Measurement of multi-differential dijet cross sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$.

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik,

Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Luis Ignacio Estévez Baños

aus

Mayabeque, Kuba

Hamburg

2022

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den 18. Januar 2022 — Luis Ignacio Estévez Baños

Gutachter der Dissertation:	Dr. Hannes Jung Prof. Dr. Johannes Haller
Zusammensetzung der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch Prof. Dr. Elisabetta Gallo Dr. Judith Katzy Prof. Dr. Johannes Haller Dr. Hannes Jung
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
Datum der Disputation:	17. May 2022
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Günter H. W. Sigl
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

ABSTRACT

The understanding of the theory of strong interactions in the context of the Standard Model is of great interest to particle physics and its developments. The measurements and the improvement of the theoretical predictions would allow us to get closer to future findings of new physics by discriminating better between backgrounds produced mainly by strong interactions. In this sense, the study of soft and hard radiation, measured as collimated "jets" of strongly interacting particles, needs to be better understood by exploring new phase space regions. In this direction, the Large Hadron Collider (LHC) at CERN is today's most powerful tool to study the strong interactions in proton-proton collisions.

The work presented in this thesis is focused on measuring new multi-differential cross sections to get more insight on how the decorrelation in dijet events is built up by extra radiation. The aim is to measure the cross section as a function of the jet multiplicity (N_{jet}), the azimuthal angle between the two leading jets ($\Delta\phi_{1,2}$), and the transverse momentum of the leading p_{T} jet (p_{T1}). In addition, the cross section as a function of the transverse momentum of the four leading p_{T} jets is also measured. The measurements were done with the data collected by the CMS experiment at the LHC in proton-proton collisions at center-of-mass energy $\sqrt{s} = 13\,\mathrm{TeV}$ during 2016, corresponding to an integrated luminosity of $36.3 fb^{-1}$.

The measurements were then compared to leading order matrix-element (ME) calculations supplemented with parton showers, multiparton interactions, and hadronization. Also, more complex next-to-leading order ME calculations with conventional parton showers and Parton Branching (PB) transverse momentum dependent (TMD) parton densities and PB-TMD initial state shower are compared to the measurements. This thesis is one of the first and very stringent tests done to the PB-TMD method using multi-differential jet observables, which help to further develop this new approach.

The multi-differential cross sections presented in this work were very insightful for the understanding of multijet radiation in the back-to-back region and showing the importance of parton showers. In addition, the importance of higher-order ME contributions in the current theoretical predictions for jets is discussed, and the measurements were found to be of great interest to test the developments of (more complex) multijet merging approaches.

ZUSAMMENFASSUNG

Das Verständnis der Theorie starker Wechselwirkungen im Kontext des Standardmodells ist für die Teilchenphysik und ihre Entwicklungen von großem Interesse. Die Messungen und die Verbesserung in theoretischen Vorhersagen würden es uns ermöglichen, zukünftigen Erkenntnissen in neuer Physik näher zu kommen, indem man besser den physikalischen Untergrund, der hauptsächlich durch starke Wechselwirkungen kommt, trennen kann. In diesem Sinne muss die Untersuchung softer und harter QCD Strahlung, messbar als kollimierte "Jets"-stark wechselwirkender Teilchen, durch Studien in neuen Phasenraumregionen besser verstanden werden. Der Large Hadron Collider (LHC) am CERN ist heute das leistungsstärkste Werkzeug, um starke Wechselwirkungen in Protonen-Protonen-Kollisionen zu untersuchen.

Die in dieser Arbeit vorgestellten Studien konzentrieren sich auf die Messung neuer multidifferenzieller Wirkungsquerschnitte, um Einblick zu erhalten, wie die Dekorrelation in Dijet-Ereignissen durch zusätzliche QCD Strahlung aufgebaut wird. Ziel ist es, den Wirkungsquerschnitt in Abhängigkeit von der Jet-multiplizität (Njet), dem azimutalen Winkel zwischen den beiden leading Jets ($\Delta\phi_{1,2}$) und dem Transversalimpuls des leading p_{T}-Jets (p_{T1}) zu messen. Darüber hinaus wurde auch der Wirkungsquerschnitt als Funktion der Transversalimpulse der vier leading Jets gemessen. Die Messungen wurden mit den Daten durchgeführt, die vom CMS-Experiment am LHC in Protonen-Protonen-Kollisionen bei einer Schwerpunktssenergie von $\sqrt{s} = 13\,\mathrm{TeV}$ im Jahr 2016, entsprechend einer Luminositaet von $36.3 fb^{-1}$, aufgenommen wurden.

Die Messungen wurden mit Berechnungen in führender Ordnung in α_s verglichen, die durch Partonschauer, Multiparton Wechselwirkungen und Hadronisierung ergänzt wurden. Des weiteren werden Rechnungen in höherer Ordnung in α_s mit herkömmlichen Parton-Shower und und mit Parton Branching (PB) transversalimpulsabhängigen (TMD) Partondichten und PB-TMD Parton-Shower mit den Messungen verglichen. Diese Arbeit beschreibt einen der ersten und sehr strengen Tests der PB-TMD-Methode zur Beschreibung multidifferenzieller Jet Messungen, welche zur weiteren Entwicklung des neuartigen und erfolgreichen PB-Ansatz führen können.

Die in dieser Arbeit vorgestellten multidifferenziellen Wirkungsquerschnitte waren sehr aufschlussreich für das Verständnis von Multijet-Strahlung in der Back-to-Back Region und zeigten die Bedeutung von Parton-Shower. Darüber hinaus wurde die Bedeutung von Beiträgen höherer Ordnung in α_s in den theoretischen Vorhersagen für Jets diskutiert, und diese Beobachtungen sind von großem Interesse für die Entwicklungen von komplexen Mulitjet Rechnungen.

CONTENTS

1 Introduction **1**

2 The Standard Model and physics at hadron colliders **3**
2.1 The Standard Model of Particle Physics 4
2.1.1 Quantum Chromodynamics sector 5
2.2 QCD and hadron collider physics 6
2.2.1 Colliding high energetic protons 6
2.2.2 Collinear factorization 7
2.2.3 DGLAP evolution 8
2.2.4 Transverse Momentum Dependent parton densities 11
2.3 Monte Carlo Event Generators 12
2.3.1 Fixed order QCD 13
2.3.2 Parton showers 14
2.3.3 Fixed order matched to parton showers 16
2.3.4 Multiparton interactions 18
2.3.5 Hadronization 22
2.3.6 Jets 24

3 Experimental Setup **27**
3.1 The Large Hadron Collider 27
3.1.1 Overview 28
3.1.2 Luminosity 30
3.2 The Compact Muon Solenoid 32
3.2.1 The coordinate system 32
3.2.2 The solenoid magnet 33
3.2.3 Tracking detectors 34
3.2.4 Calorimeters 36
3.2.4.1 Electromagnetic calorimeter 36

3.2.4.2 Hadron calorimeter . . . 38
3.2.5 Muon system . . . 38
3.2.6 Trigger system . . . 39

4 Event reconstruction and selection **41**
4.1 Particle identification and Particle Flow algorithm . . . 41
4.2 Jet reconstruction and selection . . . 43
4.3 Event selection . . . 45
4.4 Observables and phase space definition . . . 45
4.4.1 Four leading jets p_{T} spectra . . . 46
4.4.2 Multi-differential jet multiplicity . . . 46
4.5 Removing the background . . . 47

5 Analysis at detector level **49**
5.1 Data and simulated samples . . . 50
5.2 Pileup corrections . . . 51
5.2.1 Measurement of pileup in data . . . 52
5.2.2 Pileup simulation . . . 52
5.2.3 Removal of overweighted events in simulation . . . 53
5.2.4 Procedure of pileup reweighting . . . 56
5.3 Jet Energy Corrections . . . 56
5.3.1 Jet Energy Scale . . . 57
5.3.2 Jet Energy Resolution . . . 58
5.4 Trigger strategy . . . 61
5.4.1 Trigger efficiency determination . . . 62
5.5 Additional corrections . . . 65
5.5.1 Missing transverse energy . . . 65
5.5.2 Hot zones correction . . . 69
5.5.3 Prefiring correction . . . 71
5.6 Comparison data to simulation . . . 72

6 Analysis at particle level **75**
6.1 Unfolding procedure . . . 76
6.1.1 Response matrix construction . . . 77
6.1.2 Migrations in the phase space . . . 80
6.1.3 Unfolding tests . . . 83
6.2 Uncertainties after unfolding . . . 89
6.2.1 Statistical uncertainties . . . 89
6.2.2 Systematic uncertainties . . . 89
6.2.3 Correlations . . . 92

7 Theoretical predictions **97**
7.1 Comparison with previous jet measurements . . . 98
7.1.1 Inclusive jets . . . 98
7.1.2 Azimuthal correlations . . . 99
7.1.3 Azimuthal correlations in back-to-back region . . . 100

7.2 The role of Multiparton Interactions . . . 100
7.3 Study of MG5_aMC+CA3 (jj) NLO different contributions . . . 101
7.4 Dijet and three-jet predictions from MG5_aMC+CA3 NLO . . . 103
7.5 Calculations at fixed Next-to-Leading Order . . . 105

8 Results and discussion **107**
8.1 Jet multiplicity distribution . . . 107
8.2 Transverse momenta of the four leading p_{T} jets . . . 110

9 Conclusions **115**

Appendices **119**

A Introduction to MC integration **121**

B Preselection studies **125**

C Background studies on the measured observables **129**

D Pythia and Madgraph samples in slices **133**

E Determination of the maximal weight for each p_T^{rec} bin **137**

F Pileup cleaning effects **141**
F.1 PU cleaning effect on the measured observables in this thesis . . . 141

G MET filters studies **149**

H Probability matrices **153**
H.1 Jet multiplicity measurement $(N_j, \Delta\phi_{1,2}, p_{T1})$ . . . 153
H.2 Four leading jets p_T (p_{Ti}, n_i^{jet}) . . . 153

I Predictions compared to data **157**

J Publications **163**

Bibliography **211**

CHAPTER

1

INTRODUCTION

Since ancient times, we have searched for knowledge to describe the world surrounding us. The first scientists in classical antiquity, called *natural philosophers*, already studied subjects related to physics, astronomy, and mathematics. Notwithstanding, their work was based on observation and analysis, with somehow rudimentary experiments, philosophers like Aristotle (384-322 BC), Anaximenes (585-525 BC), and Heraclitus (535-475 BC) studied how the matter was made. But Leucippus (around 460 BC) and his pupil Democritus (460-370 BC) were the ones that introduced the atomistic conception: all the matter is made of atoms (indivisible particles) and void.

The scientific interest in the world surrounding us and the development of new theories did increase through the history of science. The revolutionary studies of the constituents of protons, the so-called partons, performed at Deutsches-Elecktronen Synchrotron (DESY) experiments state a solid basis into the future. Nowadays, we have big experiments like the experiments at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) that investigate the fundamental building blocks of matter, the formation of the universe, among other of the most fundamental and complex questions ever formulated in the history of science. The establishment of a theory that describes matter, The Standard Model, has been a milestone in the last century's achievements, and contributions from centers like DESY and CERN have made this happen.

From what has been studied at colliders in High Energy Physics (HEP) within the Standard Model, the theory of strong interactions is a fundamental subject. At high energies, in hadron collider experiments, the strong force is the most dominant among the four fundamental forces of nature. The theory that describes this force is called Quantum Chromodynamics (QCD). This theory has allowed us to study the structure of matter at the lowest scales (high energies) ever and understand how quarks and gluons bind together to form hadrons and mesons: particles that form all the known matter in the universe (except for Dark Matter and Dark Energy which structure remains presently unknown). The production of jets, which are reconstructed from a collimated stream of high energetic hadrons, coming

from the fragmentation of high energetic partons, are described by QCD. Jets are the most powerful objects we have to test perturbative QCD calculations. At leading order (LO) in the strong coupling α_{S}, two collinear partons from the protons scatter and produce two high p_{T} partons in the final state. The jets originating from such a process will be strongly correlated in the azimuthal plane, and the azimuthal angle between both, $\Delta\phi_{1,2}$, will be close to π. However, higher-order corrections to the lowest order process will result in a decorrelation in the azimuthal plane, and $\Delta\phi_{1,2}$ will significantly deviate from π. These higher-order corrections can be hard corrections, calculated at the matrix element level at next-to-leading order (NLO), or soft multiple parton radiation treated by parton showers. It is therefore of great interest to measure the multiplicity of such additional jets in different regions of $\Delta\phi_{1,2}$ and to measure the transverse momenta of the additional jets.

The azimuthal correlation in high p_{T} dijet events has been measured previously by the D0 Collaboration in $\mathrm{p\overline{p}}$ collisions at a center-of-mass energy of $\sqrt{s} = 1.96\,\mathrm{TeV}$ [1, 2], in pp collisions by the ATLAS Collaboration at $\sqrt{s} = 7\,\mathrm{TeV}$ [3], and by the CMS Collaboration at $\sqrt{s} = 7$, 8, and $13\,\mathrm{TeV}$ [4–7].

This work presents new measurements of multi-differential dijet event cross sections in pp at a center-of-mass energy of 13 TeV. The data used in these measurements were recorded by the CMS experiment at CERN during 2016 (Run II). The selected events contain a leading jet with transverse momenta $p_{T1} > 200\,\mathrm{GeV}$, and a subleading jet of $p_{T2} > 100\,\mathrm{GeV}$ in the range of pseudorapidity $|y| < 2.5$. The multiplicity of jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ (within $|y| < 2.5$) in addition to the two leading jets is measured in bins of p_{T1} and $\Delta\phi_{1,2}$. The jet multiplicity in bins of $\Delta\phi_{1,2}$ will provide information about the number of jets needed to create the $\Delta\phi_{1,2}$ decorrelation and how higher order corrections in perturbative QCD are important to describe it. In addition, the cross section of the first four leading jets is measured as a function of the p_{T} of each jet, which will give additional information on the nature of the higher-order corrections. The measurements are done considering correlations between p_{T1} and $\Delta\phi_{1,2}$ bins, for the jet multiplicity, and correlations among jets in the measurement of the four leading jets p_{T}, presented in the correlation matrix for each measurement.

