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

Anterpreet Kaur
Supervisor : Prof. Manjit Kaur

Department of Physics
Panjab University
Chandigarh, India

Outline

- Physics of the Thesis
- Experimental Set-up
- Theoretical Background
- Software Tools
- Analysis Strategy
- Measurement of Inclusive Multijet Cross-sections
- Theoretical Predictions
- Data-Theory Comparisons
- Cross-section Ratio R_{32}
- Determination of the Strong Coupling Constant, $\alpha_S(M_Z)$
- Summary and Outlook
- Hardware Activities

- Hadrons colliding at very high center-of-mass energies provide a direct probe to the nature of the underlying parton-parton scattering physics.
- **Scattering → Fragmentation → Hadronization → Clustering into Jets**
- Jets being the final structures observed in the detector :
  - relate experimental observations to theory predictions.
  - preserve the energy and direction of the initial partons.
  - serve as a direct test of theory of strong interactions, **Quantum Chromodynamics**.
- Inclusive multijet production cross-section provides the details of parton distribution functions (PDF) of the colliding hadrons and precise measurement of the strong coupling constant α_S :

$$\sigma_{\text{i-jet}} = \sigma(\mathrm{pp} \rightarrow \mathrm{i\ jets} + \mathrm{X}) \propto \alpha_{\mathrm{S}}^{\mathrm{i}}$$

- Cross sections ratios : $\mathrm{R_{mn}} = \frac{\sigma_{\text{m-jet}}}{\sigma_{\text{n-jet}}} \propto \alpha_{\mathrm{S}}^{\mathrm{m-n}}\ ;\ \mathrm{m} > \mathrm{n}$
- **Why cross-section ratios ?**
  - better choice w.r.t. inclusive cross-sections because of less dependency on uncertainties such as :
  - uncertainties due to luminosity,
  - scale dependence,
  - PDF dependence etc.
- A measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section as a function of the average transverse momentum, $< \mathrm{p_{T1,2}} >$, of the two leading jets in the event is performed at 7 TeV (Eur. Phys. J. C 73 (2013) 2604).

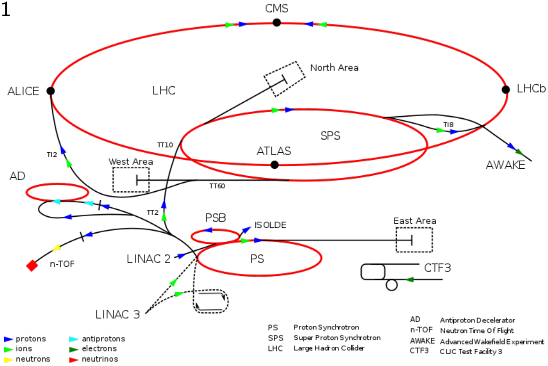
Experimental Set-up

- World's biggest and the most powerful particle accelerator and collider built by CERN (European Organization for Nuclear Research).
- Located in a 27 km long tunnel, 100 m (approx.) underground at Swiss-France border.
- **Proton-proton (p-p)**, proton-lead and lead-lead as well as xenon-xenon nuclei collisions take place at different center-of-mass energies ($\sqrt{s}$).

  - **3.5 TeV** per beam in 2010 and 2011
  - **4.0 TeV** per beam in 2012
  - **6.5 TeV** per beam at present.

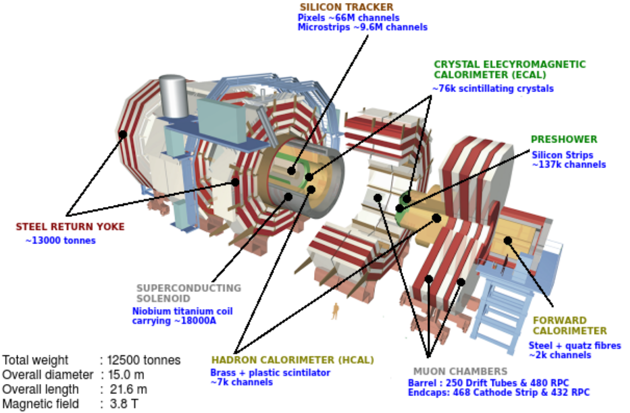
- Four big experiments located at different interaction points.

  - **ATLAS & CMS :** General purpose
  - **ALICE & LHCb :** Dedicated

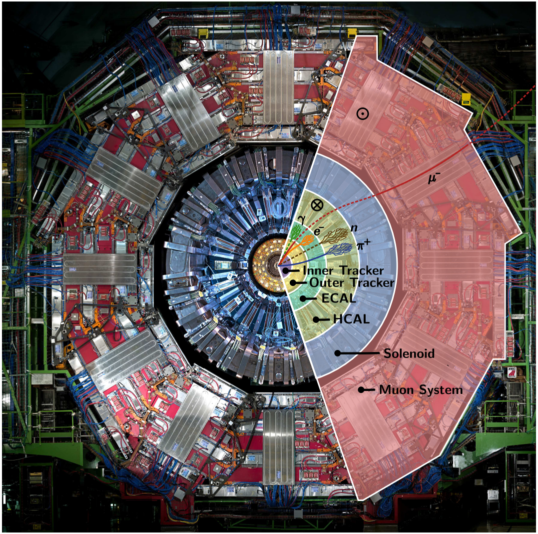


CMS Detector

- Compact Muon Solenoid (CMS) : One of the general purpose detectors which
  - is compact in size : 21 m long, 15 m wide and 15 m high weighing 12,500 tonnes
  - emphasis on the detection of muons
  - is enclosed within high solenoidal magnetic field
- Complex layered structure of sub-detectors : Tracker, Electromagnetic Calorimeter (ECAL), Hadron Calorimeter (HCAL) and Muon System.



- Major fraction of the particles produced in p-p collisions is hadrons.
- Collimated in the form of conical structures called Jets.
- Both hadronic (charged and neutral) and electromagnetic components.
- Charged Particles : energy deposits in the calorimeters.
- Neutral Particles : missing transverse energy.
- HCAL : detects neutral particles; an essential sub-system of the CMS detector and contributes to most of physics studies with CMS.



