carlo Samples

To have a comparison of data results with the simulated events, the MadGraph5 [2] Monte-Carlo event generator has been used. The MadGraph5 generates matrix elements for High Energy Physics processes, such as decays and $2 \rightarrow n$ scatterings. The underlying event is modeled using the tune Z2*. It has been interfaced to pythia6 [3] by the LHE event record [4], which generates the rest of the higher-order effects using the Parton Showering (PS) model. Matching algorithms ensure that no double-counting occurs between the tree-level and the PS-model-generated partons. The MC samples are processed through the complete CMS detector simulation to allow studies of the detector response and compare to measured data on detector level.

The cross section measured as a function of the transverse momentum p_{T} or the scalar sum of the transverse momentum of all jets H_{T} falls steeply with the increasing p_{T}. So in the reasonable time, it is not possible to generate a large number of high p_{T} events. Hence, the events are generated in the different phase-space region binned in H_{T} or the leading jet p_{T}. Later on, the different phase-space regions are added together in the data analyses by taking into account the cross section of the different phase-space regions. The official CMS MadGraph5 + pythia6 (MG+P6)

MC samples used in this analysis are generated as slices in the H_{T} phase-space are tabulated in Table 1.2 along with their cross sections and number of events generated.

Table 1.2: The official MC production samples generated in phase space slices in H_{T} with the generator MadGraph5 and interfaced to pythia6 for the parton shower and hadronization of the events. The cross section and number of events generated are mentioned for each sample.

Generator	Sample	Events	Cross Section pb
MadGraph5 + pythia 6	/QCD_HT-100To250_TuneZ2star_8TeV-madgraph-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	50129518	1.036×10^{7}
	/QCD_HT-250To500_TuneZ2star_8TeV-madgraph-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27062078	2.760×10^{5}
	/QCD_HT-500To1000_TuneZ2star_8TeV-madgraph-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30599292	8.426×10^{3}
	/QCD_HT-1000ToInf_TuneZ2star_8TeV-madgraph-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	13843863	2.040×10^{2}

1.2 Event Selection

To yield a multijet sample with high purity and high selection efficiency, the events are selected according to several quality criteria. This event selection also reduces beam induced background, detector-level noise and jets arising from fake calorimeter energy deposits.

1.2.1 Trigger Selection

CMS implements a trigger system organized in two levels, in order to reduce the amount of recorded events to a sustainable rate. This analysis deals with jets in the final state, so single jet trigger paths are used to select events in data which consists of one L1 trigger seed and multiple HLT filters. The L1 jet trigger uses transverse energy sums computed using both HCAL and ECAL in the central region ($|\eta| < 3.0$) or HF in the forward region ($|\eta| > 3.0$). A more elaborate but still very fast algorithm, the "jet finder", is then implemented on the energy cluster but with a finer segmentation in order to select the raw object for the HLT trigger : the algorithm makes use of a cone size in order to cluster in a primitive jet the calorimeter towers whose energy is larger than the seed threshold. If the primitive HLT jet has an energy above the threshold set by the trigger, the event is selected and the collection of recorded data is saved and stored in streams. The single jet triggers used for this analysis are tabulated in Table 1.3. HLT_PFJetX implies that there is at-least one jet in the event, whose $p_T > X$ (GeV). The L1 trigger has a lower threshold to ensure full efficiency versus p_T of the HLT trigger. The p_T spectrum is steeply falling and hence the rates for low-p_T jets are very high. So it is not feasible to use a single unprescaled trigger for the selection of all required events. To collect sufficient data in the lower part of the p_T spectrum, different five prescaled low-p_T trigger paths, each with different prescale value, are used. Also, one unprescaled trigger i.e. HLT_Jet320 is used in the high p_T region, in which the rate is sufficiently small to collect and store all events. During the reconstruction of the spectrum, the prescales have been taken into the account.

Table 1.3: List of all single jet trigger paths used in the analysis. The column $H_{T,2}/2$, 99% indicates the value of $H_{T,2}/2$ at which each trigger exhibits an efficiency larger than 99%. The last column reports the effective luminosity seen by each trigger. This number, divided by the total integrated luminosity of 19.71 fb^{-1}, gives the effective prescale applied on a trigger over the whole run period.

Trigger Path	**L1 threshold** GeV	**HLT threshold** GeV	$\boldsymbol{H_{T,2}/2}$**, 99%** GeV	**Eff. Lumi** fb^{-1}
HLT_PFJet80	36	80	120.0	0.21×10^{-2}
HLT_PFJet140	68	140	187.5	0.56×10^{-1}
HLT_PFJet200	92	200	262.5	0.26
HLT_PFJet260	128	260	345.0	1.06
HLT_PFJet320	128	320	405.0	19.71

The efficiency of each trigger, as a function of the measured observable, is described by the turn-on curves with a rising part, where the trigger is partly inefficient, until a plateaux region, corresponding to the region of full efficiency of the trigger. Hence it is necessary to determine the threshold above which a trigger becomes fully efficient. It is defined as the value at which the efficiency exceeds

99%. In the assumption that the reference trigger HLT_PFJetX is fully efficient in the considered region of the phase space, the trigger efficiency for HLT_PFJetY is defined as Eq. 1.3. The value of X is chosen previous to that of Y in p_{T} ordering from the trigger list so that the higher trigger condition can be emulated from the lower trigger path.

$$\epsilon_{\mathrm{HLT\_PFJetY}} = \frac{H_{\mathrm{T},2}/2\Big(\mathrm{HLT\_PFJetX} + (\mathrm{L1Object\_p_T} > \mathrm{Z}) + (\mathrm{HLTObject\_p_T} > \mathrm{Y})\Big)}{H_{\mathrm{T},2}/2(\mathrm{HLT\_PFJetX})} \tag{1.3}$$

where Y indicates the p_{T} threshold of HLT_PFJetY and Z is the L1 seed value corresponding to the trigger path HLT_PFJetY. The denominator represents the number of events for which the reference trigger path HLT_PFJetX has been fired. The numerator is the number of events for which HLT_PFJetX has been fired along the p_{T} of L1Object $\geq$ Z and the p_{T} of HLTObject $\geq$ Y. For example, in order to obtain turn-on curve for HLT_PFJet260, the reference HLT path HLT_PFJet200 is chosen, the p_{T} cut on L1Object is 128 GeV and p_{T} cut on HLTObject is 260 GeV. The uncertainty on the efficiency is indicated by error bars which represent Clopper-Pearson confidence intervals [5]. To determine the point, at which the trigger efficiency is larger than 99%, the turn-on distribution is fitted using a sigmoid function described in Eq. 1.4. The trigger turn-on curves as a function of $H_{\mathrm{T},2}/2$ can been seen in Fig. 1.1.