In the flowing, the structure of the thesis is described. First of all, in Chapter 2 an overview of the standard model of particle physics is presented and the main aspects of QCD physics used at hadron colliders (in particular proton-proton colliders) are discussed together with the Monte Carlo simulation programs used to describe high energy proton-proton collisions. Later on Chapter 3 the experimental setup used in the measurements is described. In Chapter 4 the details on the selected events and its reconstruction is presented. Afterwards, in Chapter 5 all the corrections applied to the data at detector level are described. These studies are made to understand the detector effects and make sure they are properly simulated by the Monte Carlo samples that include a detector simulation. After understanding the detector level, the next step is to unfold the data to particle level (more specifically to hadron level) and study the impact of the correlations in the measurements. The data unfolding is done to have the measured data in a sort of "universal format" or level that can be compared directly with theoretical predictions. All this analysis is described in Chapter 6. The understanding of the theoretical predictions used to compare with the unfolded data is carried on in Chapter 7. After having understood the theoretical predictions, in Chapter 8 the results are presented. Then finally, in Chapter 9, the summary and conclusions of the work are discussed.

CHAPTER

2

THE STANDARD MODEL AND PHYSICS AT HADRON COLLIDERS

Contents

2.1 The Standard Model of Particle Physics 4
2.1.1 Quantum Chromodynamics sector 5
2.2 QCD and hadron collider physics 6
2.2.1 Colliding high energetic protons 6
2.2.2 Collinear factorization 7
2.2.3 DGLAP evolution 8
2.2.4 Transverse Momentum Dependent parton densities 11
2.3 Monte Carlo Event Generators 12
2.3.1 Fixed order QCD 13
2.3.2 Parton showers 14
2.3.3 Fixed order matched to parton showers 16
2.3.4 Multiparton interactions 18
2.3.5 Hadronization 22
2.3.6 Jets 24

The field of particle physics has given significant contributions to our understanding of the universe in the last century. In that sense knowing how the matter is built and the interactions between its components have been the main subject of interest to particle physicists. Since the first experiments and discoveries on cosmic rays up to the particle accelerators such as HERA, LEP, SLC, Tevatron, and nowadays the LHC, the Standard Model has proved to be accurate in predicting new particles and describing three of the four fundamental interactions: the short-range *weak* ($\sim 10^{-18}$ m) and *strong* ($\sim 10^{-15}$ m) forces and the long (infinite)

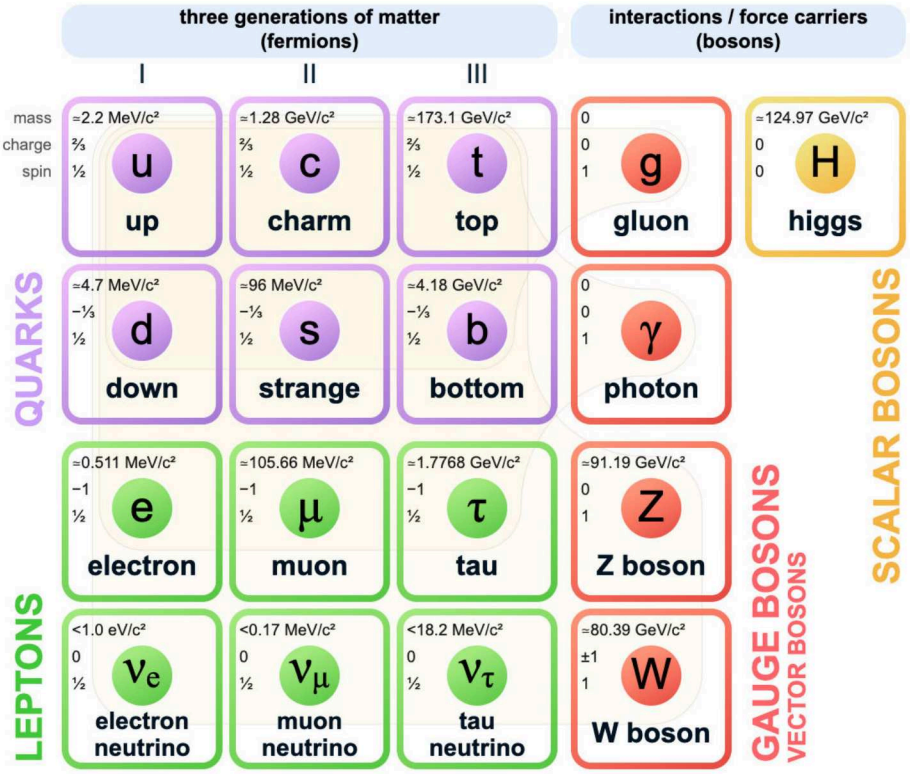


Figure 2.1: Representation of the Standard Model with all particles included and its corresponding mass, spin and electric charge. Figure taken from Ref. [11]

range *electromagnetic* force. Those three forces are respectively about 10^{32}, 10^{38}, and 10^{36} times stronger than gravity (which is the fourth fundamental force of nature). Therefore in subatomic scale measurements at colliders, those three forces mentioned above have been studied with high precision using the SM. In this chapter, the SM model will be introduced together with the QCD sector. The remaining part will focus on physics at hadron colliders and Monte Carlo Event generators. In the end, the jets will be defined as the objects measured in experiments and predicted by QCD theory.

2.1 The Standard Model of Particle Physics

The Standard Model is a Quantum Field Theory based on (gauge) symmetry principles. The SM successfully describes the strong, weak, and electromagnetic interactions between all known elementary particles. Except for gravity, the SM appears to be the theory that explains all interactions in our universe. The fundamental QFT theories that form the standard model are Quantum Chromodynamics (QCD) and the Electroweak Theory (EW). The SM is built by matter particles called fermions, which are elementary particles of spin $\frac{1}{2}$ that obey the Fermi-Dirac statistics [8] and the Pauli exclusion principle [9]. On the other hand, there are bosons that follow the Bose-Einstein (integer spin) statistics [10], being the *gauge bosons* or force carriers (spin 1) plus the scalar Higgs boson (spin 0).

As shown in Fig. 2.1 fermions are conformed by three generations of quarks and leptons, each one with a corresponding antiparticle (same properties but opposite electric charge, not

shown in the figure). All the ordinary matter is made of first-generation fermions (protons and neutrons, made of *up* and *down* quarks form atoms along with the electrons), which form particles (baryons) with a longer lifetime than the other two-generation fermions, which are only visible in very high energy environments. There are six different quark *flavors* in the SM namely: up (u), down (d), charm (c), strange (s), bottom (b), and top(t). As a consequence of the exclusion principle, quarks, e.g., in the proton (two *up* and one *down*), have another degree of freedom called *color charge*. On the other hand there are six leptons: electron (e), muon (μ), tau (τ), and three neutrinos in correspondence with them (ν_e, ν_μ, ν_τ).

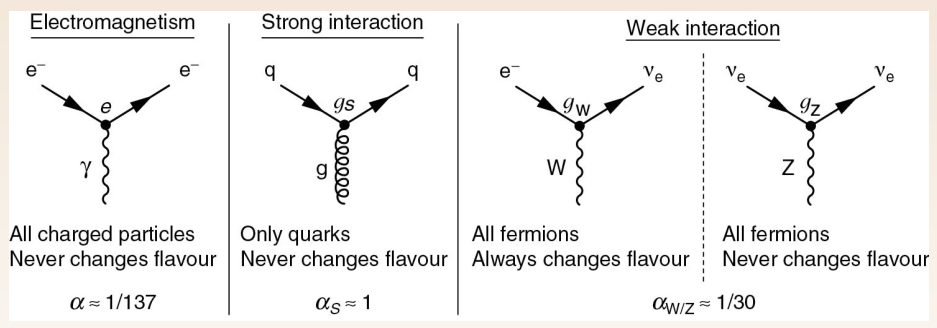


Figure 2.2: SM interaction examples. Only the Higgs weak interactions are not represented here. Taken from Ref. [12].

The forces in the SM are explained by the interchange of spin-one force mediators or gauge bosons as shown in Fig. 2.2. The carriers of the weak force are $W^{\pm}$ and Z bosons; there are eight gluon fields (g) which are the strong force mediators, and the photon (γ), which mediates the electromagnetic interactions, plus the Higgs boson. The Standard Model seems to be completed with QCD, ($SU(3)$ gauge theory) which is the theory of the strong force, and EW ($SU(2) \otimes U(1)$), which explains the electromagnetic and weak interactions. The EW theory predicted *massless* W and Z bosons, but the experimental evidence showed they have non-zero mass. This gave birth to the last piece of the SM, the Higgs mechanism [13] which explains that Z and W bosons acquire their masses from the interaction with the Higgs field, and also all quarks and leptons in consequence. The Higgs boson was discovered at the LHC by CMS [14] and ATLAS [15] experiments in 2012 as predicted by the SM several decades before.

2.1.1 Quantum Chromodynamics sector

The QCD lagrangian based on the symmetry group SU(3) explains the theory of hard interactions. This gauge invariant non-abelian theory (gluons interact with each other) explains the interaction of quarks and gluons (so called partons) that form hadrons and mesons. The lagrangian of QCD can be written as a Dirac lagrangian as [16]:

$$\mathcal{L}_{QCD} = -\frac{1}{4} G^{a\mu\nu} G^{a}_{\mu\nu} + \sum_{q=1\ldots6} \bar{\Psi}_q (i\gamma^\mu - m_q) \Psi_q \tag{2.1}$$

where the gluon field $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_3 f^{abc} G^b_\mu G^c_\nu$ encodes the allowed vertex interactions in QCD as depicted in Fig. 2.3. But this is classical gauge QCD theory, the "real" QCD lagrangian (after quantization) is far more complicated than eq. 2.1 and encodes all the

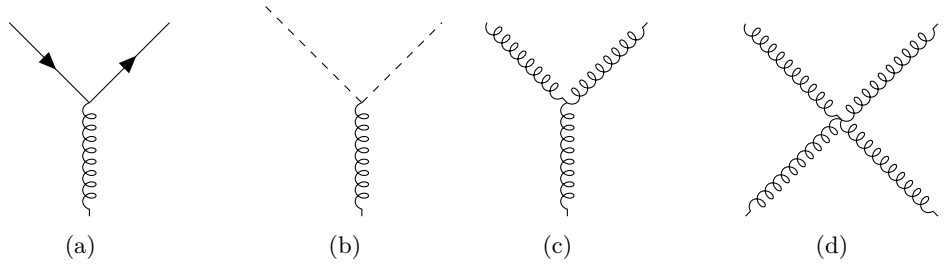


Figure 2.3: Interaction vertices of QCD: (a) basic vertex interaction between quarks and gluons, (b) ghost field coupling to gluon. (c) three-gluon vertex interaction, (d) four-gluon vertex interaction.

possible interaction vertices as shown in Fig. 2.3. In a gauge-invariant non-Abelian theory as QCD, the gauge is fixed using Faddeev-Popov [17,18] ghosts which do not affect the Lorentz invariance. The ghost field can be identified as part of a negative kinetic energy term in the Lagrangian and leads to ghost vertex interaction with gluons. The QCD Lagrangian in eq. 2.1 does not have the gauge fixing, ghost terms, and counterterms from the renormalization of the theory.

2.2 QCD and hadron collider physics

The study of high energy colliding hadrons has led us to test and improve all our knowledge of the Standard Model and improve our simulations to predict and describe the data collected. The first steps in colliders physics were dedicated to studying the internal structure of protons through the "measurement" of the so-called parton distribution functions (PDFs). In this sense, the study of Deep Inelastic Scattering (DIS) at HERA [19] electron (positron) proton collider finally consolidated the knowledge on the QCD parton model which led us later to hadron colliders and the study of proton-proton collisions at the LHC. This section will focus on the physics behind high energy collisions of protons and the evolution at early stages up to the final stage: form interacting partons (quarks and gluons) to hadrons (formed by quarks bounded by gluons) and jet production.

2.2.1 Colliding high energetic protons

As well known, the proton rest mass is about $m_p \sim 1\,\text{GeV}$ and at this scales ($Q \sim 1\,\text{GeV}$) quarks bounded by gluons interact at a time scale $\sim 1/m_p$ which lies in the non-perturbative QCD regime. In this regime, the high value of the (running) strong coupling α_S makes the proton a consistent and long living system ($\sim 10^{34} years$). When going to high energy interactions of protons, with a center-of-mass energy $\sqrt{S} \gg m_p$ the scale of the interaction $Q \gg m_p$ allows hard interactions between the partons inside those protons. Then in a head-on proton-proton collision, like the one depicted in Fig. 2.4, one has access to perturbative and non-perturbative regions. The partons from the protons carry a fraction x_i of the proton momentum P_i such that its momentum is $p_i = x_i P_i$ (with $i = 1, 2$). This information is encoded in the universal but non-perturbative PDFs, $f_{P1,i}(x, \mu^2)$ as shown in Fig. 2.5). These PDFs give the probability to find a parton with a certain fraction of the proton momentum x at certain scale μ^2, and are determined from fits to collision data [21] and could also be determined by lattice QCD calculations [22] which is a powerful tool to make predictions

in the non-perturbative regime. The partons undergo a hard scattering producing a certain number of high-energy partons. The initial and final partons radiate gluons and finally come together into hadrons, leaving signals in the detectors. This process depicted in Fig. 2.4 comes from the collinear factorization theorem, which will be introduced in the following section.