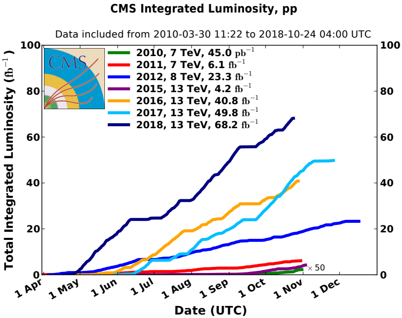
$\odot$ - Direction of magnetic field in the return yoke
$\otimes$ - Direction of magnetic field inside the solenoid

- Luminosity ($\mathcal{L}$) gives the rate at which collisions occur.
- Number of collisions produced in a detector per cm^2 and per second.
- $\mathcal{L}$ is related to total number of events N of a process over a time period T and σ as :

$$N = \int_0^T \mathcal{L}\ \sigma\ dt = \mathcal{L}_{int}\ \sigma$$

where $\int_0^T \mathcal{L}\ dt\ =\ \mathcal{L}_{int}$ gives the total integrated luminosity, expressed in barn^{-1} units*

- $\mathcal{L}_{int}$ gives a direct indication of number of events produced in a process.
  - For e.g. an integrated luminosity of 10 fb^{-1} means that 10 events are produced in a process having cross-section equal to 1 fb.
- CMS constantly monitors the instantaneous luminosity delivered by the LHC.
- The absolute luminosity is measured using van-der-Meer scans done in special runs of the LHC.
- The measured integrated luminosity for 2012 data set is 19.7 fb$^{-1} \pm \begin{matrix} 2.5\%(syst.) \\ 0.5\%(stat.) \end{matrix}$

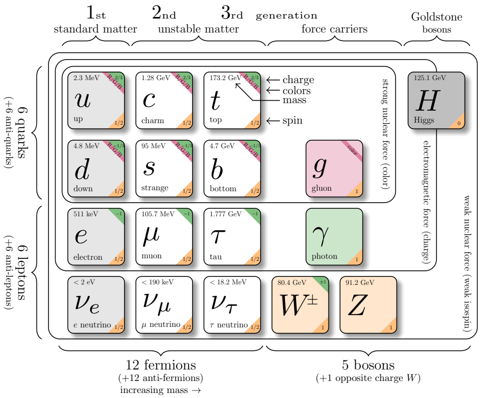


*1 barn = 10^{-28} m^2

Theoretical Background

Standard Model

- A theoretical model to describe the nature and properties of fundamental particles and their interactions.
- Fundamental particles :
  - Fermions
    - Quarks (q)
    - Leptons (ℓ)
  - Bosons
    - Gauge Bosons
    - Scalar Boson
- Forces of interactions :
  - Electromagnetic - photon (γ)
  - Strong - gluons (g)
  - Weak - $W^{\pm}$ and Z
  - Gravity* - graviton

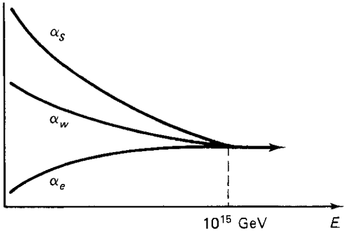


- Described by $\underbrace{SU(3)_C}_{\text{Strong}} \otimes \underbrace{SU(2)_L \otimes U(1)_Y}_{\text{ElectroWeak}}$ gauge symmetry

*not been incorporated into Standard Model yet

Quantum Chromodynamics (QCD)

- A non-abelian gauge theory describing the strong interactions between quarks and gluons.
- Color charge -
  - Quarks and gluons : colored
  - Hadrons - Baryons (qqq) and Mesons ($q\bar{q}$) : colorless
- Strong coupling constant α_S - a fundamental parameter giving the strength of interaction.
- Properties -
  - Confinement : at low energies $\rightarrow$ Non-perturbative (NP) regime
  - Asymptotic freedom : at high energies $\rightarrow$ Perturbative regime



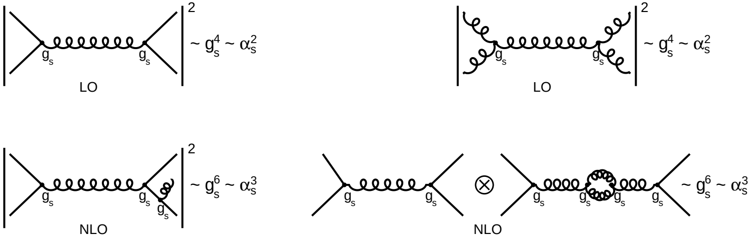
- In perturbative QCD

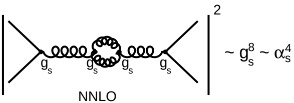
$$X = \sum_{i=0}^{N} \alpha_s^n c_i = c_0 + \alpha_s^1 c_1 + \alpha_s^2 c_2 + \ldots$$

is determined by summing over the amplitudes of all Feynman diagrams.

  - Power of α_S : the number of vertices associated with q-g or g-g interactions.

Different orders in a 2 → 2 scattering process





$\alpha_S(Q) = g_s^2/4\pi$

- Ultraviolet (UV) divergences : Calculations become complex with the loop diagrams; associated integrals are divergent.
- Renormalization : a mathematical procedure which
  - allows finite calculation of momenta integrals of virtual loop by removing UV divergences.
  - introduces a regulator for the infinities - renormalization scale μ_r.
  - redefines or renormalizes the coupling constant to absorb the UV divergences.

Running of the Strong Coupling

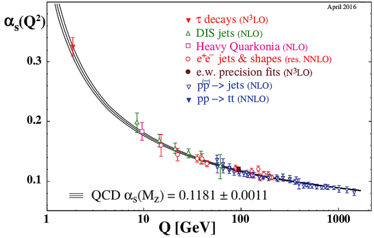
- Renormalization group equation (RGE) : exact dependence of $\alpha_S(\mu_r^2)$ on μ_r
  - Dependence of X on μ_r must cancel $\rightarrow$

$$\mu_r^2 \frac{d}{d\mu_r^2} X\left(\frac{Q^2}{\mu_r^2}, \alpha_s(\mu_r^2)\right) = 0$$

  - First order solution of RGE is :

$$\alpha_s(\mu_r^2) = \frac{1}{b_0 \ \ln\left(\mu_r^2/\Lambda_{QCD}^2\right)}$$

  - Coupling becomes large at the scale Λ_{QCD}
  - For $b_0 > 0$, coupling becomes weaker at higher scales $Q \rightarrow$ asymptotic freedom