$$f_{fit}(x) = \frac{1}{2}\left(1 + erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right) \tag{1.4}$$

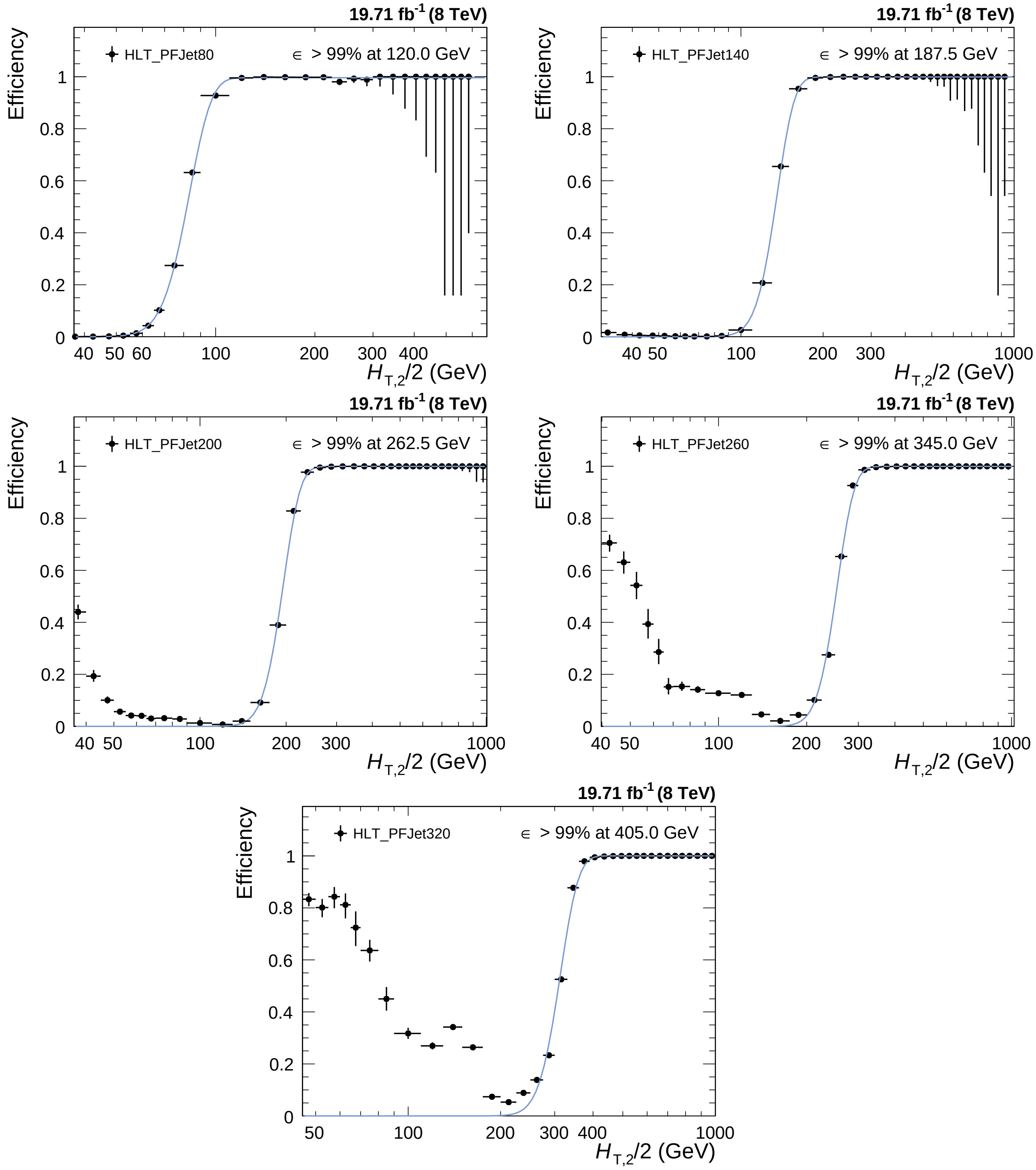


Figure 1.1: Trigger efficiencies turn-on curves for the single jet trigger paths used in the analysis.To determine the 99% efficiency threshold, the trigger turn-on curves are fitted using a sigmoid function taking into account the uncertainties using Clopper-Pearson confidence intervals.

1.2.2 Primary Vertex Selection

A primary vertex (PV) is identified by a collection of tracks, measured in the tracker with a good fit quality between the hits and compatible with the beam line. The tracks are clustered according to the z-coordinate of their point of closest approach to the beam axis. Each event is required to have at least one good PV which is well reconstructed within a distance of $\left|z(PV)\right| < 24$ cm to the nominal interaction point of the detector. Also the radial distance in x-y plane, ρ(PV) should be smaller than 2 cm. The number of degrees of freedom in vertex fit needs to be at-least four. Thus, at least four tracks must be present in order to perform a valid vertex fit.

1.2.3 Missing Transverse Energy

If all particles could be identified and perfectly measured, the transverse momentum of all particles would sum up to zero. Neutral weakly interacting particles, such as neutrinos, escape from typical collider detectors without producing any direct response in the detector elements. The presence of such particles must be inferred from the imbalance of total momentum of all visible particles. The vector momentum imbalance in the plane perpendicular to the beam direction is known as missing transverse momentum or energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). It is one of the most important observables for discriminating leptonic decays of W bosons and top quarks from background events which do not contain neutrinos, such as multijet and Drell–Yan events or searches for physics beyond the Standard Model which involve undetectable particles.