2.2.2 Collinear factorization

The factorization theorem is extensively used in high-energy cross sections calculations. The factorization theorem allows us to separate (factorize) the long-distance from short-distance effects. The long-distance effects, soft emissions, come from the PDFs evolution (see Sec. 2.2.3). Then the short-distance process, hard scattering, is calculated by perturbation theory. Factorization theorem is expected to hold in colorless final state (hadrons, mesons, leptons in final state) hand have been proven in DIS, Drell-Yan, etc. (more details on Ref. [23]). In a proton-proton collision, at high energies, one considers the partons to have a

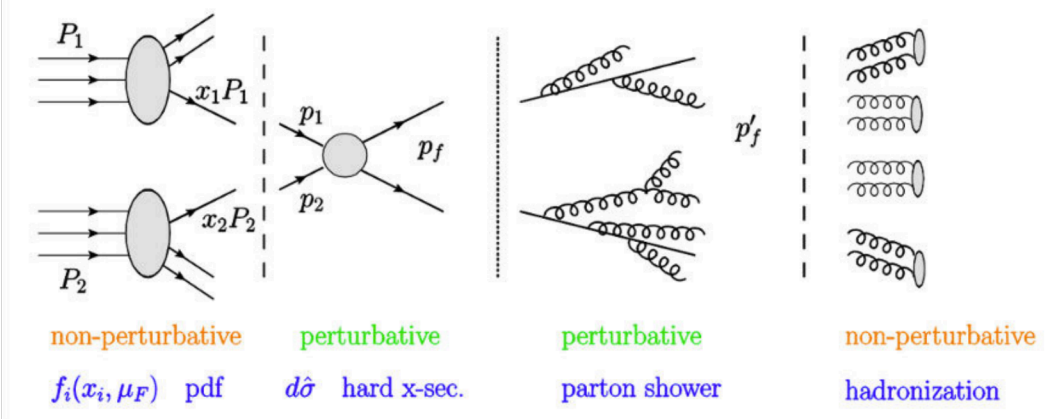


Figure 2.4: Schematic view of a proton proton collision. [20]

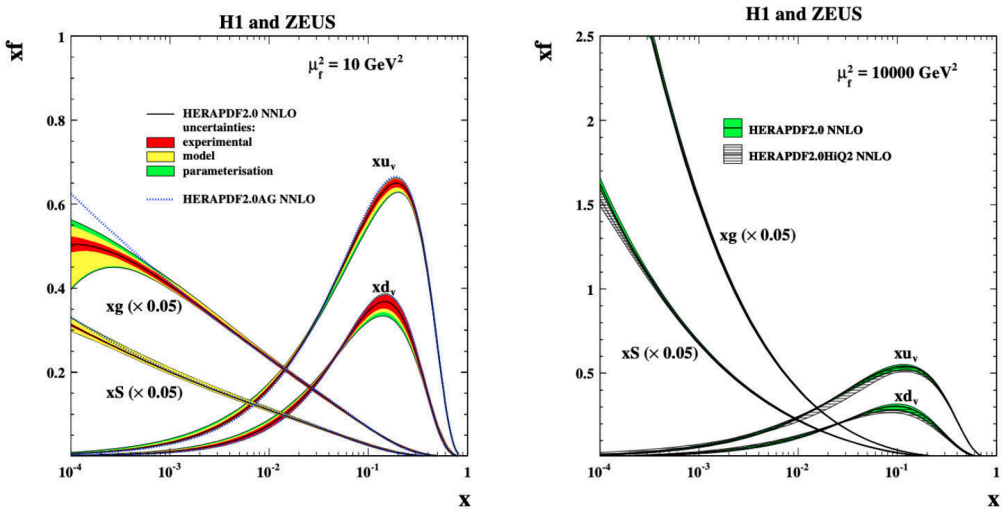


Figure 2.5: Parton distribution functions for proton valence quarks xu_v, xd_v, sea-quark $xS = 2x(\bar{U} + \bar{D})$ and gluon xg distributions. HERAPDF2.0 NNLO at a factorization scale $\mu_f^2 = 10\,\mathrm{GeV}^2$ (left) and at $\mu_f^2 = 10^4\,\mathrm{GeV}^2$ (right). Figure taken from Ref. [21]

longitudinal momentum p_i much larger than its intrinsic transverse momentum k_T, but in the low $x \to 0$ limit is more appropriate to consider $\boldsymbol{k_T}$ factorization (next Sec. 2.2.4). In the case of QCD collinear factorization, the cross section can be written as:

$$\sigma(pp \to X) = \sum_{i,j} \int_0^1 \int_0^1 dx_i dx_j f_{P_{1,i}}(x_i, \mu_f^2) f_{P_{2,j}}(x_j, \mu_f^2) \hat{\sigma}_{ij \to X}(x_i x_j S, \mu_r^2, \mu_f^2) \tag{2.2}$$

where the sum runs over the partons i and j that interact with momentum fractions x_i and x_j (from the proton momentum) and give rise to a final state X at a partonic centre-of-mass energy $\sqrt{x_i x_j S}$, where $\sqrt{S}$ is the proton-proton collision centre-of-mass energy. The partonic cross section $\hat{\sigma}_{ij \to X}$ is calculated perturbatively, introducing the renormalization scale μ_r to absorb UV divergences, this term is usually called Matrix Element (ME) calculation. The factorization scale μ_f is the link between the hard partonic cross section and the PDFs and acts as a separation for divergent contributions (avoids large logarithms in partonic cross sections) into the PDFs.

2.2.3 DGLAP evolution

The PDFs are non-perturbative objects, but only an initial distribution $f_i(x, \mu_0^2)$ at a very small scale $\mu_0 \sim 1\,\text{GeV}$ encodes the non-perturbative effects and its parametrization for every parton (i) have been widely studied by different groups like HERAPDF [24], CTEQ [25] or NNPDF [26] among others. There are different QCD evolution approaches like BFKL or DGLAP as depicted in Fig. 2.6 (right figure) in the $\ln(1/x)$-$\ln Q^2$ plane. In the intermediate region, in Fig. 2.6, is represented an attempt for unification of DGLAP and BFKL evolution, the so-called Ciafaloni-Catani-Fiorani-Marchesini (CCFM) evolution equations [27–29] which was also a motivation for the Parton Branching (PB) evolution presented in the next section (Sect. 2.2.4). This section will describe the DGLAP evolution and how it can be solved via iteration since the PB evolution will use this same solution and interpretation.

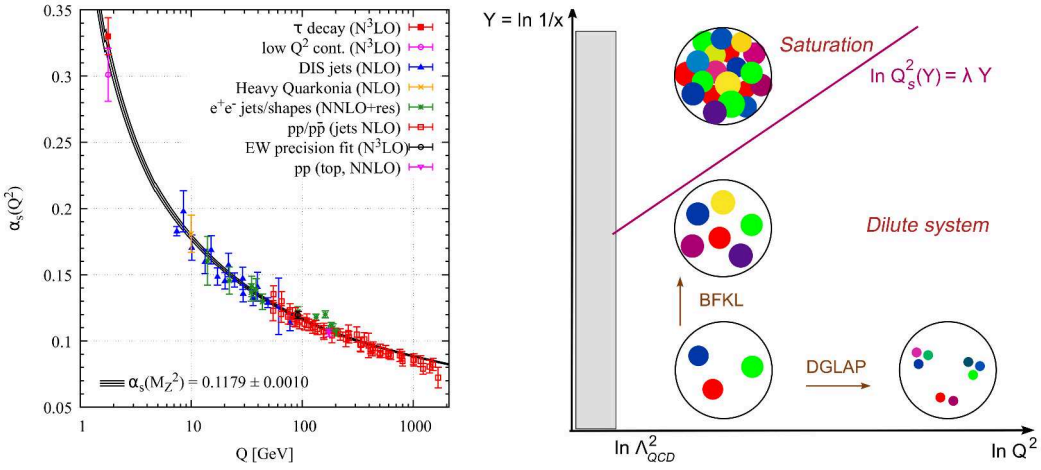


Figure 2.6: (Left) The running of α_s is presented as function of the scale (for reference). (Right) The QCD phase space with the different evolution approaches and their application region [30]

The scale dependence and therefore the evolution to higher scales is calculable in QCD perturbation theory by using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [31–

34] evolution equation:

$$\mu^2 \frac{\partial f_i(x,\mu^2)}{\partial \mu^2} = \frac{\alpha_S(\mu^2)}{2\pi} \sum_j \int_x^1 \frac{dz}{z} P_{ij}(\alpha_s(\mu^2), z) f_j(x/z, \mu^2) \tag{2.3}$$

where $f_i(x,\mu^2)$ are the parton distributions for a given parton i and $P_{ij}(\alpha_S, x)$ are the DGLAP splitting functions, depending on the running coupling α_S and the splitting variable $z = x'/x$: where z is the momentum fraction of the emitted parton and x' is the remaining parton momentum fraction. The splitting functions can be calculated in perturbation theory as:

$$P_{ij}(\alpha_s, x) = P_{ij}^{(0)}(x) + \frac{\alpha_S}{2\pi} P_{ij}^{(1)}(x) + ... + \left(\frac{\alpha_S}{2\pi}\right)^n P_{ij}^{(n)}(x) \tag{2.4}$$

here $n = 0$ corresponds to "Leading Order" (LO), $n = 1$ to "Next to Leading Order" (NLO) and so on, and the expressions for LO and NLO splitting functions can be found in [35]. The evolution equation gives us how the PDFs depend on the scale but still the momentum fraction dependence is encoded in the non-perturbative initial distribution $f_i(x, \mu_0^2)$.

One can solve the DGLAP evolution equation by introducing the regularized splitting functions $\hat{P}_{ij}(x)$ and the so called *Sudakov form factor* defined as:

$$\Delta_i(\mu^2) = \exp\left(-\sum_j \int_{\mu_0^2}^{\mu^2} \frac{d^2q}{q^2} \int_x^{z_M} dz \frac{\alpha_s}{2\pi} \hat{P}_{ij}(z)\right) \tag{2.5}$$

this Sudakov with the cut-off on z_M will give us the no-branching probability from a given scale μ_0 to scale μ and gives us loop (virtual) as well as parton (real) emissions resummation at all orders. The Sudakov is built on the assumption that the virtual and real emissions sum up one (unitarity). Then introducing f/Δ_i in eq. 2.3 one gets:

$$\mu^2 \frac{\partial}{\partial \mu^2} \frac{f_i(x,\mu^2)}{\Delta_i(\mu^2)} = \frac{\alpha_S(\mu^2)}{2\pi} \sum_j \int_x^1 \frac{dz}{z} \frac{P_{ij}(\alpha_s(\mu^2), z)}{\Delta_i(\mu^2)} f_j(x/z, \mu^2) \tag{2.6}$$

and one arrives to the following integral from of the DGLAP equation:

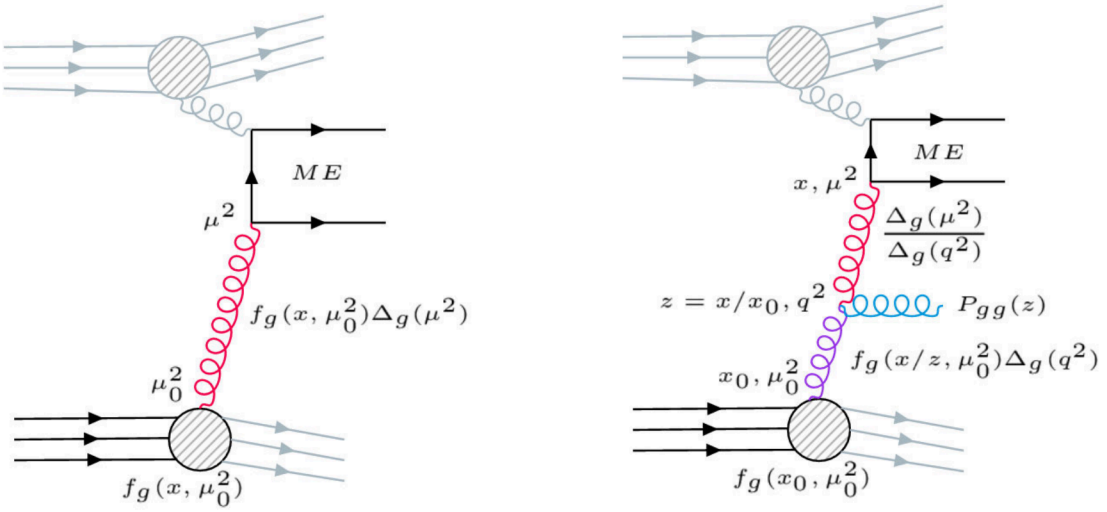
$$f_i(x,\mu^2) = f_i(x,\mu_0^2)\Delta_i(\mu^2) + \sum_j \int_{\mu_0^2}^{\mu^2} \frac{d^2q}{q^2} \frac{\Delta_i(\mu^2)}{\Delta_i(q^2)} \frac{\alpha_S(q^2)}{2\pi} \int \frac{dz}{z} \hat{P}_{ij}(z) f_j(x/z, q^2) \tag{2.7}$$

This way of introduce a resolution scale z_M to separate the possible emissions for $z < z_M$ from the "non-resolvable" or not detectable emissions (branchings) for $z_M < z < 1$ and treat them with the Sudakov from factor is the Parton Branching approach to solve DGLAP evolution. This way the eq. 2.7 have a simple and very physical interpretation, the first term (on the right hand side of eq. 2.7) $f_i(x,\mu_0)\Delta_i(\mu^2)$ gives the no-branching contribution from scale μ_0 to scale μ, and the second encodes the evolution through branchings up to scale μ where the factor $\Delta_s(\mu^2)/\Delta_i(q^2)$ represent the probability of evolving from q to μ without branching. This eq. 2.7 can be solved iteratively [36]:

$$
\begin{aligned}
f_i^{(0)}(x,\mu^2) &= f_i(x,\mu_0^2)\Delta_i(\mu^2) \\
f_i^{(1)}(x,\mu^2) &= f_i^0(x,\mu_0)\Delta_i(\mu^2) + \frac{\alpha_S}{2\pi}\sum_j \int_{\mu_0^2}^{\mu^2} \frac{d^2q}{q^2}\frac{\Delta_i(\mu^2)}{\Delta_i(q^2)} \int \frac{dz}{z}\hat{P}_{ij}(z)\Delta_i(q^2) f_j^{(0)}(x/z,q^2) \\
&= f_i^{(0)}(x,\mu_0^2)\Delta_i(\mu^2) + \log\frac{\mu^2}{\mu_0^2} A_{ij} \circledast \Delta_i(\mu^2) f_j(x/z,\mu_0^2) \\
f_i^2(x,\mu^2) &= f_i^{(1)}(x,\mu^2) + \frac{1}{2}\log^2\left(\frac{\mu^2}{\mu_0^2}\right) A_{ij} \circledast A_{ij} \circledast \Delta_i \mu^2 f(x/z,\mu_0^2) \\
&\vdots \\
f_i(x,\mu^2) &= \lim_{n\to\infty} f_i^{(n)}(x,\mu^2) = \lim_{n\to\infty}\sum_n \frac{1}{n!}\log^n\left(\frac{\mu^2}{\mu_0^2}\right)\sum_j A_{ij}^n \circledast \Delta_i(\mu^2) f_j(x/z,\mu_0^2)
\end{aligned}
\tag{2.8}
$$

here $A_{ij} = \frac{\alpha_S}{2\pi}\int \frac{dz}{z}\hat{P}_{ij}(z)$ is a representation of the integral over the multiple branching variables and $\circledast$ indicates a convolution, such that for A_{ij} (as example):

$$
A_{ij} \circledast \Delta_i(\mu^2) f_j(x/z,\mu_0^2) = \frac{\alpha_S}{2\pi}\int \frac{dz}{z}\hat{P}_{ij}\Delta_i(\mu^2) f_j(x/z,\mu_0^2) \tag{2.9}
$$



(a) Evolution with no detectable emission.