- It is always convenient to express α_S at some fixed scale -
  - Some of the best measurements come from Z decays : the strong coupling is determined at the scale of the Z boson mass $\alpha_S(M_Z)$ given by

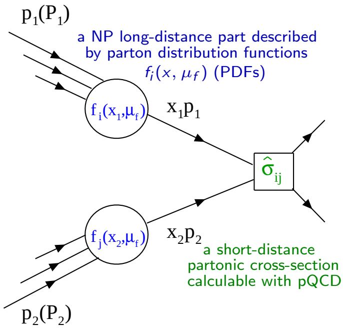
$$\alpha_S(\mu_r, \alpha_S(M_Z)) = \frac{\alpha_S(M_Z)}{1 + \alpha_S(M_Z) b_0 \ln(\mu_r^2/M_Z^2)}$$

- α_S : a free parameter of QCD theory
  - must be extracted from experimental measurements
  - evolved to the scale of the Z boson
  - current world average value*is $\alpha_S(M_Z) \ = 0.1181 \pm 0.0011$

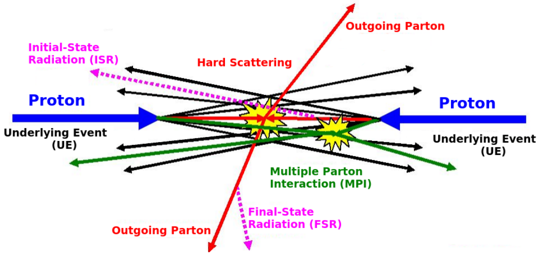
*Particle Data Group (PDG)

- Proton - three valence quarks (uud), gluons and the sea quarks.
- Cross-section (σ) of a certain process : the probability that the two hadrons interact and give rise to that final state.
- In a hadronic collision, the perturbation theory is only valid at the parton-level.
- Factorization theorem - allows the calculation of σ by separating into two parts :

$$\sigma_{P_1P_2\to X} = \sum_{i,j} \int dx_1 dx_2 f_{i,P_1}(x_1,\mu_f) f_{j,P_2}(x_2,\mu_f) \times \hat{\sigma}_{ij\to X}\left(x_1 p_1, x_2 p_2, \alpha(\mu_r^2), \frac{Q^2}{\mu_f^2}\right)$$



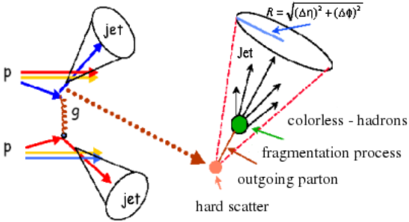
  - Proton PDFs f_i and f_j : probability to find a parton i with momentum fraction x within a hadron.
  - Factorization scale, μ_f : corresponds to the resolution with which the hadron is being probed.
    - Particles emitted with $p_{\mathrm{T}} > \mu_f$ are considered in the calculation of hard scattering perturbative coefficients.
    - Particles emitted with $p_{\mathrm{T}} < \mu_f$ are accounted for within the PDFs.



- **Hard Scattering :**
  - Parton Shower : collinear parton splitting and the soft gluon emissions (pQCD).
  - Hadronization : confinement of colored quarks and gluons into the color-neutral composite particles called hadrons (non-pQCD).
- **Underlying Event :** Event structure is significantly more complex than that of the lepton collisions.
  - **Initial and Final-State Radiations (ISR, FSR) :** two incoming as well as outgoing partons can also develop parton showers.
  - **Multiple Parton Interactions (MPI) :** remaining two partons which do not participate in a hard collision may also interact.

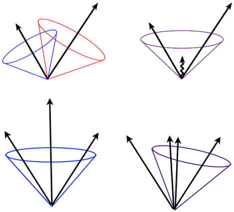
Jets

- Hadrons produced in p-p collisions get collimated in the form of conical structures called "jets" of radius 'R'.
- Direction of the jet is towards the direction of the initial partons that originated them.
- Detectable objects which relate experimental observations to theory predictions formulated in terms of partons.



- Act as a bridge between the elementary quarks and gluons of QCD and the final hadrons produced in high energy collisions.
- At the LHC, **Dijet Events** : $2 \rightarrow 2$, **Multijet Events :** $2 \rightarrow 3$, $2 \rightarrow 4$
- Jets and their observables are the best tools to test the predictions of pQCD :
  - Jet production cross-section : helps to extract the value of α_S and to reduce the uncertainties of the PDFs of the proton.
  - Inclusive multijet event cross-sections permits more elaborate tests of QCD.
  - A precise study of jet variables helps to understand the signal and background modelling for new physics searches in hadronic final states.

- Jet algorithms provide a set of rules which determine how the particles can be clustered into a jet.
- Parameters involved indicate how close two particles must be to belong to the same jet :
  - Cone algorithms $\rightarrow$ closeness in coordinate space
  - Sequential Recombination algorithms $\rightarrow$ closeness in momentum space
- Recombination Scheme
- Infrared safety : Addition of a soft emission should not change the number of hard jets found in an event.
- Collinear safety : Collinear splitting should not modify the number of jets formed in an event.
- Three types of Sequential Recombination algorithms :
  - k_t : involves clustering of soft particles first; susceptible to the underlying and pileup events
  - Cambridge/Aachen (C/A) : involves energy independent clusterings.
  - **anti-k_t :** cluster hard particles first; less sensitive to underlying and pileup events.



Software Tools

Software Tools

- CMS software (CMSSW) framework : a dedicated data structure and software tools.
  - Performs calibration, event generation, detector simulation, event reconstruction etc. by implementing the codes either in C++ or Python languages.
- ROOT : an open source object-oriented data analysis framework.
  - Consists of a huge C++ library which store and analyze large amounts of the data.
  - Provides histrogramming methods in 1, 2 and 3 dimensions, curve fitting functions, minimization procedures, graphics and visualization classes.
- FastJet : a software C++ package.
  - Provides a broad range of jet finding, determination of jet areas, estimation of pileup and underlying-event noise levels etc.
- NLOJet++ : a C++ program to evaluate jet production cross-sections at leading order (LO) and next-to-leading order (NLO).
  - Calculations of pQCD cross-sections using Monte Carlo integration methods are very time consuming.
  - PDF fits or estimations of uncertainties becomes difficult where the calculations of the cross-sections are needed to be repeated.
- fastNLO : fast re-evaluations of cross-sections in interface with NLOJet++.
  - The strong coupling constant and the PDFs can be changed without a re-calculation of the perturbative coefficients.