The ratio of missing transverse energy to the total transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}/\sum E_{\mathrm{T}}$, shown in Fig. 1.2 for $\mathrm{n_j} \geq 2$ (left) and $\mathrm{n_j} \geq 3$ (right), shows a discrepancy between data and MC at the tail part of the distribution. This is because of a finite contribution from Z($\rightarrow \nu\bar{\nu}$) + jet events which gives rise to non-zero E_{T} in the events in data. Such events are absent in QCD simulated events in MC. Hence

$E_{\mathrm{T}}^{\mathrm{miss}}/\sum E_{\mathrm{T}}$ is required to be less than 0.3 to reject events with high $E_{\mathrm{T}}^{\mathrm{miss}}$.

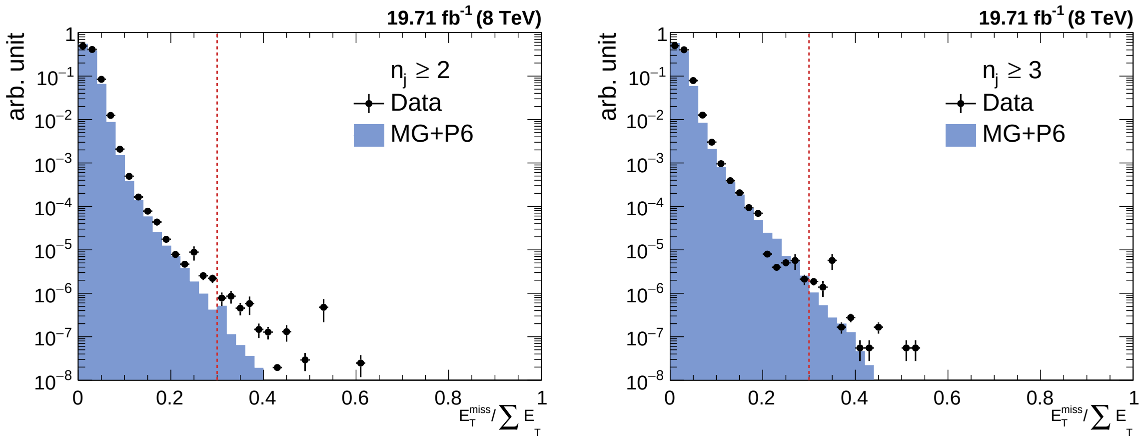


Figure 1.2: Missing transverse energy fraction of the total transverse energy per event in data and simulated events in inclusive 2-jet (left) and 3-jet events (right). To remove background and noise, events with a fraction exceeding a certain threshold, here indicated with the red dashed line, are rejected.

1.2.4 Jet Identification

In order to suppress fake jets, arising from detector noise or misreconstructed particles, jet identification criteria (ID) has been applied. Instead of applying it eventwise, it is applied it on each jet. The algorithm works on reconstructed jets using information of the clustered particle candidates. The official tight jet ID [6], recommended by JETMET group [1] is used. Due to pileup and electronic noise the jet constituent fractions may vary from event to event. In order to reject the noisy jets, some jet selection criteria are optimized to select only good quality jets. The selection criteria are implemented as selection cut on jet fractions. Table 1.4 summarizes the properties of the reconstructed jets and their respective cuts. Each jet should contain at least two particles, one of which should be a charged hadron. The cut on the fraction of neutral hadrons and photons removes HCAL noise and ECAL noise, respectively. Muons that are falsely identified and clustered as jets are removed by the muon fraction criterion. Based on information of the tracker, additional selection cuts are enforced in the region $|\eta| < 2.4$. The charged electromagnetic fraction cut removes the jets clustered from misidentified electrons. Furthermore, the fraction of charged hadrons in the jet must be larger than zero and jets without any charged hadrons are very likely to be pileup jets. The Figures 1.3 and 1.4 show the distributions of the jet constituents observed in data and simulated events for $n_j \geq 2$ and $n_j \geq 3$, respectively.

Table 1.4: The jet ID removes noise and fake jets based on the properties of the reconstructed jets and the clustered particle candidates. All the selection cuts which are recommended by the JETMET group are applied [1].

	Property	**Loose ID**	**Tight ID**
	neutral hadron fraction	< 0.99	< 0.90
Whole	neutral EM fraction	< 0.99	< 0.90
η region	number of constituents	> 1	> 1
	muon fraction	< 0.80	< 0.80
	charged hadron fraction	> 0	> 0
only $\lvert\eta\rvert < 2.4$	charged multiplicity	> 0	> 0
	charged EM fraction	< 0.99	< 0.90

1.2.4.1 Jet ID Efficiency

The efficiency of the jet ID as a function of $H_{\mathrm{T},2}/2$ is studied using a tag-and-probe technique with dijet events. The two leading jets are required to be back-to-back in the azimuthal plane such that $|\Delta\phi - \pi| < 0.3$. One of the dijets is selected randomly as a "tag" jet which is required to fulfill the tight jet ID criteria. The other jet is called "probe" jet for which it is examined, whether it also passes the tight jet ID. The ID efficiency is defined as the ratio of events where the probe jet passes the ID requirements, over the total number of dijet events. Figure 1.5 shows the ID efficiency as a function of $H_{\mathrm{T},2}/2$ for $n_j \geq 2$ (left) and $n_j \geq 3$ (right) ?. As expected, the jet ID efficiency is larger than 99%. The QCD cross section decreases as a function of $H_{\mathrm{T},2}/2$ and hence the number of events decrease on moving to higher $H_{\mathrm{T},2}/2$. Consequently the statistical fluctuations for ID efficiency are larger at higher $H_{\mathrm{T},2}/2$.