(b) Evolution with one emission.

Figure 2.7: Example of a gluon PDF $f_g(x,\mu^2)$ evolution.

In Fig. 2.7a is depicted the evolution of a gluon PDF up to $f_g^0(x,\mu^2)$, where the PDF evolves from scale μ_0^2 up to the interaction scale μ^2 without branching (no detectable branching). In Fig. 2.7b the gluon PDF evolves from (x_0,μ_0^2) to (x,μ^2), by emitting a gluon with momentum fraction $z = x/x_0$.

2.2.4 Transverse Momentum Dependent parton densities

A more realistic approach is to consider the transverse momentum $\boldsymbol{k_T}$ of the partons inside the proton. This approach is in general called TMD (or k_T) factorization and introduces a more general parton distribution function which depends on $\boldsymbol{k_T}$ called Transverse Momentum Dependent (TMD) PDF, also known as unintegrated PDF (uPDF) (for a general review see Ref. [37]). There are several approaches to calculate the TMDs like CCFM [38] which describe the low x region but uses off-shell matrix elements (considering k_T in the ME calculation). The parton densities from this approach lead to high k_T, which can be larger than the one calculated in the matrix elements and is also limited to gluon-induced processes. This is where the novel Parton Branching (PB) method [39, 40] enters. This approach will be discussed in this section and has been studied to be valid in a wide range of longitudinal momentum fractions (from low to high x, about five orders of magnitude) at different scales in the evolution of the TMD (see a detailed study on Ref. [36]) and is valid for quark and gluon induced processes.

The factorization formula on eq. 2.2 can be generalized to:

$$\sigma(pp \to X) = \\ \sum_{i,j} \int d^2\boldsymbol{k_{Ti}} d^2\boldsymbol{k_{Tj}} \int_0^1 \int_0^1 dx_i dx_j \mathcal{A}_{P_{1,i}}(x_i, \boldsymbol{k_{Ti}}, \mu_f^2) \mathcal{A}_{P_{2,j}}(x_j, \boldsymbol{k_{Tj}}, \mu_f^2) \hat{\sigma}_{ij\to X}(\hat{s}, \mu_r^2, \mu_f^2) \tag{2.10}$$

here $\mathcal{A}_{P_{1,i}}(x_i, \boldsymbol{k_{Ti}}, \mu_f^2)$ is now the TMD PDF and the integration to calculate the cross section is also done over $\boldsymbol{k_T}$ of the incoming patrons. In this approach the TMD is related to the collinear PDF as follows:

$$f_{P_{1,i}}(x_i, \mu_f^2) = \int \mathcal{A}_{P_{1,i}}(x_i, \mu_f^2, \boldsymbol{k_{Ti}}) \frac{d^2\boldsymbol{k_{Ti}}}{\pi} \tag{2.11}$$

The evolution of a TMD $\mathcal{A}_{P_{1,i}}(x_i, \boldsymbol{k_{Ti}}, Q^2)$, in the PB approach, from a scale Q_0 to a scale Q for a parton i obeys the following equation [36]:

$$\mathcal{A}_{P_{1,i}}(x_i, \boldsymbol{k_{Ti}}, Q^2) = \Delta_i(Q^2)\mathcal{A}_{P_{1,i}}(x_i, \boldsymbol{k_T}, Q_0^2) + \sum_j \int \frac{d^2\boldsymbol{q}}{\pi \boldsymbol{q}^2} \frac{\Delta_i(Q^2)}{\Delta_i(\boldsymbol{q}^2)} \Theta(Q^2 - \boldsymbol{q}^2)\Theta(\boldsymbol{q}^2 - Q_0^2) \\ \times \int_x^{z_M} \frac{dz}{z} P_{ij}^R(\alpha_s, z) \mathcal{A}_{P_{1,j}}\left(\frac{x_i}{z}, \boldsymbol{k_{Ti}} + (1-z)\boldsymbol{q}, \boldsymbol{q}^2\right) \tag{2.12}$$

where $z_M < 1$ is the resolution scale which separates *resolvable* from *unresolvable* (involves the emission of an undetectable soft parton) branchings,P_{ij}^R are the resolvable splitting probabilities (similar to DGLAP splitting functions, more details in Ref. [36]) and $\Delta_i(Q^2)$ is the Sudakov from factor:

$$\Delta_i(Q^2) = exp\left(-\sum_j \int_{Q_0^2}^{Q^2} \frac{dq^2}{q^2} \int_0^{z_M} dz\ P_{ji}^R(\alpha_S(q^2), z)\right) \tag{2.13}$$

In addition, in the PB method for the evolution scale is chosen to follow the angular

ordering condition $Q = q_T/(1-z)$, where q_T is the transverse momentum of the emitted parton with respect to the beam axis, ensuring quantum coherence of soft radiated partons. Also, in the PB TMD determination in Ref. [39], two different options according to the α_S renormalization scale Q and q_T are chosen in the fit to DIS data.

The eq. 2.12 can be solved iteratively for any number of consecutive branchings up to the scale of the hard interaction similar to the collinear case. In this method, the kinematics are determined in each branching by energy-momentum conservation, not being the case in parton shower implementations on the collinear factorization approach where it gives kinematic shifts in the longitudinal momentum distributions [41]. The PB-TMD method discussed in this section is implemented in the CASCADE3 [42] Monte Carlo (MC) program, and it can be successfully interfaced to on-shell calculations by adding k_T to ME without altering the momentum energy conservation (more details on Ref. [42]). This approach has been applied to Drell-Yan production [43, 44] at NLO, where the PB-TMDs are interfaced to MADGRAPH5_AMC@NLO [45], with very successful results at high and low mass and at different center-of-mass energies, which any other MC approach has not ever achieved without changing the intrinsic k_T depending on the center-of-mass energy. This method has also been successful in $Z + b\bar{b}$ [46] for four and five flavor schemes at the different center of mass energies. It also has been tested for azimuthal correlations of high transverse momentum jets at NLO [47], some related studies will be further discussed (on Chapter 7) given the link they have with the measurements on this thesis. The CASCADE3 MC event generator loads the PB-TMDs from the TMDlib package. The PB-TMDs can be visualized and compared to other TMDs by using the online tool called TMD-plotter (further details on TMDlib and TMD-plotter last versions are available on Ref. [48]).

2.3 Monte Carlo Event Generators

The PB approach, as discussed in the last sections (Sec. 2.2.3-2.2.4), relies on the Sudakov form factor, which is well suited to parton shower Monte Carlo (MC) simulations of QCD processes. The usage of MC techniques is not only related to the evolution of the partons but also to solving complex integrals, for instance, appearing in the hard scattering cross section, in a pretty efficient way compared to other standard numerical methods as shown in Tab. 2.1. In App. A MC techniques and integration methods are discussed. This section will

Uncertainties as function of N	for 1 dim	for d dim
Monte Carlo	$N^{-1/2}$	$N^{-1/2}$
Trapezoid method	N^{-2}	$N^{-2/d}$
Simpson's method	N^{-4}	$N^{-4/d}$

Table 2.1: MC integration efficiency compared to other integration methods.

describe all MC programs' elements for simulating QCD processes from the PDF evolution up to the hard process and further hadronization. Finally, jets are the fundamental QCD objects measured by collision experiments. The following sections will be focused on the discussion of important aspects related to the MC predictions used in this thesis for comparison with

the measured data at hadron level (further characterized on Chapter 7) and at detector level (described on Chapter 5).

2.3.1 Fixed order QCD

The final and most crucial goal of MC event generators and simulations, in general, is to compute cross sections. These cross sections are at the end compared with experimental data (in most cases after unfolding, see Chapter 6). This can be achieved in the perturbative regime where α_S is small (for scale $Q^2 \gg m_p^2$) and one can calculate the (partonic) cross section as:

$$\hat{\sigma}_{ij\to n} = \alpha_S^n(\mu_R^2) \sum_{m=0}^{\infty} \hat{\sigma}^{(m)} \alpha_S^m(\mu_R^2) \tag{2.14}$$

where the renormalization scale (μ_R) deals with the UV divergences which appear in loop integrals and the cross section is expanded at leading order (LO) for $m = 0$, next-to-leading order (NLO) for $m = 1$ and next-to-next-to-leading order (NNLO) for $m = 2$ and so on, then the cross section can be calculated as truncated series at order $\mathcal{O}(\alpha_S^m)$. The dependence on the factorization scale (μ_F) is still encoded in the cross section at all orders for make use the factorization formula to account for the perturbative and non-perturbative effects. The inclusion of the non-perturbative effects (by the factorization formula) allows us to compare with the data extracted from the experiment. The final state in eq. 2.14 is not specified, in general it can be formed by n partons. In perturbative QCD one calculate the fix order cross sections (ME) at a given order (LO, NLO ...) and for a given final number of partons ($2 \to 2, 2 \to 3, \cdots 2 \to n$). The calculation of the cross section can be done by using Feynman diagrams. The LO cross section is:

$$\hat{\sigma}^{LO} = \int d^4\Phi_n \mathcal{B}(\Phi_n) \tag{2.15}$$

where Φ_n represents the n body (partons) phase space and $\mathcal{B}(\Phi_n)$ is the so-called Born cross section. Nevertheless, the scale dependence in α_S and the PDFs (μ_R and μ_F) starts to be compensated by loop correction at NLO. In that sense, NLO cross section are reliable predictions for rates and uncertainties (also NNLO and so on). In addition it opens new partonic channels from real emissions and brings a better description of final state by the inclusion of extra radiation (real correction). The NLO cross section can be written as:

$$\hat{\sigma}^{NLO} = \int d^4\Phi_n \mathcal{B} + \int d^4\Phi_n \mathcal{V} + \int d^4\Phi_{n+1} \mathcal{R} \tag{2.16}$$

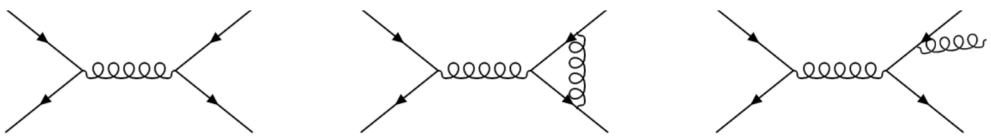


Figure 2.8: From left to right examples of Born, Virtual and Real (emission) corrections at NLO for dijet production in $q\bar{q} \to q\bar{q}$ channel are shown.

At NLO, the real ($\mathcal{R}$) and virtual ($\mathcal{V}$) corrections are not separately finite (because of soft

and collinear divergences), but their sum is (KLN theorem). In Fig. 2.8 examples of Real (emission) and Virtual corrections together with Born cross section diagrams are shown. The divergences need to be subtracted before performing the numerical MC integration. For the cancellation of divergences, one first should perform dimensional regularization to identify the divergences as poles $\sim \frac{1}{\epsilon}$ in the regularization parameter ϵ. The main and more efficient method for eliminating the poles is the subtraction method introduced in App. A. This method has two main variants: dipole subtraction [49] and FKS subtraction [50] automated in MadGraph5_aMC@NLO. The main idea is to cancel the poles from $d-$dimensional integration for the virtual correction in the $n-$body phase space, and the poles from integrating in $4-$dimension the real (emission) in the $n+1-$body phase space by adding and subtracting the so-called subtraction terms (represented as $\mathcal{C}$):

$$\hat{\sigma}^{NLO} = \int d^4\Phi_n \mathcal{B} + \int d^4\Phi_n \left(\mathcal{V} - \int d^d\Phi_1 \mathcal{C} \right)_{\epsilon \to 0} + \int d^4\Phi_{n+1} (\mathcal{R} - \mathcal{C}) \tag{2.17}$$

The subtraction terms do not change the final result but make the real and virtual correction finite separately. Now we can integrate numerically in $d = 4$ the terms in brackets in eq. 2.17 independently. The subtraction terms $\mathcal{C}$ should be chosen such that it matches the singular (divergent) behavior of $\mathcal{R}$ and they can be integrated numerically in a convenient way. The $\mathcal{C}$ terms should also be integrable exactly in d dimension leading to the soft and collinear poles in the dimensional regulator. They should be process independent (overall factor times $\mathcal{B}$).

2.3.2 Parton showers

The formulation of parton branching in terms of the Sudakov form factor is the basis of the parton shower (PS) Monte Carlo programs for simulating QCD jets (more in Sec. 2.3.6). In this section, we will discuss the basics of PS algorithms, starting from the most straightforward and intuitive PB evolution implementation and from there introducing the final-state (timelike) and initial-state (spacelike) showers, also called final-state radiation (FSR) and initial-state radiation (ISR).