Monte Carlo (MC) Simulations

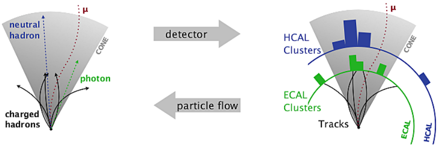
- PYTHIA: most widely used program to generate the collisions at high energies.
  - Uses LO calculations to derive the colored partons from the hard interaction.
  - Uses the Lund string hadronization model to describe hard and soft interactions and parton showers.
  - PYTHIA6 with tune Z2* and PYTHIA8 with tunes CUETS1 and CUETM1 have been used.
- MADGRAPH : generates LO matrix elements for high energy physics processes.
  - Stores the event information of the particles in the Les Houches format.
  - MADGRAPH5 has been interfaced to PYTHIA6 with tune Z2* - used here mainly for general comparisons to the data and calculating the detector resolution.
- HERWIG (Hadron Emission Reactions With Interfering Gluons) : a multi-purpose LO event generator.
  - Uses angular ordering for parton showers and cluster model for hadronization.
  - HERWIG++ with the default tune of version 2.3 has been used to study the non-perturbative effects.
- POWHEG (Positive Weight Hardest Emission Generator) : performs the fixed NLO calculations merged with parton showers.
  - Uses POWHEG BOX to implement NLO calculations in shower MC programs.
  - POWHEG has been interfaced to PYTHIA8 with tunes CUETS1 and CUETM1 to include the parton shower and hadronization.

Simulation and Reconstruction

- Detector Simulation - a computer program which takes the particles generated by MC event generators.
    - Defines the detector system, its geometry, material and electronics properties.
    - Full Simulation - based on a C++ simulation toolkit GEANT4 (GEometry ANd Tracking); handles the interactions of particles with matter over a wide range of energy.
    - Fast Simulation - detector effects are parametrized instead of simulating; events are produced at much faster rates.
- Digitization : The simulated detector response is then transformed into a digital signal with the help of electronics.
- Event Reconstruction : identifies the particles passing through the detector by interpreting the electrical signals produced in digitization.
    - Analysis-level objects are created by combining recorded signals from the tracker, energy deposits from calorimeters and muon detectors.
    - Particle Flow (PF) Algorithm combines the information from individual sub-detectors.
- Track Reconstruction : CMS uses an iterative tracking algorithm.
    - Quality criteria is used to reject the badly reconstructed tracks and to decrease the fake rate.
- Primary Vertex Reconstruction : identification of the primary vertex of the main hard interaction is crucial.
    - Track assigned to only one vertex $\rightarrow$ **hard interaction**.
    - Track assigned to more than one vertex $\rightarrow$ **soft interaction**.

Jet Reconstruction

- PF event reconstruction algorithm converts the detector signals back to physical objects and their energy is measured :
  - **Electrons :** from the track momentum, corresponding ECAL energy deposits and the energy sum of all bremsstrahlung photons associated with the tracks.
  - **Muons :** from the curvature of the tracks in tracker and muon chambers.
  - **Photons :** directly from the ECAL measurements.



  - **Charged hadrons** : from track momentum and corresponding energy clusters in ECAL and HCAL.
  - **Neutral hadrons** : from calibrated ECAL and HCAL energies.
- **Jets :** are reconstructed from the collection of PF objects.
  - **Generator Jets (GenJets) :** stable particles generated by the MC event generators.
  - **Calorimetric Jets (CaloJets) :** energy deposits in the ECAL and HCAL towers.
  - Particle Flow Jets (PFJets) : detector level jets from particle flow candidates.
    - Use of the tracker and high granularity of the ECAL gives better energy resolution.
    - PFJets perform better than CaloJets and are the standard jets used at CMS.

Measurement of the Differential Inclusive Multijet Cross-sections and their Ratio R_{32}

Analysis Strategy

- Jets are reconstructed from particle flow objects using the anti-k_t clustering algorithm with the size parameter R $= 0.7$.
- Event Selection : Appropriate selection criteria has been designed for choosing the best event for analysis. This measurement uses jets with $p_{\mathrm{T}} > 150$ GeV and $|y| < 2.5$.
- Pileup reweighting.
- Detector level comparison of differential inclusive 3-jet and 2-jet cross-sections in terms of defined observable for full Data, MC and theory predictions.
- Unfolding set-up and the measured cross-sections are corrected for detector smearing effects by using the Unfolding technique.
- Evaluated cross-section ratio, R_{32}.
- Evaluated experimental and systematic uncertainties from different sources.
- Included NLO pQCD calculations obtained using different PDF sets and corrected them for MPI and HAD effects by applying non-perturbative (NP) corrections as well as for electroweak (EW) effects.
- Evaluated theoretical uncertainties from different sources.
- Compared theoretical results with that of data.
- Extracted the value of $\alpha_S(M_Z)$ by fitting cross-sections as well as cross-section ratio, R_{32}.
- Studied the running of $\alpha_S(Q)$ as a function of Q.

- The scale based on the transverse momentum of the jets is used :

$$\langle p_{\mathrm{T}_{1,2}} \rangle = \frac{p_{\mathrm{T},1} + p_{\mathrm{T},2}}{2} = H_{\mathrm{T},2}/2$$

- Inclusive differential multijet event cross section (pb/GeV) is defined as :

$$\frac{d\sigma_{n-\mathrm{jet}}}{d(H_{\mathrm{T},2}/2)} = \frac{1}{\epsilon\ \mathcal{L}_{\mathrm{int,eff}}} \frac{N_{\mathrm{event}}}{\Delta(H_{\mathrm{T},2}/2)}$$

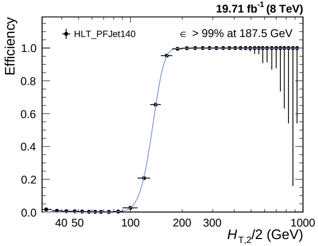
- **Inclusive *n*-jet** event samples include the events with **number of jets ($\mathrm{n_j}$)** $\geq n$.
  - $n = 2 \rightarrow$ Inclusive 2-jet events ($\mathrm{n_j} \geq 2$)
  - $n = 3 \rightarrow$ Inclusive 3-jet events ($\mathrm{n_j} \geq 3$)

- Cross-section ratio is defined as :

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

- Samples used :
  - **Data** collected at $\sqrt{s} = 8$ TeV during 2012 run; Integrated Luminosity : 19.7 fb^{-1}
  - **Simulated Monte-Carlo (MC)** samples using MADGRAPH5 + PYTHIA6 Tune Z2* (MG5+P6 Z2*), HERWIG++ and PYTHIA6 generators.
  - **Theoretical NLO** calculations using the NLOJET++ program (v4.1.3) within the framework of the FASTNLO package (v2.3) using different PDF sets.