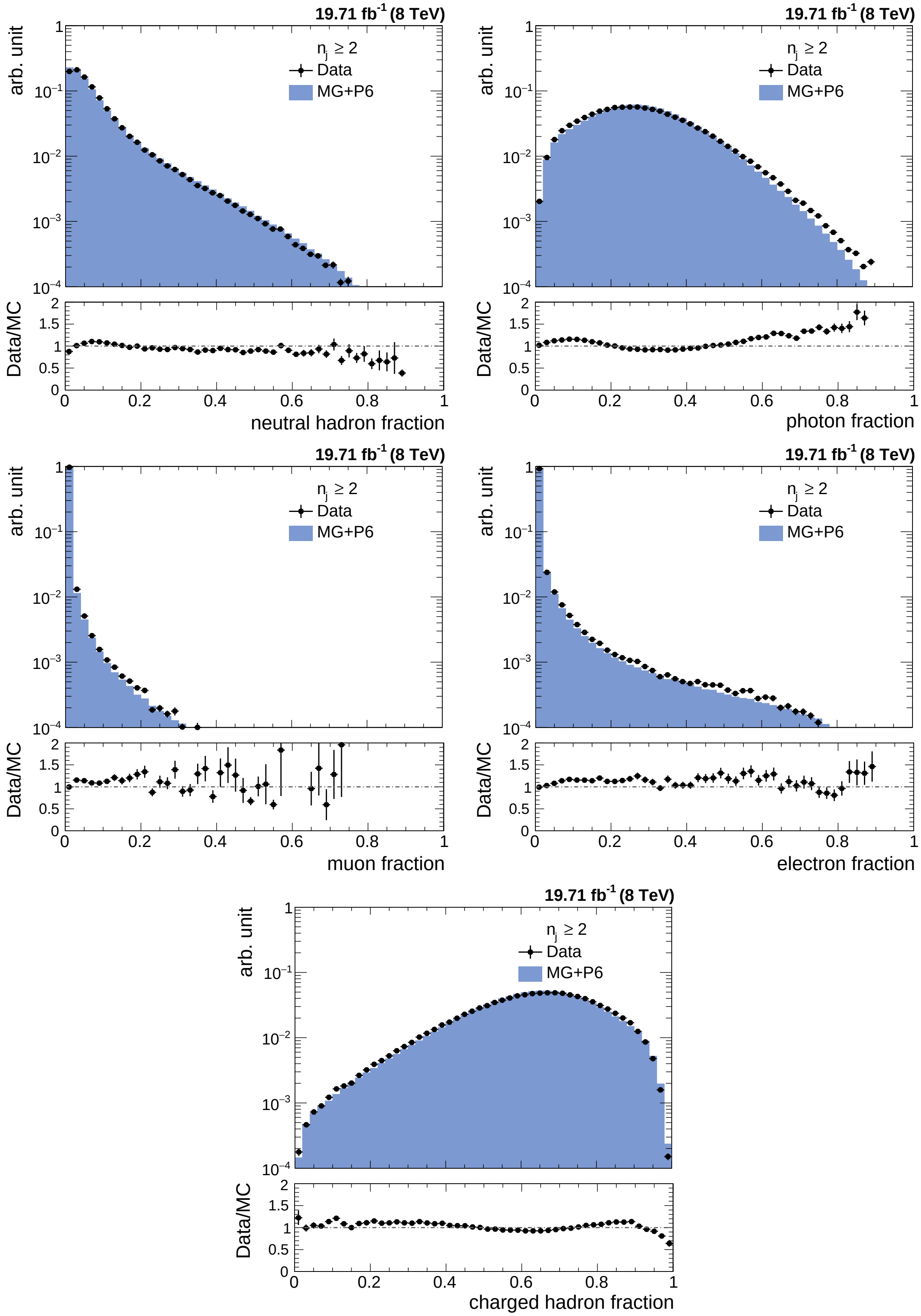


Figure 1.3: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 2-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.

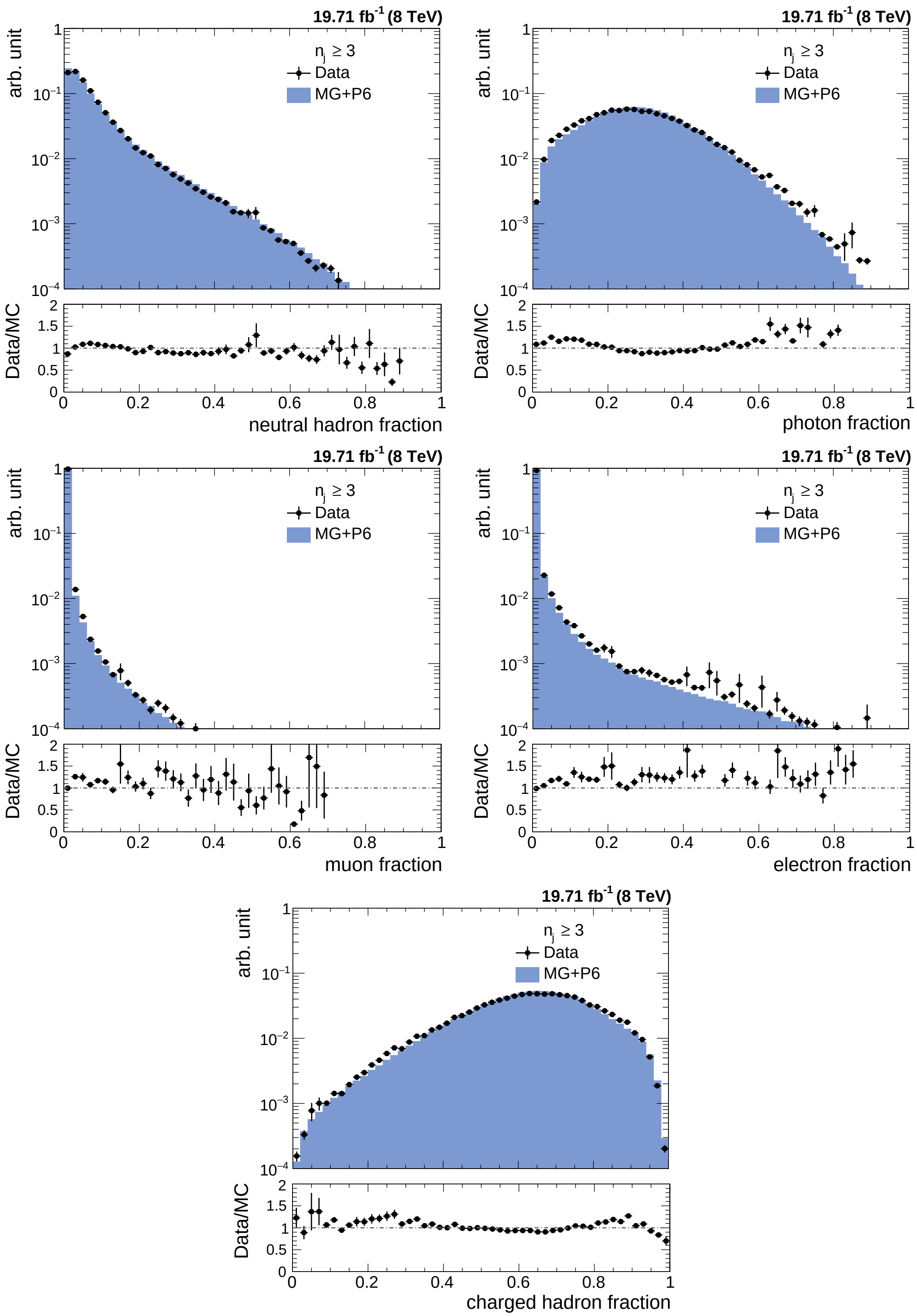


Figure 1.4: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates for inclusive 3-jet events. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.