Parton branching and spacelike showers

The evolution from a given scale and momentum fraction (μ_k^2, x_k) up to the subsequent (μ_{k+1}^2, x_{k+1}) is done using MC techniques. The first quantity to be generated is the value of μ_{k+1}^2, and as introduced in Sec. 2.2.3, the evolution from μ_k^2 to μ_{k+1}^2 without (resolvable) branching is given by the ratio between the Sudakov form factors as $\Delta(\mu_{k+1}^2)/\Delta(\mu_k^2)$. This way the second scale (μ_{k+1}^2) can be generated from the initial scale (μ_k^2) by solving:

$$\frac{\Delta_i(\mu_{k+1}^2)}{\Delta_i(\mu_k^2)} = \mathcal{R}_1 \tag{2.18}$$

where $\mathcal{R}_1$ is a random number in $[0, 1]$. Notice than if the scale μ_{k+1}^2 is higher than the hard process scale Q^2 the process stops, and no further branchings are generated. The splinting variables z are generated from the (resolvable) splitting functions $P_{ji}^{(R)}$ as:

$$\int_{z_{min}}^{z_{k+1}} dz' P_{ji}^{(R)}(z', \alpha_S(\mu_{k+1})) = \mathcal{R}_2 \int_{z_{min}}^{z_M} dz' P_{ji}^{(R)}(z', \alpha_s(\mu_{k+1})) \tag{2.19}$$

where z_{min} is the lowest kinematically allowed value and z_M is the resolution parameter introduced in the PB method. For generating the azimuthal angle of the emissions (not discussed here), one should consider the coherence effects when using p_T or angular ordered emissions (for a detailed discussion, see Ref. [35]).

Timelike showers

This previous example corresponds to the spacelike forward evolution since the partons increase their virtuality (mass scale μ^2) up to the hard scale, also defining their momentum fractions by the successive application of the algorithm described above.

In the timelike case, used in FSR implementations, the parton evolves downwards from the hard scattering to the hadronization scale $\mu_h^2 = \mu_0^2 \sim 1\,\mathrm{GeV}^2$ and the process of generating the branchings is very similar. In this case, the probability of evolving downwards with no (resolvable) branching from scale μ_k^2 to scale μ_{k+1}^2 is:

$$\frac{\Delta_i(\mu_k^2)}{\Delta_i(\mu_{k+1}^2)} = \mathcal{R}_1 \tag{2.20}$$

notice that here the process scales are is actually decreasing ($\mu_k > \mu_{k+1}$) if one compares with the previous case on eq. 2.18. In this case, eq. 2.20 has no solution for $\mu_{k+1}^2 > \mu_0^2$ since $\Delta(\mu_0^2) = 1$ (from the definition of the Sudakov in eq. 2.5) and the evolution is stopped at μ_0. This way one generates the branchings going to the hadronization, starting from the hard scattering. The procedure for generating the momentum fraction is the same described for the spacelike case in eq. 2.19.

Backward evolution

In practice, the forward evolution is used in FSR when evolving downwards to the hadronization scale. But using forward evolution in the initial state for the implementation of ISR is not efficient at all since the evolution up to the hard scattering scale can lead to configurations of (μ^2, x) in most cases incompatible with the matrix element calculation of the cross section. In this sense, what is used is a *backward evolution* from the hard scale to the PDFs.

This process is not simply to run the forward evolution in reverse as done for the FSR since this will lead to inconsistencies with the parton density functions. Instead, what is done is to consider the parton density for choosing the next value of the evolution scale. The probability of evolving backwards from (μ_n^2, x) to (μ_{n-1}^2, x) with $\mu_n^2 > \mu_{n-1}^2$ is then:

$$\mathcal{S}_{ISR}(\mu_{n-1}^2, \mu_n^2; x) = \frac{f(x, \mu_{n-1}^2)\Delta(\mu_n^2)}{f(x, \mu_n^2)\Delta(\mu_{n-1}^2)} \tag{2.21}$$

This is similar to the forward evolution, but instead of using the Sudakov one uses $\Delta(\mu_i^2)/f(x, \mu_i^2)$, and now the next value in the evolution is calculated from:

$$\mathcal{S}_{ISR}(\mu_{n-1}^2, \mu_n^2; x_n) = \mathcal{R} \tag{2.22}$$

The next step would be to generate the momentum fraction $x_{n-1} = x_n/z$. For this we have

to consider a probability distribution like:

$$\frac{\alpha_S}{2\pi}\frac{P(z)}{z}f(x_n/z,\mu^2_{n-1}) \tag{2.23}$$

and generate the values with a MC method as in eq. 2.19 for the forward evolution. This backward evolution will then evolve from the matrix elements (μ^2_n, x_n) to the proton scale and ensures that the final momentum fraction (from the starting PDF) will be consistently smaller than unity as expected.

This "simple" MC method (based on the concepts introduced in App. A) helps us to obtain a solution to the evolution equations. These methods are the core of the evolution and the PS QCD jets simulation used in the CASCADE3 MC program (for more details see Ref. [36, 42]).

Further discussions on ISR and FSR with more emphasis on coherence effects and ordering conditions can be found in Ref. [51, 52] where the two main approaches from PYTHIA8 and HERWIG++ are discussed. The differences are mainly on the treatment of the coherence effects by using p_T ordering in PYTHIA8 (ensuring angular ordering in the FSR by a veto algorithm) and in HERWIG++ by using angular ordering.

2.3.3 Fixed order matched to parton showers

Parton shower MC matched with fixed LO have been widely used in MC programs like PYTHIA and HERWIG with quite nice predictions and overall shape description of many observables. The LO+PS cross section can be written as:

$$\frac{d\sigma_{\text{"}LO+PS\text{"}}}{d\mathcal{O}} = \mathcal{B}d\Phi_n I^n_{MC}(\mathcal{O}) \tag{2.24}$$

where I_{MC} represents the shower operator. Here because of the unitarity of the shower, when one integrates over $\mathcal{O}$, one gets back the Born cross section (LO cross section). Still, while in the non-perturbative region, this LO plus PS predictions are (in most cases) quite reliable and simulate events at the hadron level, they do not bring us a reasonable estimate of uncertainties given the big scale-dependence at LO. This situation is cured when going to NLO (and further) accuracy. Then NLO matched to PS can achieve better predictions with more reliable uncertainties from the scale variations (renormalization and factorization scales). This way, we get the accuracy of fixed NLO in the perturbative regime and the shower MC prediction in the non-perturbative regime. Nowadays, this is possible with packages like MADGRAPH5_AMC@NLO, which deals with technical issues like avoiding double-counting between the real and virtual emissions and the PS. This has been implemented in other approaches, but here we describe what will be used and studied for the MC predictions to compare with the final results (see Ch. 7).

MC@NLO matching to PS

A first (naive) attempt to match NLO to PS could be to assume that we can generate events separately from Born virtual and real emissions and we pass them to a parton shower such that the differential cross section for a given observable $\mathcal{O}$ would be:

$$\frac{d\sigma_{\text{"}NLO+PS\text{"}}}{d\mathcal{O}} = [\mathcal{B}+\mathcal{V}]d\Phi_n I^n_{MC}(\mathcal{O}) + d\Phi_{n+1}\mathcal{R}I^{n+1}_{MC}(\mathcal{O}) \tag{2.25}$$

where I^n_{MC} is the PS, and n represents the final partons from NLO, which are then passed through the PS. This first attempt is not NLO, and this can be seen if we expand the shower operator at order α_S (which gives one or zero emissions):

$$I_{MC} = \Delta_i(Q^2) + \Delta_i(Q) d\Phi_1 \frac{\alpha_S(Q^2)}{2\pi} P_{ij} \tag{2.26}$$

the Sudakov at NLO would be:

$$\Delta_i(Q^2) = \exp\left[-\int_{Q_0^2}^{Q^2} d\Phi_1 \frac{\alpha_S(q^2)}{2\pi} P_{ij} \right] \approx 1 - \int d\Phi_1 \frac{\alpha_S(Q^2)}{2\pi} P_{ij} \tag{2.27}$$

and using this in eq. 2.26 we finally get the PS at order α_S:

$$I_{MC} \approx 1 - \int d\Phi_1 \frac{\alpha_S(Q^2)}{2\pi} P_{ij} + d\Phi_1 \frac{\alpha_S(Q^2)}{2\pi} P_{ij} \tag{2.28}$$

Then using eq. 2.28 in eq. 2.26 one obtains the following expression:

$$\frac{d\sigma_{\text{"}NLO+PS\text{"}}}{d\mathcal{O}} = [\mathcal{B}+\mathcal{V}] d\Phi_n + d\Phi_{n+1}\mathcal{R} - \mathcal{B} d\Phi_n \int d\Phi_1 \frac{\alpha_S(Q^2)}{2\pi} P_{ij} + \mathcal{B} d\Phi_n d\Phi_1 \frac{\alpha_S(Q^2)}{2\pi} P_{ij} \tag{2.29}$$

If we compare this expression (eq. 2.29) with eq. 2.16 it is clear that we do not recover the NLO cross section, the last two terms in eq. 2.29 come from a wrong matching. In this case, double counting is not avoided, and we do not recover the NLO cross section when expanding at order α_S. This issue is fixed in the $MC@NLO$ formalism where double counting is cured by the so-called Monte Carlo counterterms (also known as subtraction terms) defined as:

$$C_{MC} = \left| \frac{\partial \Phi_1^{MC}}{\partial \Phi_1} \right| \frac{\alpha_S(Q^2)}{2\pi} P_{ij} \tag{2.30}$$

now the Sudakov form factor is defined according to C_{MC} as:

$$\Delta_i(Q^2) = \exp\left(-\int d\Phi_1 C_{MC} \right) \tag{2.31}$$

Then the MC@NLO cross section is defined as:

$$\frac{d\sigma_{\text{"}MC@NLO\text{"}}}{d\mathcal{O}} = \left[\mathcal{B} + \mathcal{V} + \int d\Phi_1 C_{MC} \right] d\Phi_n I^n_{MC}(\mathcal{O}) + (\mathcal{R} - C_{MC}) d\Phi_{n+1} I^{n+1}_{MC}(\mathcal{O}) \tag{2.32}$$

This way, if we expand the PS operator up to order α_S we get:

$$I_{MC} \approx 1 - \int d\Phi_1 C_{MC} + d\Phi_1 C_{MC} \tag{2.33}$$

and using this on eq. 2.32 we can recover the NLO expression:

$$\frac{d\sigma_{\text{"}MC@NLO\text{"}}}{d\mathcal{O}} = [\mathcal{B} + \mathcal{V}] d\Phi_n + d\Phi_{n+1}\mathcal{R} \tag{2.34}$$

Since we cannot integrate the counterterms analytically to extract the poles, in practice

one uses the subtraction technique as in eq. 2.17 which leads to:

$$\begin{aligned}\frac{d\sigma_{"MC@NLO"}}{d\mathcal{O}} = &\left[\mathcal{B} + \left(\mathcal{V} + \int d\Phi_1 C\right) + \int d\Phi_1 \left(C_{MC} - C\right)\right] d\Phi_n I^n_{MC}(\mathcal{O}) \\ &+ (\mathcal{R} - C_{MC}) d\Phi_{n+1} I^{n+1}_{MC}(\mathcal{O})\end{aligned} \tag{2.35}$$

The C_{MC} counterterms not only avoid double-counting but also reproduce the singular behavior of the real emission in the ME. They also ensure a "smooth" matching since NLO+PS has the same shape as the shower in the soft and collinear region, and in the hard region, it approaches NLO. Nevertheless, this method is PS-dependent since, for each PS, we need its own MC counterterms. In addition, it is essential to state that the results from this approach are only physical after showering the events so that the process can be treated in two steps: first, calculate the NLO with the subtraction (which is saved in LHE [53] format), and then shower those events. More detailed discussions about MadGraph5_aMC@NLO method to interface NLO to PS can be found on Ref. [54].

The subtraction terms are available for shower MC herwig and pythia. In this thesis, we will also use the MadGraph5_aMC@NLO method with PB TMDs from Cascade3 by using herwig6 subtraction terms given that the herwig6 shower delivers angular ordered emission (this could also be done with other herwig versions) as in Cascade3 (more details and discussion can be found in Ref. [42]) this approach has been used and justified already in the referenced publications (see Sec. 2.2.4) of PB TMDs interfaced with NLO in the MadGraph5_aMC@NLO framework.

2.3.4 Multiparton interactions

In proton-proton collisions the extra activity not related to the hard scattering and the shower (ISR and FSR) evolution is known as underlying event(UE). This extra activity is greater than the co-called minimum-bias events (collisions without an identifiable hard process). The UE activity is related to the collisions between the partons on the proton that do not participate in the hard interaction. The QCD $2 \to 2$ interaction cross section, having into account all gluons and quarks interaction channels is $\sigma_{int}(p_{Tmin}) \propto 1/p^2_{Tmin}$ where p_{Tmin} is considered since $p_{Tmin} \to 0$ gives a divergent σ_{int}. The falling spectrum of σ_{int} ($\sim 1/p^2_{\mathrm{T}}$) is shown in Fig. 2.9 where one can notice that the total cross section (σ_{total}) is smaller than the interaction cross section already for $p_{Tmin} < 5\,\mathrm{GeV}$. This happens by the fact that the interaction cross section is inclusive. If an event (pp collision) has two partonic interactions is counted twice in σ_{int} but only once in σ_{total} (and so on for a higher number of partonic interactions). Therefore one can introduce the multiparton interaction (MPI) interpretation considering that more than one partonic interaction happens in the same pp collision, then the total cross section would be:

$$\sigma_{total} = \frac{\sigma_{int}(p_{Tmin})}{\langle n \rangle (p_{Tmin})} \tag{2.36}$$

where $\langle n \rangle (p_{Tmin})$ is the average MPI per event above p_{Tmin}.

An interesting approach is to consider double parton scattering (DPS) [56] where two separate hard (partonic) interactions occur. Considering the factorization formula with double parton distribution functions and the corresponding two partonic cross sections one can get [57]:

$$\sigma_{DPS} = \frac{m}{2}\frac{\sigma_A \sigma_B}{\sigma_{eff}} \tag{2.37}$$

where σ_A and σ_B are the individual interaction cross sections, m is a symmetry factor (considered $m = 1$ if process A and B are identical and $m = 2$ for different processes) and σ_{eff} is the so-called "effective cross section" [56, 57] which can be interpreted as an effective interaction area and considered constant with respect to the center-of-mass energy of the collision. If one considers a large separation $y \gg 1/Q$ between the two partonic interactions then $\sigma_{total} \approx \sigma_{DPS}$, otherwise:

$$\sigma_{total} = \sigma_{DPS} + \sigma_{SPS} - \sigma_{sub} \tag{2.38}$$

where σ_{SPS} is the single parton scattering (one scattering per pp collision) cross section and σ_{sub} is a subtraction term (see Ref. [58] for more details) for avoiding double counting between DPS and SPS. A possible DPS is depicted in Fig. 2.10 with also the corresponding color reconnection, which needs to be taken into account for correct simulation of colorless particles (hadrons and mesons) at the end when going to the hadronization in the non-perturbative regime.