- CMS implements a two-level trigger system to reduce the amount of recorded events to a sustainable rate.
- Five **single-jet high-level triggers** : select an event in which at least one jet has the transverse momentum above the threshold.
- Trigger efficiency for HLT_PFJetY :



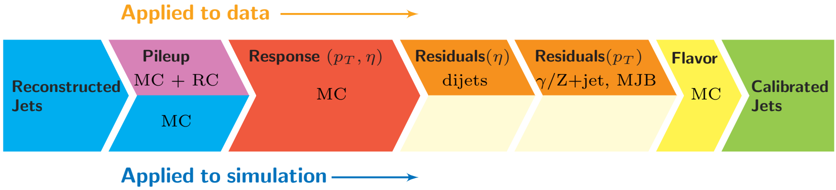
$$\epsilon_{\mathrm{HLT\_PFJetY}} = \frac{\mathrm{H_{T,2}/2}\big(\mathrm{HLT\_PFJetX} + (\mathrm{L1Object\_p_T} \geq \mathrm{Z}) + (\mathrm{HLTObject\_p_T} \geq \mathrm{Y})\big)}{\mathrm{H_{T,2}/2}\big(\mathrm{HLT\_PFJetX}\big)}$$

  - the value of X is chosen previous to that of Y in p_{T} ordering
  - Z is the L1 seed value corresponding to the trigger path
- Trigger regions defined as ranges of the $H_{\mathrm{T},2}/2$ for every single-jet trigger used :

Trigger Path	**L1 threshold (GeV)**	**HLT threshold (GeV)**	**$H_{\mathrm{T},2}/2$, 99% (GeV)**	**Eff. Lumi (fb^{-1})**
HLT_PFJet80	36	80	120.0	0.0021
HLT_PFJet140	68	140	187.5	0.0560
HLT_PFJet200	92	200	262.5	0.2600
HLT_PFJet260	128	260	345.0	1.0600
HLT_PFJet320	128	320	405.0	19.7100

- Measured energy of jets cannot be directly translated to the energy at true particle or parton level.
  - Non-linear and non-uniform response of the calorimeters, effects of pileup and small residual effects in the data remaining after the corrections based on MC simulations.
- CMS follows a factorized approach :

$$p_T^{\mathrm{corr}} = c_{\mathrm{res}}(\eta, p_T'') \cdot c_{\mathrm{mc}}(\eta, p_T') \cdot c_{\mathrm{pileup}}(\eta, \rho, \mathrm{A}_j, p_T^{\mathrm{raw}}) \cdot p_T^{\mathrm{raw}}$$



- Corrections are applied to jets in both the data# as well as in simulated events*.
  [Details in Back-up slide 40]

Winter_V8 jet energy corrections
* START53_V27 jet energy corrections

- To observe the extremely rare events, the event rate in a collider should be very high.

High delivered luminosity

↓

Increase in the number of bunches or protons per bunch

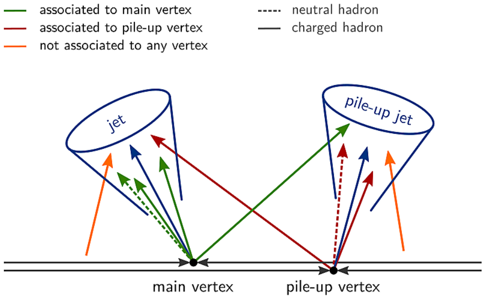
↓

Pileup (PU) interactions

↓

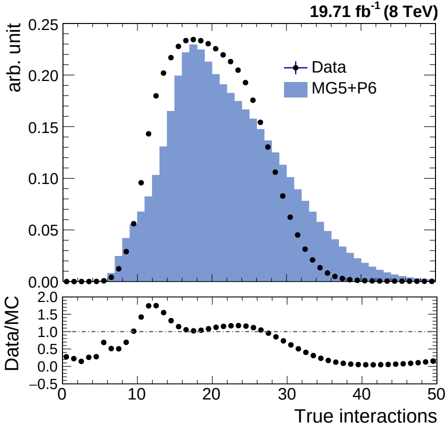
Low p_{T} jets

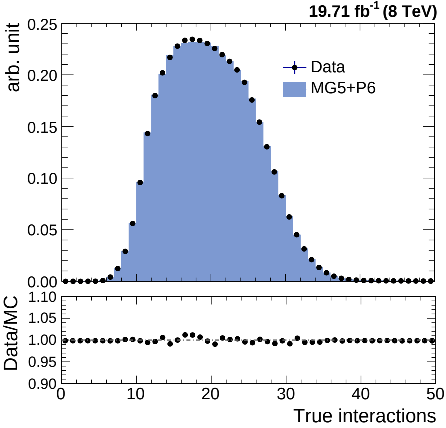
- **In-time pileup :** Additional collisions within a single bunch crossing.
- **Out-of-time pileup :** Additional collisions coming from other bunch crossings.