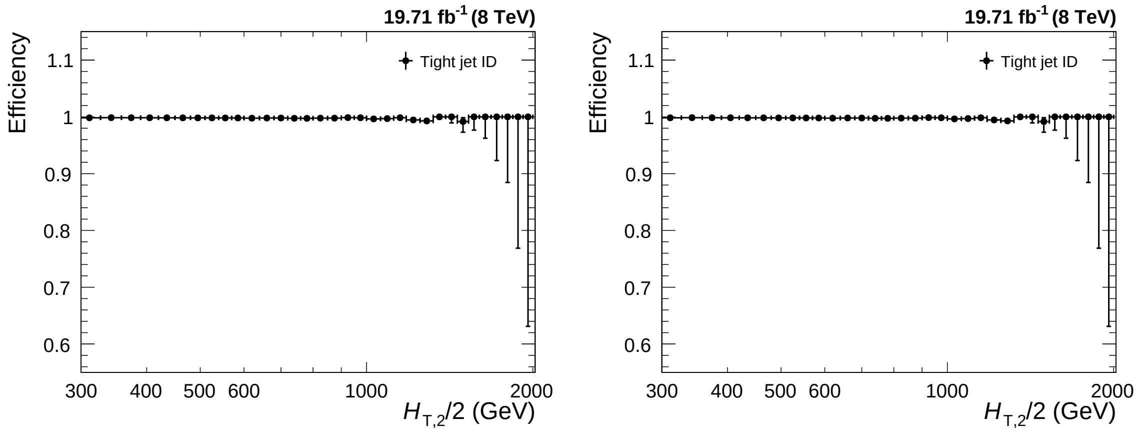


Figure 1.5: The jet ID efficiency studied using a tag-and-probe technique on dijet event topologies, is shown as a function of $H_{T,2}/2$ for inclusive 2-jet (left) and 3-jet events (right) and it always exceeds 99%.

1.2.5 Jet Energy Corrections and Selection

The measurement presented in this thesis is based on jets clustered from PF candidates using the anti-k$_t$ jet algorithm with a size parameter of 0.7. All the jet energy corrections, described in section ? and recommended by CMS, are applied prior to this selection in order to have the correct energy scale of the jets. These comprises different correction levels for jets in data[2] and for jets in simulated events[3]. The jet selection, based on phase space cuts on transverse momentum and rapidity of jets in an event, is as follows :

- All jets having $p_T > 150$ GeV and $|y| < 5.0$ are selected.
- Events with at least two jets are selected.
- The two leading jets should have $|y| < 2.5$ and further jets are counted only, if they lie within the same central rapidity range of $|y| < 2.5$.

These cuts assure high detector acceptance and exactly same selection is applied in the measurement, simulated events as well in theoretical calculations for a consistent

[2]The JEC version applied on data is internally referred to as Winter14_V8
[3]The latest JEC for run-independent Monte Carlo Samples are called START53_V27

comparison.

1.3 Comparison with Simulated Events

1.3.1 Pile-up Reweighting

The official Monte-Carlo samples are generated with distributions for the number of pileup interactions which is meant to roughly cover the conditions expected for each data-taking period. But the number of pile-up events implemented in the simulation $N_{\mathrm{MC}}(N_{\mathrm{PU,truth}})$, is not exactly the same as the one measured in data $N_{\mathrm{data}}(N_{\mathrm{PU,est.}})$. To match the pile-up distributions in data and simulated events, the simulated events are reweighted with a weight w_{PU}, given by :

$$w_{\mathrm{PU}} = \frac{N_{\mathrm{data}}(N_{\mathrm{PU,est.}})/\sum N_{\mathrm{data}}}{N_{\mathrm{MC}}(N_{\mathrm{PU,truth})}/\sum N_{\mathrm{MC}}} \tag{1.5}$$

Figure 1.6 shows the number of reconstructed vertices before and after reweighting. The significant mismatch of the pile-up distributions in data and simulated events, which is present before the reweighting, completely vanishes.

1.3.2 Cross Section Comparison

The measured data distribution of differential cross section at detector level is compared to the predictions of Monte Carlo simulation using MadGraph5 generator interfaced with pythia6 (MG+P6) and including the detector simulation as well as to a fixed-order prediction of NLOJet++. Figure 1.7 shows the comparison of differential cross section as a function of $H_{\mathrm{T,2}}/2$ for $\mathrm{n_j} \geq 2$ (left) and $\mathrm{n_j} \geq 3$ (right), for data (black solid circles), MG+P6 MC (red hollow circles) and NLO (histogram). The bottom panel in each plot shows the ratio of data to the MC predictions (red line) as well as to the NLO predictions (blue line). The NLO predictions on par-