In this thesis, two MPI models from PYTHIA and HERWIG are used (for the corresponding PYTHIA and HERWIG prediction). The impact of MPI has been studied with PYTHIA8 (interfaced with NLO by using MADGRAPH5_aMC@NLO) in Chapter 7 and is found to be not relevant for the measured cross sections. But in the following, the PYTHIA and HERWIG MPI models will be described mainly based on Ref. [59] and Ref. [60] respectively.

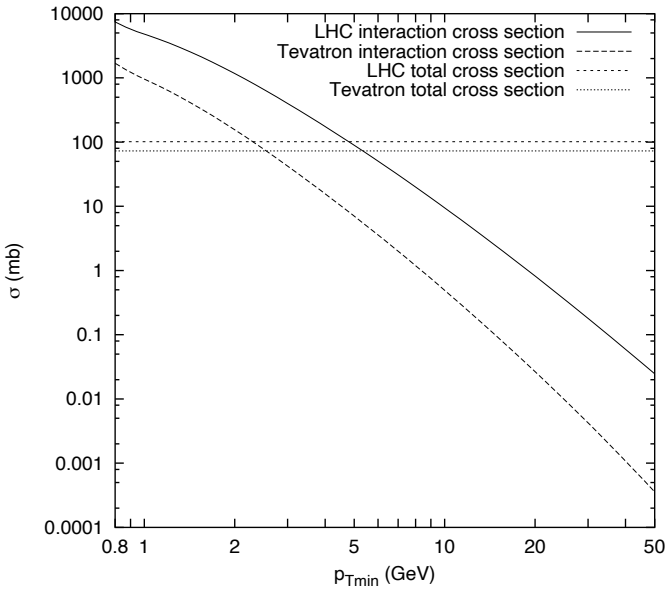


Figure 2.9: The integrated interaction cross section σ_{int} above p_{Tmin} as predicted for Tevatron ($p\bar{p}$ collisions) at 1.8 TeV and the LHC (pp collisions) at 14 TeV compared to the respective total cross section (horizontal lines). Figure taken from Ref. [55].

Pythia MPI model

The MPI model in PYTHIA8 is based on the interleaved evolution for MPI with its own ISR and FSR evolution. The master formula which describes this model is the following:

$$\begin{aligned}\frac{d\mathcal{P}}{dp_{\mathrm{T}}} = &\left(\frac{d\mathcal{P}_{MPI}}{dp_{\mathrm{T}}} + \sum \frac{d\mathcal{P}_{ISR}}{dp_{\mathrm{T}}} + \sum \frac{d\mathcal{P}_{FSR}}{dp_{\mathrm{T}}}\right) \\ &\times \exp\left[-\int_{p_{\mathrm{T}}}^{p_T^{max}} \left(\frac{d\mathcal{P}_{MPI}}{dp'_{\mathrm{T}}} + \sum \frac{d\mathcal{P}_{ISR}}{dp'_{\mathrm{T}}} + \sum \frac{d\mathcal{P}_{FSR}}{dp'_{\mathrm{T}}}\right) dp'_{\mathrm{T}}\right]\end{aligned} \tag{2.39}$$

This gives the probability of an MPI to happen at a given scale (p_T) and the proper evolution of the partons through the ISR and FSR.

The MPIs are ordered in p_T using a sudakov form factor as:

$$\frac{d\mathcal{P}_{MPI}}{dp_{\mathrm{T}}} = \frac{1}{\sigma_{nd}}\frac{d\sigma}{dp_{\mathrm{T}}} \exp - \left(\int_{p_T^i}^{p_T^{i-1}} \frac{1}{\sigma_{nd}}\frac{d\sigma}{dp'_{\mathrm{T}}} dp'_{\mathrm{T}}\right) \tag{2.40}$$

since this approach is not intended to cover elastic or diffractive physics the σ_{int} is distributed among the non-diffractive inelastic cross section σ_{nd} ($\sigma_{nd} \approx 0.6\sigma_{total}$ [61]) .

The cross section dependent of p_T diverges in the limit when $p_{\mathrm{T}} \to 0$ and this is usually treated with a cut-off parameter p_{T0} [59]. This allows to define the partonic cross section as:

$$\frac{d\sigma}{dp_{\mathrm{T}}^2} \propto \frac{\alpha_S^2(p_{\mathrm{T}}^2)}{p_{\mathrm{T}}^4} \to \frac{\alpha_S^2(p_{T0}^2 + p_{\mathrm{T}}^2)}{(p_{T0}^2 + p_{\mathrm{T}}^2)^2} \tag{2.41}$$

This regularization parameter, introduced to make the cross section finite is energy dependent on the PYTHIA8 model. It scales in a similar manner to the total cross section:

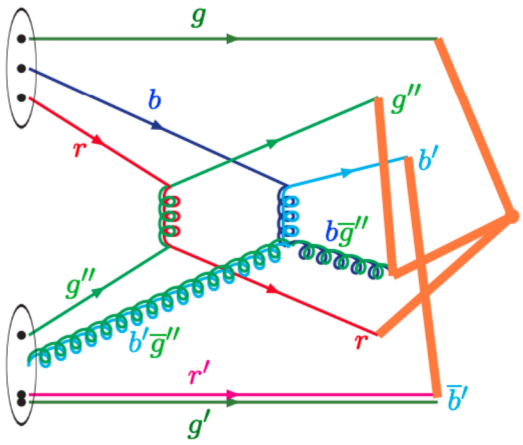


Figure 2.10: Example of a DPS with a possible color flow indicated. Thick orange lines indicate possible colorless particles. Quarks are represented with one color and gluons with two colors. Figure taken from Ref. [59].

$$p_{T0}(E_{CM}) = p_T^{ref} \times \left(\frac{E_{CM}}{E_{CM}^{ref}}\right)^{E_{CM}^{pow}} \tag{2.42}$$

where E_{CM}^{pow}, p_{T0}^{ref} are parameters fitted from data at different energies and E_{CM}^{ref} is a reference center-of-mass energy taken as starting point. Another important dependence is the impact parameter b of the collision, given that events with more central collisions ($b \to 0$) have a bigger UE activity and consequently more MPI occur. This is made by using an overlapping function $O(b)$ between the colliding protons (considered spherically symmetric) depending on the impact parameter. For instance in CUETMP8M1 [62] (PYTHIA8 UE tune used in this thesis) UE tune an exponential function is used:

$$O(b) = \exp\left(-b^a\right) \tag{2.43}$$

here for instance $a = 2$ would be a Gaussian profile.

The last ingredient in this MPI model is the color reconnection (CR) model. This ensures the correct color flow and recombination of the partons to form colorless particles in the non-perturbative regime. The CR model reconnect the different MPI subsystems at certain p_{T}^2 with a probability:

$$\mathcal{P}(p_{\mathrm{T}}) = \frac{\left(R_{rec}p_{T0}\right)^2}{\left(R_{rec}p_{T0}\right)^2 + p_{\mathrm{T}}^2} \tag{2.44}$$

here p_{T0} is the standard cut-off parameter of the MPI cross section, and R_{rec} (reconnection range) is a free parameter fitted to data.

Herwig MPI model

The HERWIG model starts from the same idea of considering independent partonic interactions were the average MPI satisfies also eq. 2.36 and depends on the overlapping between the interacting protons resulting:

$$\langle n \rangle \left(b, p_T^{min}\right) = A(b, \mu^2)\sigma_{int}(p_T^{min}) \tag{2.45}$$

where $A(b, \mu^2)$ is the overlap function with μ^2 characterizing the inverse proton radius and p_T^{min} been the minimum transverse momentum of the additional hard scattering, this model relies on calculating the additional hard scatterings according to a poissonian probability distribution with $\langle n \rangle$ calculated as in eq. 2.45 (more details can be found on Ref. [60]). The CR model in HERWIG is based on the idea that any color-anticolor pair should end up closely in the phase space. Then the color length is defined as:

$$\lambda = \sum_{pairs\ ij} m_{ij}^2 \tag{2.46}$$

here the closeness is defined by pairs having a small invariant mass or small cluster mass (given the HERWIG hadronization model). Then the color length is minimized such that the color neutral cluster (ij) is formed.

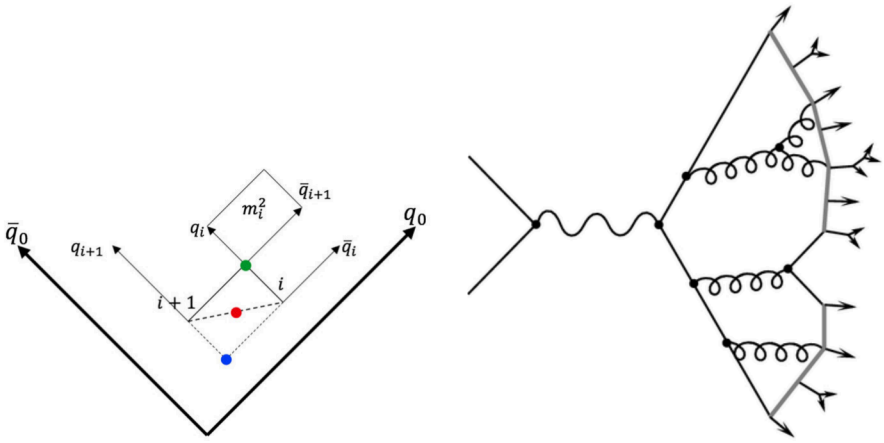
2.3.5 Hadronization

At the very end of ISR and FSR evolution, the virtuality of the partons goes down to the cut-off scale μ_0. At this point, we enter the non-perturbative regime at low momentum transfer. At that stage, color interaction (strong force) binds partons into hadrons; this process is called hadronization. Unstable hadrons will decay into further stable (hadrons and mesons) particles according to their decay width and available decay channels.

We believe that color confinement, given the rising of the strong coupling at low energies, doesn't allow us to find partons outside hadrons, and so far, we have no evidence of that happening. The hadronization process is not currently make-able with today's non-perturbative calculation techniques. Therefore we rely on models based on general features of QCD learned from measurements and lattice calculations [63]. The following two main hadronization models from PYTHIA and HERWIG QCD event generators will be discussed.

String Model

The (Lund) String Model (details on Ref. [64]) is implemented in PYTHIA. This model is based on the confinement idea and in the fact that in QCD (lattice calculations [63]) for large color charge separations the attraction potential is linear $V(r) = -kr$ creating a constant strength similar to a string tension $F(r) = k \approx 1\,\mathrm{GeV}/fm$. The string breaks into hadron-size pieces through "spontaneous" $q\bar{q}$ pairs. This fragmentation process in the string approach is depicted in Fig. 2.11. The $q\bar{q}$ pair creation, or string break, is assumed to follow a Gaussian



(a) (b)

Figure 2.11: Figure (a) represents hadron formation from the string model in a $q\bar{q}$ system [65] where blue, red and green dots represents "early", "middle" and "late" definitions of hadron production points. Figure (b) is the string model on an electron positron interaction [66] going to hadrons.

probability function [65]:

$$\mathcal{P}(m^2, p_{\mathrm{T}}^2) = \exp\left(-\frac{\pi m^2}{k}\right) \exp\left(-\frac{\pi p_{\mathrm{T}}^2}{k}\right) \tag{2.47}$$

here p_T is transverse momentum (momentum perpendicular to the direction of the traveling string) "kick" given to the produced quark by the string breaking, and m is the quark's mass. To some extent, this eq. 2.47 gives the probability of a $q\bar{q}$ pair to "tunnel out" into the classical region and become real. The creation of heavy quarks such as charm ($m_c \approx 1.27\,\text{GeV}$) and bottom ($m_b \approx 4.9\,\text{GeV}$) are suppressed since bigger mass gives a smaller probability as eq. 2.47 states. This probabilistic model can tell us what particles to expect at the end of a full fragmentation process depending on the energy of the original $q\bar{q}$ pairs and help us to sample the process using MC methods.

Cluster Model

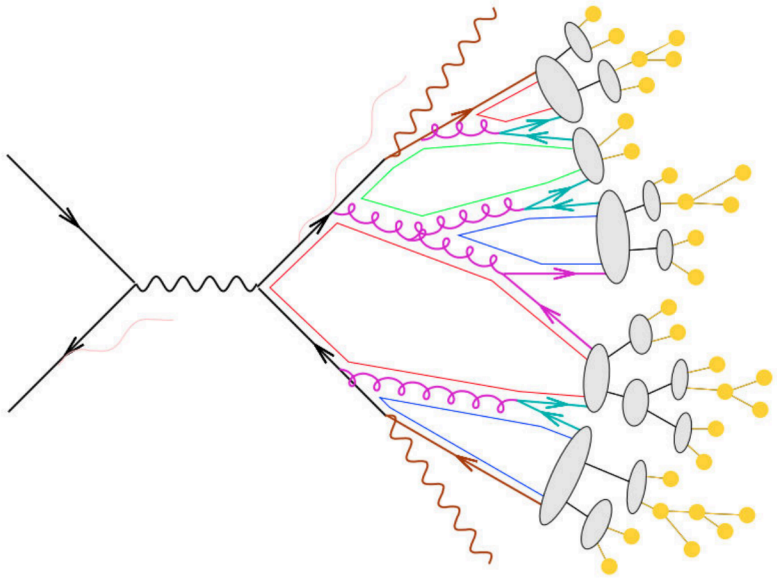


Figure 2.12: Cluster model representation in an electron positron scattering going to hadrons. The grey ovals represent the (pre-confinement) clusters and the yellow circles represent the final hadrons. Also the color flow of quarks and gluons is depicted. Figure taken from Ref. [67]

The Cluster Model is implemented in HERWIG [68] and is based on "pre-confinement" of color [69] which consists of finite mass colorless clusters of quarks and gluons formed when the non-perturbative scale is reached. These then decay into hadrons as shown in Fig. 2.12. The color flow is local in a coherent shower evolution, and this is well suited for the angular ordered shower used in HERWIG. This model also introduces the forced $g \to q\bar{q}$ branchings together with the formation of singlet clusters. The clusters (of mass M) decay to two lighter clusters (of mass m_1 and m_2) when:

$$M^c \geq M_0^c + (m_1 + m_2)^c \tag{2.48}$$

where c and M_0 are parameters of the model which are specific for light and heavy quarks. The cluster decay model described by eq. 2.48 is the so called "fission model" established in HERWIG6 [70] and used in the latest versions. The last step is the lighter clusters from fission decaying isotropicaly into two hadrons according to the corresponding density of states which is proportional to $(2s_1 + 1)(2s_2 + 1)$ where $s_{1,2}$ are the spins of the resulting hadrons.