- Number of pileup interactions are taken into account in generating the official MC samples.
- Pileup events implemented in the simulation $N_{\mathrm{MC}}(N_{\mathrm{PU,truth}})$ does not match exactly the one measured in the data $N_{\mathrm{data}}(N_{\mathrm{PU,est.}})$.
- A reweighting factor w_{PU} is applied to the simulated events :

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

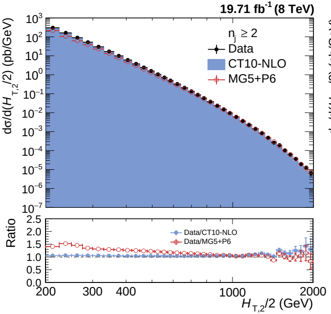


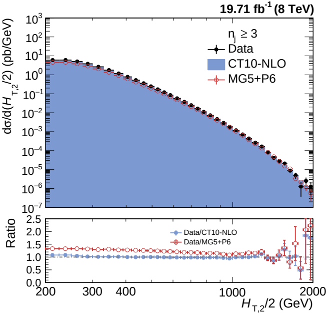


Detector Level Comparisons

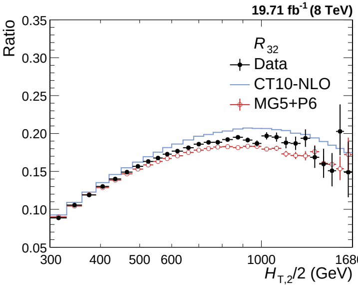
- Data are compared to the sample of **MG5+P6 Z2*** simulated events.

- LO MC generator roughly describes the spectrum on detector level.





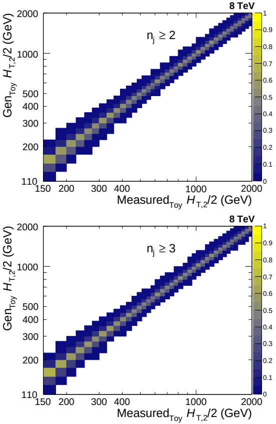
- Cross-section ratio* R_{32} :
  - Numerator and denominator are not independent samples.
  - Statistical uncertainty is calculated using the Wilson score interval method which takes into account the correlation between the numerator and the denominator.



*Due to kinematical constraints, the minimum cut on $H_{\mathrm{T},2}/2$ is 300 GeV.

Unfolding

- Jet energy resolution (JER) : Finite value of the resolution of the detector because of differences of the measured quantity from its true value.
    - Given by the width of a Gaussian distribution, centered around the true value of the measured quantity.[Details in Back-up slides 42-41].
- The finite detector resolution along with the steeply falling jet p_{T} spectrum distorts the measured cross sections.
- Each p_{T} bin content contains the migrated events from neighbouring bins along with the original events.
- Unfolding of the data allows direct comparison of experimental measurements with theory predictions or with the results from other experiments.
- Unfolding uses a **response matrix (RM)** that maps the true distribution onto the measured one.
    - **Fitting** the theoretically predicted NLO spectrum to get true $H_{\mathrm{T},2}/2$ spectrum. [Details in Back-up slides 44-45].
    - **Forward Smearing** is performed using the additionally smeared MC JER to obtain measured $H_{\mathrm{T},2}/2$ spectrum.
- The measurements are unfolded by using the iterative D'Agostini method with **4 iterations**, implemented in the RooUnfold software package.



Unfolding

- Unfolding R_{32} :

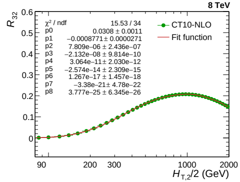
  - **Method I :** First unfold separately the inclusive 2-jet and 3-jet measured cross-sections and then construct R_{32}.

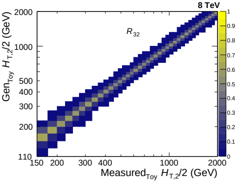
  - **Method II :** Unfold directly the cross-section ratio R_{32}.

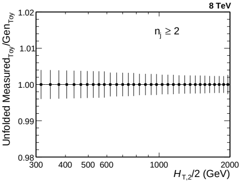
- Closure test : Ratio of the unfolded data to the measured data with $\mathrm{H}_{\mathrm{T},2}/2 > 300$ GeV

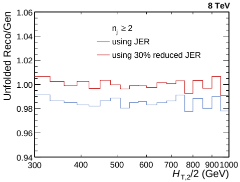
- Unfolding Cross-sections Statistical errors in the unfolded distributions are slightly higher

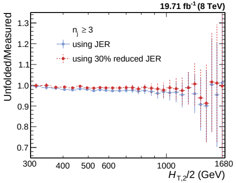
- Unfolding Ratio : Correlation of the statistical uncertainty introduced by the unfolding procedure

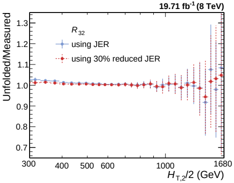


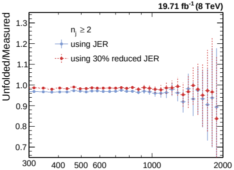


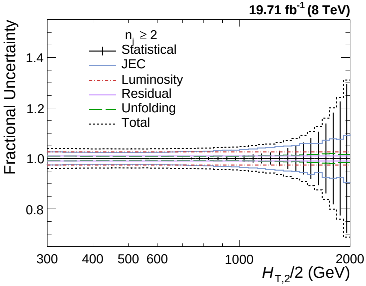


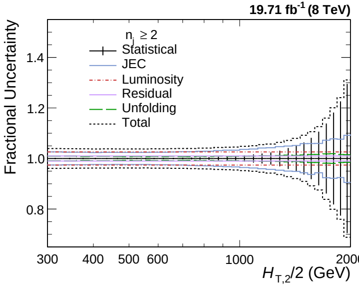




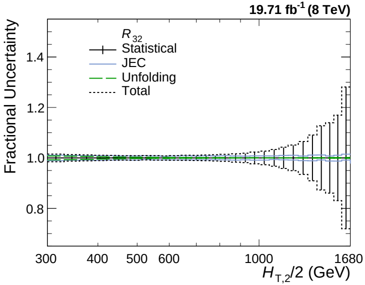








Uncertainty Source	**Inclusive 2-jet**	**Inclusive 3-jet**	R_{32}
Statistical	<1 to 30%	<1 to 30%	<1 to >50%
JEC	3 to 10%	3 to 8%	1 to 2%
Unfolding	1 to 2%	1 to 2%	1%
Luminosity	2.6%	2.6%	cancels
Residual uncorrelated	1%	1%	cancels



- Unfolding systematic uncertainty :
  - JER uncertainty : varying scale factors
  - Additional uncertainty : 30% reduced resolution
  - Model dependence : different functions to fit NLO predicitions to obtain the true spectrum (Back Up slide no. 32)