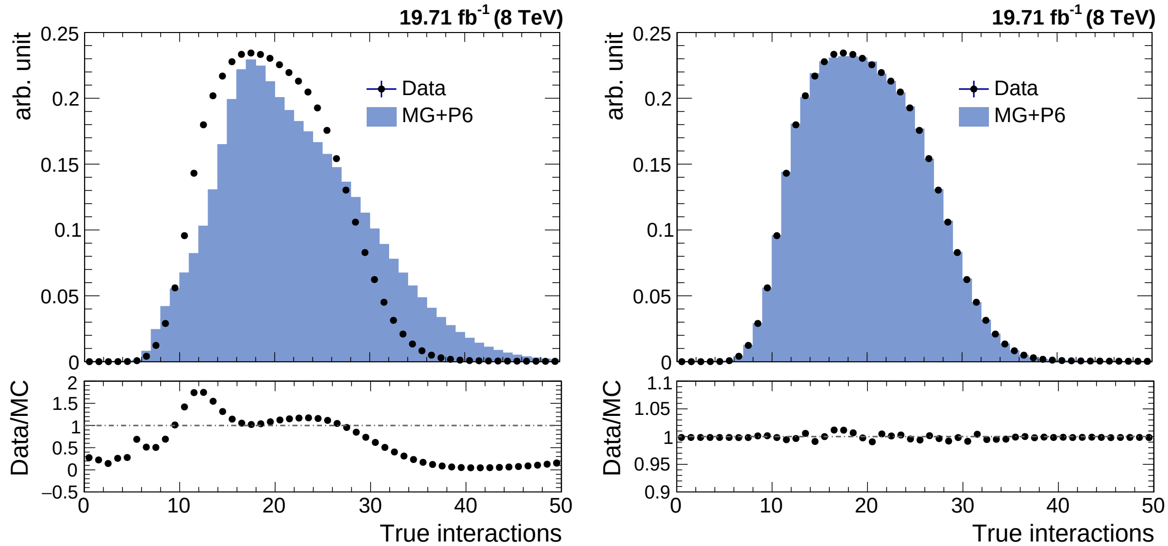


Figure 1.6: Number of reconstructed vertices in data and simulated events before (left) and after (right) the pile-up reweighting.

ton level are not corrected for non-perturbative effects. Still the NLO predictions describe the data better as compared to the LO MC simulations.

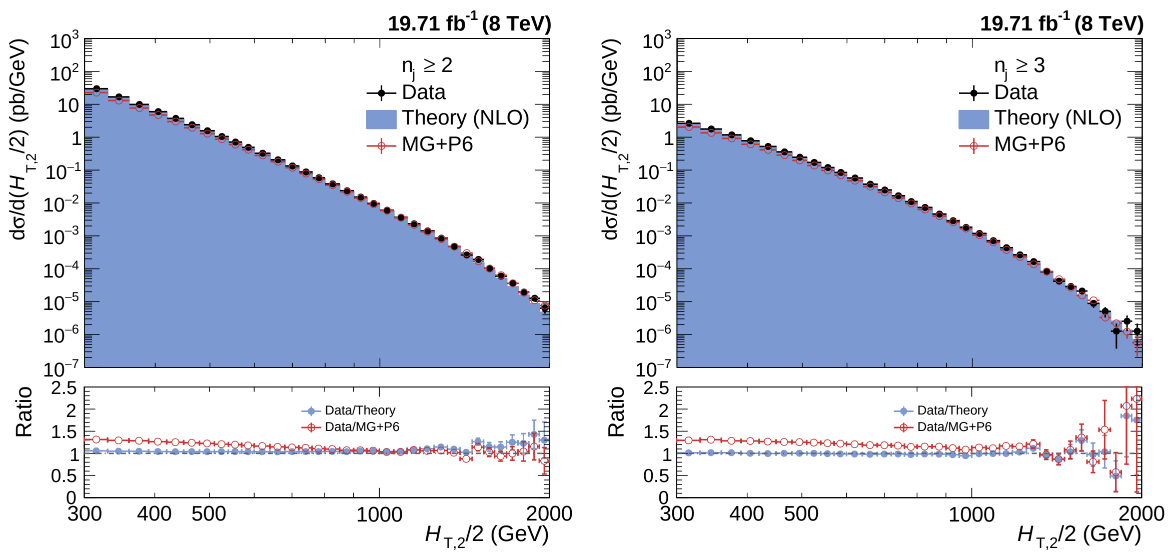


Figure 1.7: The differential cross sections are compared for data (black solid circles) and LO MadGraph5 + pythia6 (MG+P6) MC (red hollow circles), at reconstructed level with NLO theory predictions (histogram), as a function $H_{\mathrm{T},2}/2$ for inclusive 2-jet events (left)and 3-jet events (right). Ratios of data to the MC predictions (red line) as well as to the NLO predictions (blue line) are shown in bottom panel of each plot.

1.4 Jet Energy Resolution (JER)

In an ideal experiment, the value of a physical quantity would be determined exactly with an infinite precision. For e.g. whenever a particle with energy E passes an ideal calorimeter having infinite resolution, the measured energy should always be equal to E. But in real detector, the measured energy of the above mentioned particle might differ from the value E. This shift of the measured quantity from its true value may be due to detector noise, uncertainties in the calibration, non-linearity of the response etc. Hence this results in the finite value of the detector resolution (JER). In such case, the measured values of energy of different particles, crossing the same detector with same energy E, will be different. The set of measurements of this type would in the form of a gaussian distribution, centered around the true value of the measured quantity, whose width is generally interpreted as detector resolution. The resolution of a detector indicates that how precise it is able to measure a given physical observable. The narrower the distribution, the higher the resolution is and hence the more precise is the dectector. The importance of the measurement of the detector resolution lies in the fact that it indicates how much the measured value of the observable differs from the true one.

Due to finite resolution of the CMS detector, the measured transverse momentum of jets gets smeared. Since the observable in this study i.e. $H_{\mathrm{T},2}/2$ is the average sum of transverse momentum of leading and sub-leading jets, the resolution of the detector has to be studied in terms of the observable. CMS detector simulation based on MG+P6 MC event generators is used to determine the resolution as both the particle and reconstructed level information is available. The jets clustered from stable generator particles called Gen jets as well as from particle flow candidates reconstructed from the simulated detector output called Reco jets, are used. The studies of the JETMET working group at CMS has shown that the jet energy resolution in data is actually worse than in simulation [7]. So the reconstructed jet transverse momentum needs to be smeared additionally to match the resolution in data. Table 1.5 shows the scaling factors (c) which need to be applied on the transverse momentum of simulated reconstructed jets. The scaling factors depend on the absolute η of the jet. The uncertainty on these measured scaling factors needs to be taken into account in a physics analysis. This is done by smearing the reconstructed jets with two additional sets of scaling factors, c_{up} and c_{down}, that correspond to varying the factors up and down respectively, by one sigma and evaluating the impact of these new sets.

Table 1.5: JETMET working group at CMS has shown that the jet energy resolution in data is actually worse than in simulation [7]. The scaling factors need to be applied to the reconstructed jet transverse momentum in simulated events to match the resolution in data. The uncertainty on the resolution is given by an upwards and downwards variation c_{up} and c_{down} of the smearing factor $c_{central}$.