2.3.6 Jets

The QCD evolution of ISR and FSR creates a high number of final state particles which hadronize and are then measured in the detector. Those particles are detected as well collimated groups of hadrons called jets. These grouped hadrons are called jets since all the hadrons in a hard interaction tend to travel in the same direction as the hard parton from where they were originated. The jets are a natural probe of the existence of gluons and quarks and our most powerful tool to test perturbative QCD predictions since they can be measured at collider experiments.

Jets are defined using a jet clustering algorithm that determines how objects (partons, hadrons, particles, or detector objects) are grouped into jets and specifies how momentum is assigned to a jet. In collider experiments, the jets are built from clusters of energy deposition in the calorimeters and by combining this information with the trajectories detected as tracks in the tracking detectors (more details in Chapter 4). In the following, the anti-k_T clustering algorithm will be introduced and described (as the one used in this thesis measurements) based on Refs. [71, 72].

The anti-k_T jet clustering algorithm relies on the following distances defined in the transverse momentum (k_t) space as follows:

$$\begin{aligned} d_{iB} &= k_{ti}^{-2} \\ d_{ij} &= \min\left(k_{ti}^{-2}, k_{tj}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2} \end{aligned} \tag{2.49}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with y_i and ϕ_i being the rapidity and the azimuthal angle of a given i particle. with transverse momentum k_{ti}. The parameter d_{iB} defines the distance in k_{ti} of particle i with respect to the direction of the interacting protons (the beam pipe at the LHC) and d_{ij} defines the relative separation in transverse momentum and angular space (ΔR_{ij}^2) for a given radius parameter R (in our studies we use $R = 0.4$ which is one of the standards at the LHC in Run-II). The algorithm runs over all pairs of possible particles ij and if $d_{ij} < d_{iB}$ the two particles are recombines into a new particle n such that $\vec{k}_{tn} = \vec{k}_{ti} + \vec{k}_{tj}$, then particles ij are replaced by particle n and if $d_{iB} < d_{ij}$ then the particle i is a jet and its removed from the list and the algorithm continues up to all particles are combined into the jets. Since d_{ij} distance between a hard and a soft particle is determined only by the k_t of the hard particle and their space separation ΔR_{ij} soft particles will tend to cluster to hard one before they cluster among themselves. In addition, it can also be noticed that a hard particle (with no other hard particle within $2R$ distance) will cluster all soft particles within a circle of radius R in the transverse plane resulting in a conical jet. The conical shape of the jets is only affected by hard particles being closer than R to each other where they will cluster. Therefore, all their neighboring soft particles will cluster, giving partial conical jets.

The anti-k_T algorithm is an infrared and collinear safe (IRCS) jet algorithm, where collinear particles recombine early on, and soft particles do not influence the clustering sequence. An IRCS jet clustering algorithm ensures that cross sections can be computed at higher-order in perturbative QCD with no divergences. Another nice feature from the anti-k_T algorithm is that the jet area in the transverse plane ($\phi - y$) plane is not changed by the influence of soft particles and is πR^2 evidencing the independence of the area on the jet p_{T}. In addition, the smearing effects due to the UE and pileup are suppressed by the anti-k_T algorithm and make

UE, pileup, and noise corrections easier. All these excellent theoretical and experimental (practical) features made the anti-k_T the main algorithm used at LHC for measurements in jet physics.

CHAPTER

3

EXPERIMENTAL SETUP

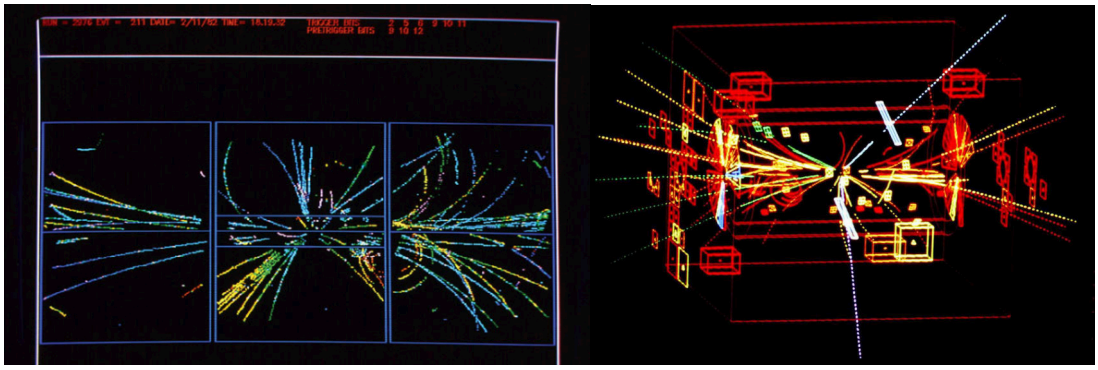
Contents

3.1 The Large Hadron Collider 27
3.1.1 Overview 28
3.1.2 Luminosity 30
3.2 The Compact Muon Solenoid 32
3.2.1 The coordinate system 32
3.2.2 The solenoid magnet 33
3.2.3 Tracking detectors 34
3.2.4 Calorimeters 36
3.2.4.1 Electromagnetic calorimeter 36
3.2.4.2 Hadron calorimeter 38
3.2.5 Muon system 38
3.2.6 Trigger system 39

The experimental setup in collision experiments is, in general, very complex, usually developed by teams or collaboration of scientists. However, it is mainly composed by the particle accelerator and the detector. In this chapter, we will describe the Large Hadron Collider [73], and the Compact Muon Solenoid(CMS) [74] detector, dedicated to collecting data from proton-proton collisions in the context of this thesis.

3.1 The Large Hadron Collider

After the Second World War, the most prominent laboratory and advanced center in the world was about to be born. The talented minds from Europe led by scientists like Niels Bohr or Louis de Broglie were enough for such an enterprise. Louis de Broglie officially put the



(a) W boson discovery (b) Z boson discovery

Figure 3.1: On the left (a), an event in UA1 from the historic run in October to December 1982, corresponding the W bosons discovery. The right image (b), corresponds to and event, by the same experiment at CERN on 30 April 1983, later confirmed to be the first detection of a Z particle decaying to an electron–positron pair (two white tracks). Figures taken from Ref. [78]

idea of creating an advanced European laboratory during a European Cultural Conference in December 1949. Later with the push from UNESCO in Paris in December 1951, the European Council for Nuclear Research (CERN from its french acronym) was established. A few weeks later, eleven countries created the first council and agreed on the creation of CERN. The first stone was there already; CERN gave birth to the most giant machine ever constructed, the Large Hadron Collider(LHC), but not after more than fifty years of scientific research and building engineering.

3.1.1 Overview

Already just ten years after the creation of CERN, in 1959, the Proton Synchrotron (PS) accelerator was started up; it was the first big project from CERN. In the 1970s, CERN built new accelerators, and the PS became the supply of particles to the Intersecting Storage Rings (ISR) [75], and in 1971 the first proton collisions were recorded. These projects gave born to all the knowledge and experience necessary for the ultimate creation of the LHC, but still, a long way of achievements was ahead. In 1976 the Super Proton Synchrotron(SPS) [76] was put in operation, probing protons' inner structure and investigating matter and antimatter, searching for exotic forms of matter. Further achievements like the discovery of W and Z bosons (Fig. 3.1) by colliding protons and antiprotons, from the UA1 (Fig. 3.2) detector, made particle physics an exciting field. The recently discovered particles raised the interest of the scientific community in the study of the electroweak sector, and more precise measurements were carried out with the help of the Large Electron-Positron collider(LEP) [77]. After LEP, the CERN Council planned the construction of the LHC on the same tunnel; this machine would lead the search of the missing particle on the Standard Model, later called the Higgs boson, which was already spotted to be on a specific part of the phase space after the LEP studies. Finally, after the shut down of LEP in November 2000, the construction of the LHC was started.

In Fig. 3.3 the LHC complex is depicted, there we can see two the two main parts of the complex namely the LHC 27 km ring where the collisions happen at CMS, LHCb [81], ATLAS [82] and ALICE [83] detectors, and the pre-acceleration facilities. The protons that

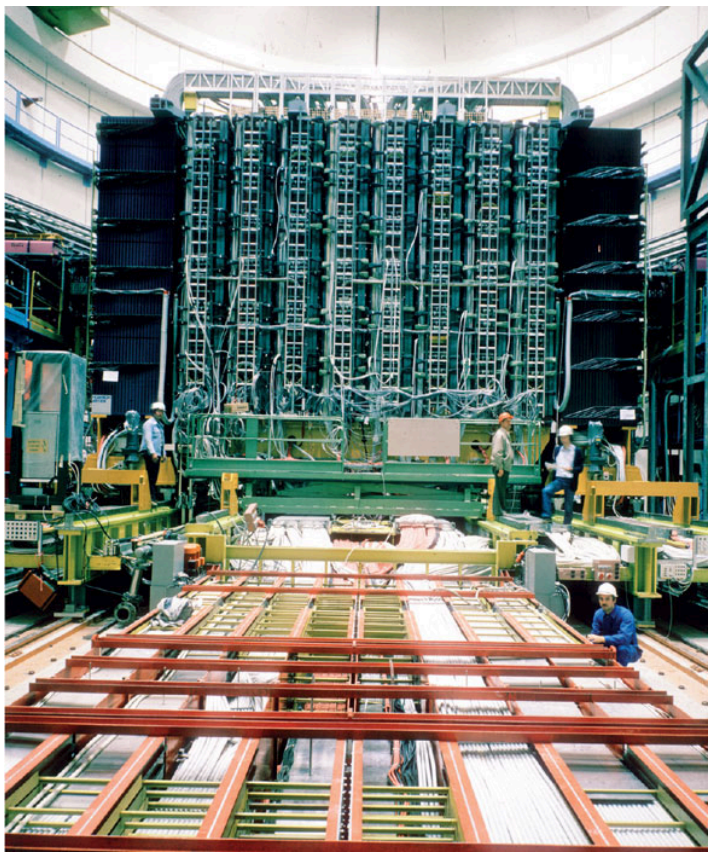


Figure 3.2: The UA1 detector, shown here in its "garage" position, in April 1981. Figure taken from Ref. [79].

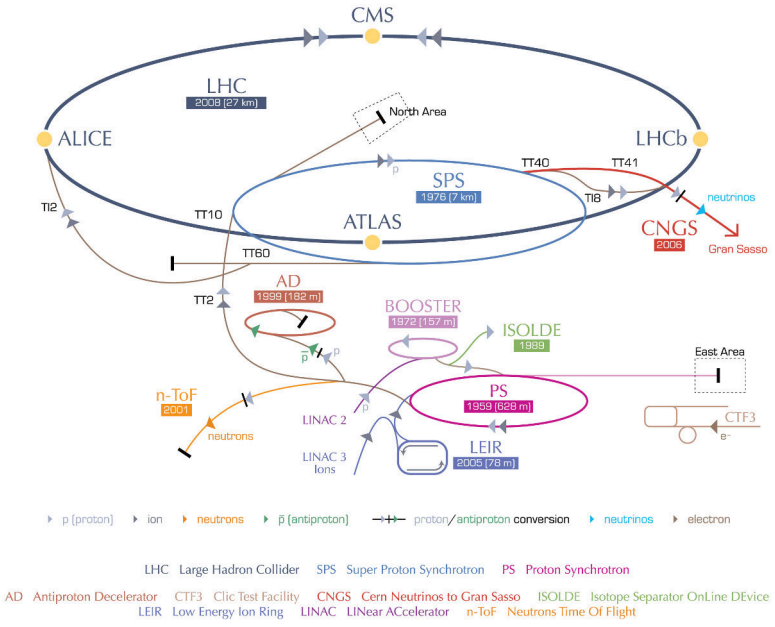


Figure 3.3: CERN accelerator complex(figure taken from Ref. [80])

take part in the collisions are extracted from a hydrogen bottle by using an electric field to remove the electrons from the atom and leave the proton that forms the hydrogen nucleus. These protons are arranged in bunches and then inserted in the pre-acceleration facilities formed by the linear accelerator (LINAC2), the Proton Synchrotron Booster (PSB), the PS, and the SPS, where the protons are accelerated to energies of 50 MeV, 1.4 GeV, 25 GeV and finally 450 GeV when they leave the SPS and enter the LHC ring. The protons from the SPS are inserted into the two pipelines that form the LHC, in which they rotate in opposite directions, the clockwise rotating protons are inserted coming from the SPS through TT60 and T12 lines into ALICE location, and anti-clockwise rotating protons are inserted just before LHCb in the LHC ring through T18 insertion line. Once inside the LHC pipelines, the protons can reach an energy of 6.5 TeV, translating into a center-of-mass energy of 13 TeV during Run-II. In future Run-III, the LHC will try to reach center-of-mass energy of 14TeV, which is the limit by the design of the LHC.

3.1.2 Luminosity

The luminosity is an essential magnitude in high-energy physics collider experiments. This magnitude is defined as the number of interactions per unit time per unit area (expressed in $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$), measuring the ability of a particle accelerator to produce the required number of interactions. The luminosity ($\mathcal{L}$) is the proportionality factor between the number of events per unit time ($\frac{dN}{dt}$) and the cross section of the process (σ_p):

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma_p \tag{3.1}$$

At the LHC, the luminosity is determined from machine parameters, considering two Gaussian beams, the luminosity can be expressed as:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi \Sigma_x \Sigma_y} \tag{3.2}$$

In this well known expression for the luminosity of two Gaussian colliding beams [84], N_1 and N_2 are the number of particles ($\sim 10^{11}$ in proton-proton collisions at the LHC) per bunch for each beam, f is the revolution frequency ($\sim 10^9$ Hz for the LHC), N_b is the number of bunches ($\sim 10^3$ proton bunches in the LHC) in each beam, and $4\pi\Sigma_x\Sigma_y$) is the beam effective area in the transverse plane, as illustrated in Fig. 3.4($\Sigma_x = \Sigma_y = 17\,\mu$m for the LHC beams).