Back-up slides

- Based on transverse momentum p_{T} of the particles.
  1. Distance d_{ij} between two particles i and j and distance d_{iB} of the particle to the beam are calculated as

$$d_{ij} = \min\left(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}j}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = p_{\mathrm{T}i}^{2p}$$
$$\text{where} \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{1}$$

  2. If $d_{ij} < d_{iB}$, particles i and j are merged into a new single jet object k, summing four-momenta of two initial particles by recombination scheme and step 1 is repeated.
  3. If $d_{iB} < d_{ij}$, particle i is declared as a final-state jet and the particle gets removed from the list.
- Value of the parameter p defines the three different sequential algorithms :
  - k_t algorithm : $p = \mathbf{1}$
  - Cambridge/Aachen (C/A) algorithm : $p = \mathbf{0}$
  - anti-k_T algorithm : $p = \mathbf{-1}$

Event Selection

- anti-k_t particle flow (PF) jets with R $= 0.7$.
- At least one good primary vertex : $|$z(PV)$| < 24$ cm, ρ(PV) < 2 cm, *ndof* > 4
- Official tight jet ID recommended by JETMET group is used.
- All jets having $p_{\mathrm{T}} > 150$ 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 $|y| < 2.5$.
- $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum E_{\mathrm{T}}} < 0.3$ to protect against mismeasured or background events with large missing E_{T}.

Datasets & MC Samples

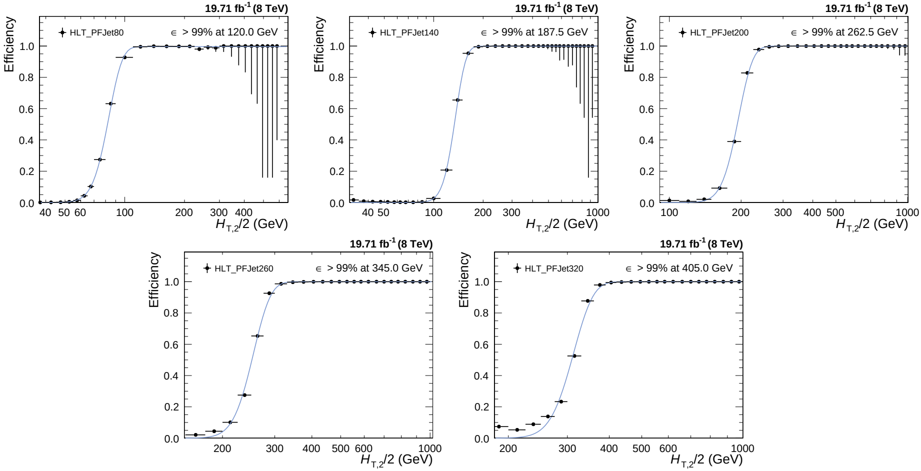
- **Data :** Collected at $\sqrt{s} = 8$ TeV during 2012 run; Integrated Luminosity : **19.7 fb^{-1}**

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

- **Simulated Monte-Carlo (MC) Samples :**

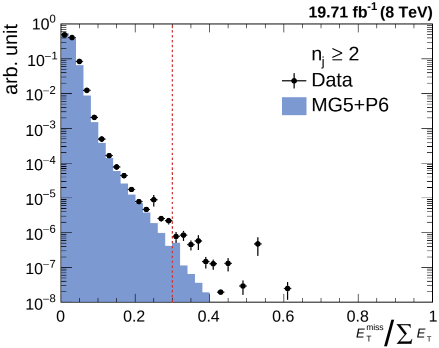
Generator	Sample	Events	Cross-section (pb)
MADGRAPH5 + PYTHIA6 Tune Z2* (MG5+P6 Z2*)	/**QCD_HT-100To250**_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	50129518	1.036×10^7
	/**QCD_HT-250To500**_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	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
HERWIG++	/**QCD_Pt-15to3000**_TuneEE3_Flat_8TeV_herwigpp/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM		
PYTHIA6	/**QCD_Pt-15to3000**_TuneZ2star_Flat_8TeV_pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM		

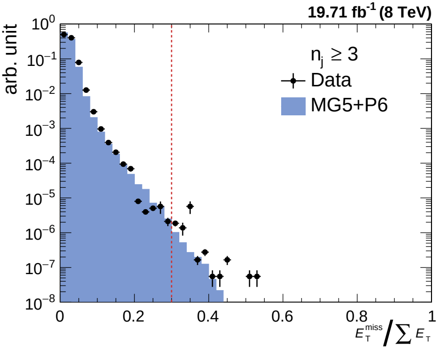
- **Theoretical NLO** calculations using the NLOJET++ program (v4.1.3) within the framework of the FASTNLO package (v2.3) using different PDF sets.



Uncertainty on the efficiency is calculated using Clopper-Pearson confidence intervals :

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


19.71 fb-1 (8 TeV)
arb. unit
$n_j \geq 2$
Data
MG5+P6
$E_T^{miss} / \sum E_T$


19.71 fb-1 (8 TeV)
arb. unit
$n_j \geq 3$
Data
MG5+P6
$E_T^{miss} / \sum E_T$

Jet Energy Corrections

- Pileup Corrections
  - Due to additional p-p collisions within the same bunch-crossing.
  - Corrections are determined by simulating a sample of QCD dijet events with and without pileup effects.
  - c_{pileup} is calculated from jet area method using the pileup density ρ in the event and the jet area A_j.
- MC Corrections
  - Due to the inefficiencies introduced by the detector simulation.
  - Based on MC simulated QCD events.
  - c_{mc} is derived by comparing the measured jet p_{T} to the particle level jet p_{T}.
- Residual Data Corrections
  - Due to remaining small differences between the data and MC simulations.
  - Applied only to the data.
  - c_{res} is derived from data-driven methods using dijet events in which a probe jet is calibrated using a tag jet.
- Flavor Corrections
  - Correct the jets for flavor dependence (b, τ etc.) and are optional.
  - Extracted using Z+jet and photon+jets simulated events.