η	0.0 – 0.5	0.5 – 1.1	1.1 – 1.7	1.7 – 2.3	2.3 – 2.8
$c_{central}$	1.079	1.099	1.121	1.208	1.254
c_{down}	1.053	1.071	1.092	1.162	1.192
c_{up}	1.105	1.127	1.150	1.254	1.316

The reconstructed jet p_{T} is smeared randomly using a gaussian width widened by the scaling factor ($c_{central}$)

$$p_{\mathrm{T}} \rightarrow Gauss\left(\mu = p_{\mathrm{T}}, \sigma = \sqrt{c^2_{central} - 1} \cdot \mathrm{JER}(p_{\mathrm{T}})\right) \tag{1.6}$$

where $\mathrm{JER}(p_{\mathrm{T}})$ is the resolution determined as a function of jet p_{T} using MG+P6 MC simulated events. After smearing transverse momentum of each reco jet, $H_{\mathrm{T},2}/2$ is calculated from both generator particle jets (Gen $H_{\mathrm{T},2}/2$) as well as the particle flow or reconstructed jets (Reco $H_{\mathrm{T},2}/2$). Then the response is calculated as defined in the Eq. 1.7.

$$R = \frac{\text{Reco } H_{\mathrm{T},2}/2}{\text{Gen } H_{\mathrm{T},2}/2} \tag{1.7}$$

The width of the response distribution in a given Gen $H_{\mathrm{T},2}/2$ bin is interpreted

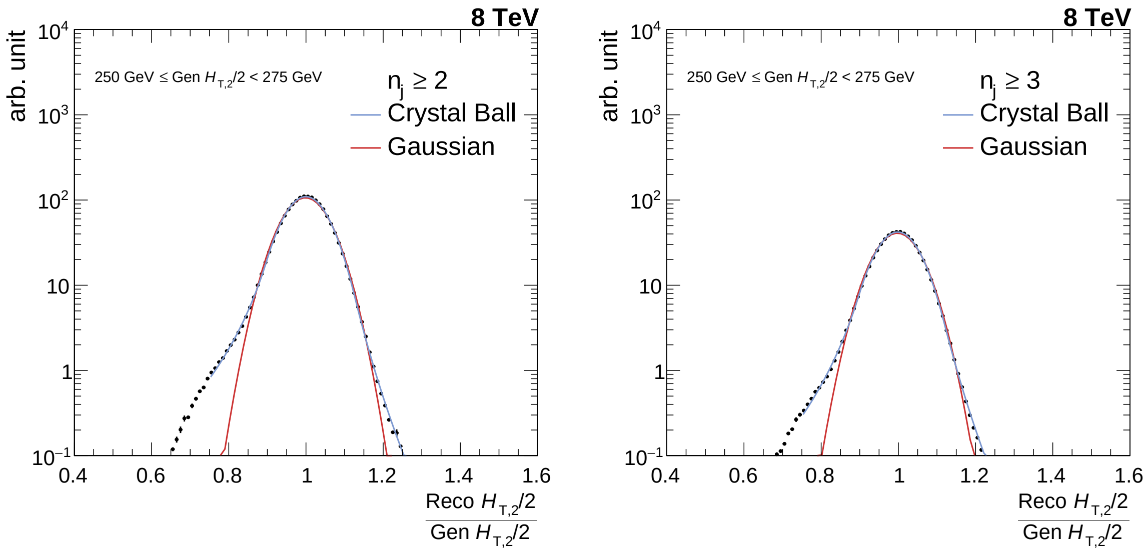


Figure 1.8: Fitting of the resolution distribution as a function of $H_{\mathrm{T},2}/2$ for inclusive 2-jet (left) and for inclusive 3-jet events (right). The blue line shows the double-sided Crystal Ball function fit of $\frac{\text{Reco } H_{\mathrm{T},2}/2}{\text{Gen } H_{\mathrm{T},2}/2}$ in each Gen $H_{\mathrm{T},2}/2$ bin overlayed by Gaussian fitting the core of the resolution (red line).

as the resolution which can in good approximation be described by the σ of a Gaussian fit to the core of distribution. To take into account the non-Gaussian tails of the jet response distribution, a double-sided Crystal-Ball function is used. The resolution as a function of $H_{\mathrm{T},2}/2$ is calculated separately for both $\mathrm{n_j} \geq 2$ and $\mathrm{n_j} \geq 3$ events. A fit example for one Gen $H_{\mathrm{T},2}/2$ bin is shown in Fig. 1.8 for $\mathrm{n_j} \geq 2$ (left) and inclusive 3-jet events(right). Here the black dots represent the jet response distribution and the double-sided Crystal-Ball fit (blue line) is overlayed by the Gaussian fit (red line). The resolution is then plotted as a function of Gen $H_{\mathrm{T},2}/2$. As expected, it has been observed from Fig. 1.9 that the Crystal Ball function better describes the measured distributions, especially in the low-$H_{\mathrm{T},2}/2$ region where the non-Gaussian tails are more pronounced. Hence the Crystal Ball function is preferred to determine the resolution.

Figure 1.10 shows the final relative resolution which is described by a modified version of the NSC formula as mentioned in Equation 1.8. Also, the extrapolation of the fit function upto 2 TeV is shown. The formula is based on the usual NSC

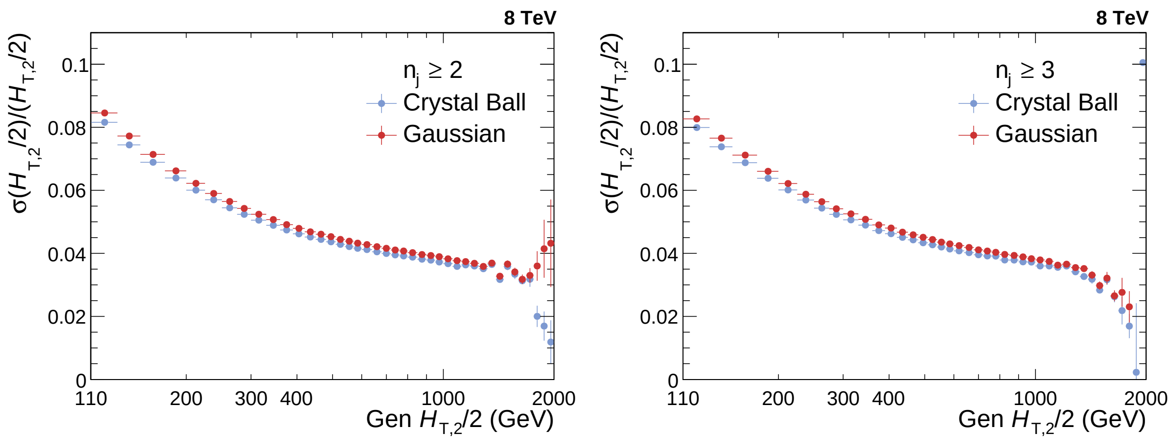


Figure 1.9: Comparison of jet energy resolution calculated using Crystal-Ball fit and Gaussian fit functions for inclusive 2-jet events (left) and for inclusive 3-jet events (right).