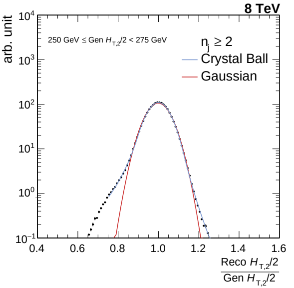
Jet Energy Resolution (JER)

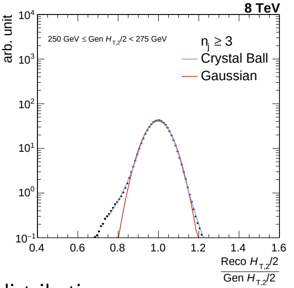
- Jet energy resolution (JER) : Finite value of the resolution of the detector because of differences of the measured quantity from its true value.
- Resolution is given by the width of a Gaussian distribution, centered around the true value of the measured quantity.
- Due to finite resolution of the CMS detector, the measured p_{T} of jets get smeared.
- Measurements show that JER in data is worse than in the simulation and the reconstructed jet p_{T} in MC needs to be smeared more to describe the data.
- Reconstructed jet p_{T} is smeared randomly using a Gaussian function, $f(p_{\mathrm{T}})$ with a width widened by the scaling factor ($c_{central}$) :

$$f(p_{\mathrm{T}}) = a \times \exp\left(-\frac{1}{2}\left(\frac{p_{\mathrm{T}}-\mu}{\sigma}\right)^2 \right)$$

where a is a constant, mean $\mu = 0$,
width $\sigma = \sqrt{c^2_{central} - 1} \cdot \mathrm{JER}(p_{\mathrm{T}}) \times p_{\mathrm{T}}$

- Width of the response, $R = \frac{\text{Reco } H_{\mathrm{T},2}/2}{\text{Gen } H_{\mathrm{T},2}/2}$, gives JER.
- Double-sided Crystal Ball function describes the jet response distribution.

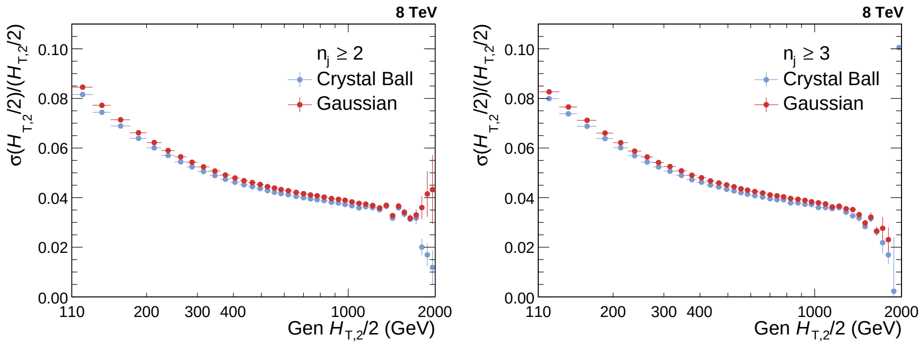




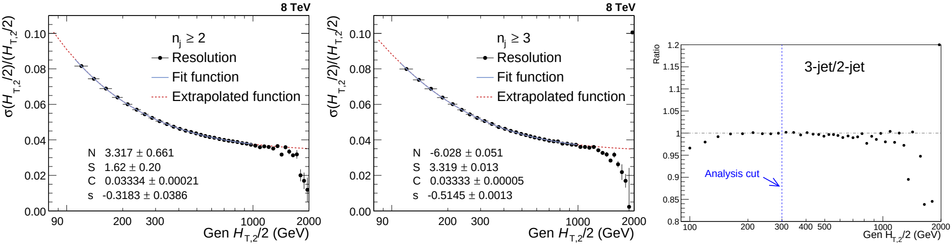
Jet Energy Resolution (JER)

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

- To match the resolution in the data, the reconstructed jet transverse momentum in simulated events need to be smeared by applying the scale factors.
- The uncertainty on the resolution is given by an upwards and downwards variation c_{up} and c_{down} of the measured scaling factor $c_{central}$.



- Fit using NSC-formula $\frac{\sigma(x)}{x} = \sqrt{sign(N) \cdot \frac{N^2}{x^2} + S^2 \cdot x^{s-1} + C^2}$; where $x = H_{\mathrm{T},2}/2$.
- Fits at high $H_{\mathrm{T},2}/2$ start lacking 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

- Resolution is similar in inclusive 3-jet and 2-jet (Right Fig.)
- Fit parameters are used for smearing of NLO spectrum to construct response matrices.

Unfolding : Fitting NLO predictions

- Fitting the NLO $H_{\mathrm{T},2}/2$ spectrum by the function (Function I)

$$f(\mathrm{H_{T,2}}/2\ ) = N(x_T)^{-a}(1 - x_T)^b \times \exp(-c/x_T)$$

where N is normalization factor and a, b, c are fit parameters.

  - This function is derived from the below function from **"Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s}$=7 TeV" (Phys.Rev.Lett. 107, 132001 (2011))**

$$f(p_T; \alpha, \beta, \gamma) = N_0(p_T)^{-\alpha}\left(1 - \frac{2\ p_T\ \ cosh(y_{min})}{\sqrt{s}}\right)^{\beta} \times \exp(-\gamma/p_T)$$

using

$$\alpha = a, \quad \beta = b, \quad \gamma = c * \sqrt{s}/2, \quad x_T = \frac{2{*}H_{\mathrm{T},2}/2{*}cosh(y_{min})}{\sqrt{s}} = \frac{2{*}H_{\mathrm{T},2}/2}{\sqrt{s}}$$

where transverse scaling variable x_T corresponds to the proton fractional momentum x for dijets with rapidity $y = 0$, $\sqrt{s} = 8000$ GeV and y_{min} is low-edge of the rapidity bin y under consideration. (Here y_{min} is taken equal to 0).

- Fitting the NLO $H_{\mathrm{T},2}/2$ spectrum by the function (Function II) (**CMS AN-12-223**) :

$$f(H_{\mathrm{T},2}/2) = A_0\left(1 - \frac{H_{\mathrm{T},2}/2}{A_6}\right)^{A_7} \times 10^{F(H_{\mathrm{T},2}/2)}, \text{where} \quad F(x) = \sum_{i=1}^{5} A_i\left(log\left(\frac{x}{A_6}\right)\right)^i$$

where the parameter A_6 is fixed to $\frac{\sqrt{s}}{2cosh(y_{min})}$, where $\sqrt{s} = 8000$ GeV and y_{min} is the minimum rapidity. The other parameters are derived from the fitting.

- First fit the NLO spectrum with function and then using the obtained fit parameters extrapolated it to lower value.

Function I (Left) and Function II (Right)