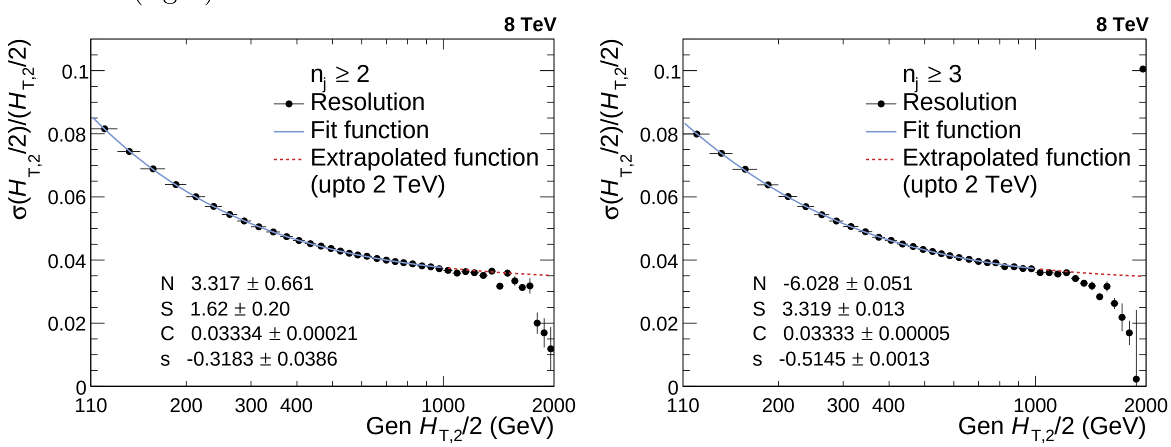


Figure 1.10: Jet energy resolution (JER) as a function of Gen $H_{\mathrm{T},2}/2$ for inclusive 2-jet events (left) and for inclusive 3-jet events (right). The solid line gives the results of the fits using the NSC-formula.

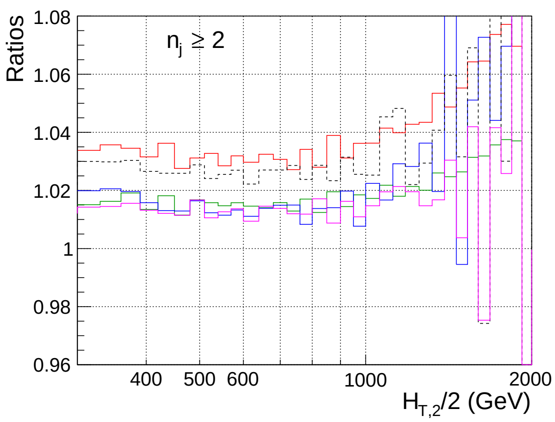
formula which describes the resolution in terms of noise N, a stochastic component S and a constant term C. Especially in the low $H_{\mathrm{T},2}/2$ region in which the tracking has a non-negligible influence on the resolution due to the particle flow algorithm, a slightly better fit is obtained by using the modified resolution formula. Table 1.6 gives the parameters of the fit for the inclusive 2-jet and inclusive 3-jet events.

$$\frac{\sigma(x)}{x} = \sqrt{sgn(N) \cdot \frac{N^2}{x^2} + S^2 \cdot x^{s-1} + C^2} \tag{1.8}$$

While calculating JER where we are using one large rapidity bin, it has been observed that from Figure 1.11, when JER extracted from simulated MG+ P6 Reco/MG+ P6 Gen is used in toyMC for smearing, it smears the fastNLO too much (red curve). The extracted JER also smears MC Gen more as seen from Smeared MG+ P6 Gen/MG+ P6 Gen (black dashed curve). When the 30% reduced JER is used to smear MG+ Gen, the Smeared MG+ Gen/MG+ P6 Gen, matches with simulated MG+ P6 Reco/MG+ P6 Gen. Also, toyMC Gen smeared with 30% reduced JER, Smeared FastNLO/Gen FastNLO. So an additional unfolding uncertainty is attributed by comparison to 30% reduced JER.

Table 1.6: The fit parameters of the resolution for inclusive 2-jet and inclusive 3-jet events.

	N	S	C	s
Inclusive 2-jet	3.32	1.62	0.0333	-0.318
Inclusive 3-jet	-6.03	3.32	0.0333	-0.515



Plots_HT_2_150/Ratio_all_3_crystal.pdf

Figure 1.11: Simulated MG+ P6 Reco/MG+ P6 Gen to extract JER, Smeared FastNLO/Gen FastNLO using extracted JER, Smeared MG+ P6 Gen/MG+ P6 Gen using extracted JER, smeared MG+ P6 Gen/MG+ P6 Gen using 30% reduced extracted JER, Smeared FastNLO/Gen FastNLO using 30% reduced extracted JER; for inclusive 2-jet (left) and inclusive 3-jet events (right).

Bibliography

[1] C. Collaboration, "Jet Identification at 8 TeV." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID, 2012. (accessed on 2017-10-31).

[2] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," *JHEP*, vol. 06, p. 128, 2011.

[3] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[4] J. Alwall *et al.*, "A Standard format for Les Houches event files," *Comput. Phys. Commun.*, vol. 176, pp. 300–304, 2007.

[5] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934.

[6] C. Collaboration, "Jet Performance in pp Collisions at 7 TeV," 2010.

[7] C. Collaboration, "Jet Energy Resolution at 8 TeV." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution, 2012. (accessed on 2017-10-31).

Selected Reprints