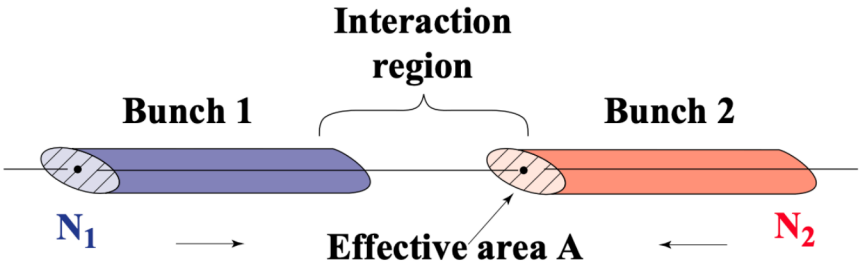


Figure 3.4: Sketch of two colliding beams [85].

In general, any detector which measures particle hit rates with a linear response can be

used as a luminometer. The effective area can be determined by the so called *van der Meer scan* [86] first used in ISR at CERN, which consists in separating the beams in the transverse plane and performing scans in x and y directions by moving one of the beams with respect to the other to determine the effective area. This method is based on measuring the rates from the luminometer as a function of the beam displacement as shown in Fig. 3.5.

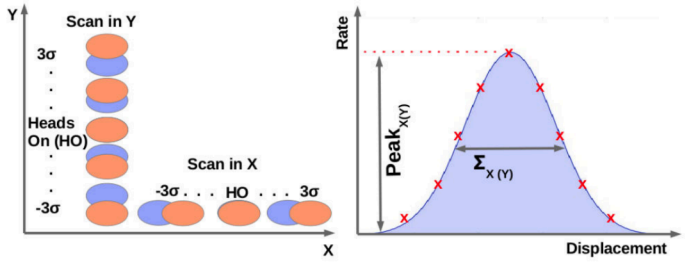


Figure 3.5: Illustrative sketch of the VdM scans in x and y planes (left). Sketch of the luminometer rates as function of the displacements in a VdM scan(right). This figure is taken from Ref. [87].

This formula, described above (eq. 3.2) corresponds to the instantaneous luminosity, and by integrating over time the integrated luminosity ($\mathcal{L}_{int}$) can be obtained:

$$\mathcal{L}_{int} = \int_0^T \mathcal{L}(t)\ dt \tag{3.3}$$

The integrated luminosity, expressed in units of inverse cross-section (e.g., fb^{-1}), measures the collected data in a given time interval (T). In Fig. 3.6, the integrated luminosity delivered by the LHC and recorded by CMS experiment during 2016 proton-proton collisions at a center-of-mass energy of 13 TeV is shown.

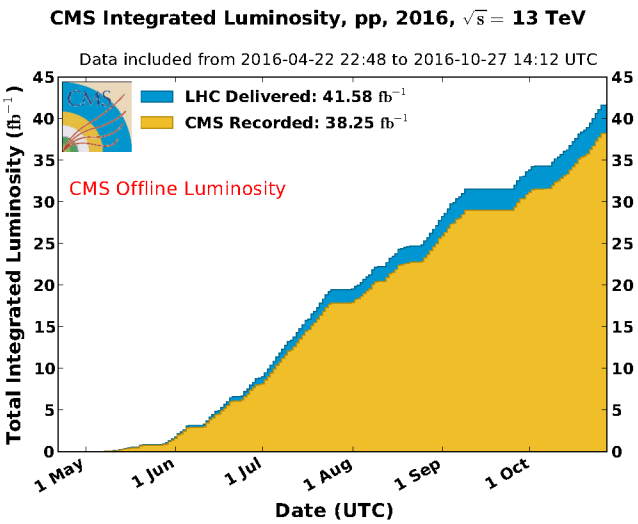


Figure 3.6: Total integrated luminosity in 2016 proton-proton collisions [88]

3.2 The Compact Muon Solenoid

The CMS experiment, located in one of the four interaction points of the LHC ring, together with ATLAS, LHCb, and ALICE experiments, constitutes the most advanced experimental setup ever built to search for the most fundamental questions in the history of science. The CMS detector is like our time travel machine, which can show us the conditions of the quark-gluon plasma of the early universe. It allows us to study the properties of matter at the highest densities and temperatures, testing our fundamental predictions from the standard model and beyond. The three-dimensional sketch of the detector with its components is represented in Fig. 3.7. The detector is shaped like a cylindrical onion with several layers. It

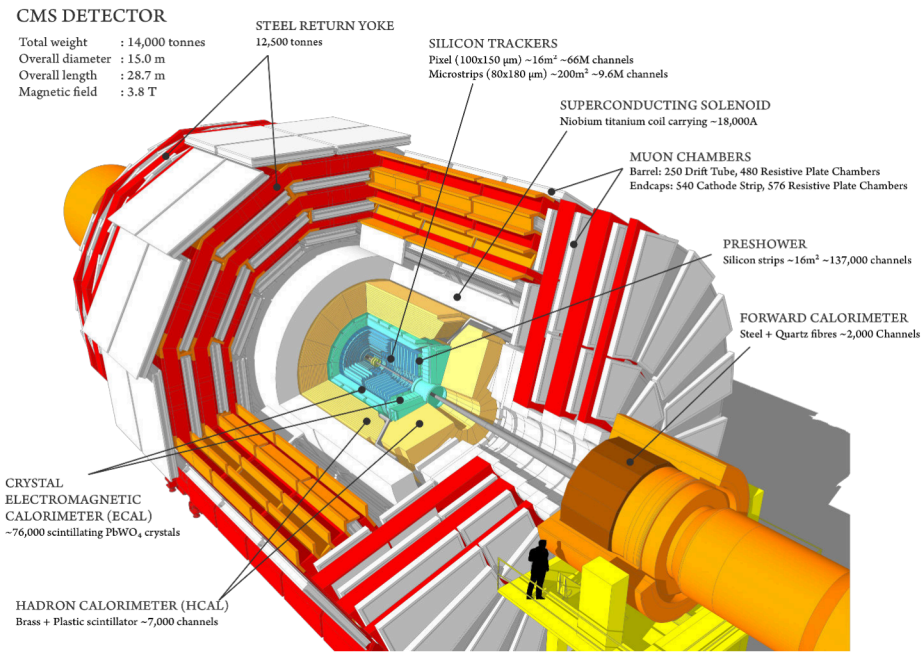


Figure 3.7: The CMS detector and all its sub-detectors. The figure give us general information about the detector size and components, and a small description of the materials used in each sub-detector(figure taken from Ref. [89]).

is composed of a superconducting solenoid magnet, with a homogeneous magnetic field of 3.8 T designed to bend the trajectory of the high energetic particles from the collisions. It is a distinctive element of the detector as the last word in its name. The volume of the solenoid magnet is large enough to have inside the tracking system and the calorimeters. Outside the solenoid, the dedicated muon stations are embedded in an iron yoke which returns the magnetic flux. The detector is designed to reconstruct most of the outgoing particles of the collisions: charged leptons and hadrons, neutral hadrons and photons, by measuring their momenta and energies.

3.2.1 The coordinate system

The coordinate system adopted by convention in CMS [74] has the origin centered at the collision point inside the detector, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC ring. The z-axis points along the beam

direction counter-clockwise. The azimuthal angle ϕ is measured from the x-axis in the xy plane, and the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the z-axis.

In the context of proton-proton collisions, the center-of-mass of the colliding partons inside the protons is affected by boosts, depending on the momentum fractions from the protons carried by the partons at the moment of the interactions. This fact makes it more suitable to use quantities invariant under Lorenz boots along the z-axis. This leads us to use the pseudorapidity(η) instead of θ since variations of η (for massless particles) are invariant under Lorentz boosts along the z-axis, it is defined as:

$$\eta = -\ln\left(\tan\frac{\phi}{2}\right) \tag{3.4}$$

The pseudorapidity can be written as function of the three momentum as:

$$\eta = \frac{1}{2}\ln\left(\frac{|\mathbf{p}| + \mathrm{p_z}}{|\mathbf{p}| - \mathrm{p_z}}\right) \tag{3.5}$$

The pseudorapidity, is equivalent to the rapidity in high energy collisions, where the momentum of the particles is much bigger than its mass, this mean $|\mathbf{p}| \approx E$ and $\eta \approx y$, which is expressed as:

$$y = \frac{1}{2}\ln\left(\frac{\mathrm{E} + \mathrm{p_z}}{\mathrm{E} - \mathrm{p_z}}\right) \tag{3.6}$$

In this context a coordinate system given by (r, ϕ, η) can be used. Another important quantity in high energy physics is the transverse momentum, defined as the component of the three momentum in the transverse plane (x-y plane):

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{3.7}$$

The CMS detector hermeticity and the high performance of the different detector subsystems allow the accurate measurement of the missing transverse energy (MET) associated with particles with almost no interaction with the detector materials, such as neutrinos and possible new physics particles. The MET is ideally defined as the magnitude ($E_{\mathrm{T}}^{\mathrm{miss}}$) of the momentum vector imbalance, in the x-y plane, of all the particles reconstructed in the event:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\mathrm{i}\in\mathrm{particles}} \vec{p}_{T\mathrm{i}} \tag{3.8}$$

3.2.2 The solenoid magnet

The solenoid magnet is one of the key features of the CMS detector. This powerful magnet is the largest superconducting solenoid ever constructed and allows the tracker and calorimeters to be placed inside the coil and forms a compact detector, as the first letter in its name indicates. To contain and return the magnetic field flux from the solenoid, a return yoke made of a 12-sided iron structure surrounds the magnet (see Fig. 3.8). The yoke is made of three layers reaching out fourteen meters in diameter, providing most of the detector's structural support. Also, as it is in the outer part of the detector, it acts like a filter that allows only weakly interacting particles as neutrinos or muons to pass through. This 12

Figure 3.8: The CMS magnet during construction. The cylindrical magnet coil and the return yoke (12-sided red structures) are shown (this figure is taken from [90]).

500 tons component, formed by the magnet coil and the return yoke, is the heaviest in the CMS detector and generates a magnetic field a hundred thousand times stronger than Earth's magnetic field. The CMS *superconducting* magnet allows electricity to flow without resistance and creates a powerful magnetic field of about 3.8 Tesla (100 000 times higher than the Earth's magnetic field). This high magnetic field allows bending the trajectories of high momentum charged particles identifying the positively and negatively charged particles which turn in opposite directions in the same magnetic field. It also allows measuring the momentum of the charged particles by measuring their curvature radius.

3.2.3 Tracking detectors

The CMS tracker records the charged particles path allowing the measurement of the curvature radius in the magnetic field and, finally, the momentum of the particles:

$$p_{\mathrm{T}} = 0.3(B \cdot R) \tag{3.9}$$

where B is the intensity of the magnetic field strength in Tesla and R is the curvature radius in meters, and the transverse momentum p_{T} is in GeV/c. Finally the momentum is obtained from the polar angle θ using trigonometry as follows:

$$p = \frac{p_{\mathrm{T}}}{\sin\theta} \tag{3.10}$$

The CMS tracker is also crucial for the efficient and precise primary (point of a pp interac-

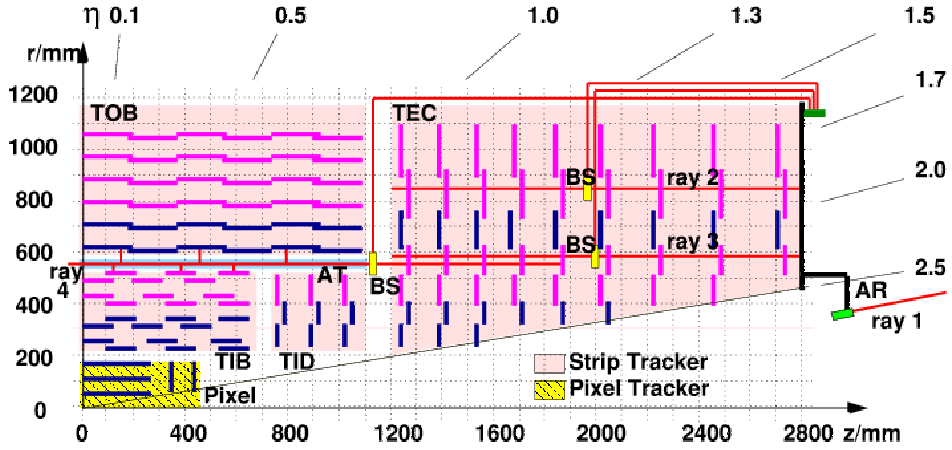


Figure 3.9: Sketch of a quarter of the CMS tracker (this figure is taken from Ref. [91]). The silicon pixel and silicon strip trackers are shown. The strip tracker is divided in the Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOC), Tracker Inner Disks (TID) and Tracker End Caps (TEC) by the magenta shaded regions. The strips made of double (single) sided silicon sensors are represented in blue (magenta) color.

tion) and secondary (displaced vertex located outside the beam profile) vertex reconstruction in the collisions, given its fine granularity. The CMS tracker is composed of an innermost pixel detector and outermost silicon strips tracker. A schematic view of the CMS tracking system is shown in Fig. 3.9

Silicon pixel tracker

When collecting 2016 data used in this work, the silicon pixel detector was the Phase-0 pixel detector. The pixel detector is the closest to the beam pipe, it is composed by three barrel layers (BPIX) at radii of 4.4 cm, 7.3 cm, and 10.2 cm, and two forward/backward disks (FPIX) at longitudinal positions of $z = \pm 34.5$ cm and $z = \pm 46.5$ cm and in a radius from about 6 to 15 cm. The BPIX are formed by 48 million pixels in an area of $0.78\,\mathrm{m}^2$, and the FPIX had 18 million pixels in $0.28\,\mathrm{m}^2$. The size of a pixel module is 100×150 $\mu\mathrm{m}^2$, about two hairs width; this ensures measuring the 3-D path of charged particles with a single hit resolution between $10 - 20\,\mu$m.

Silicon strip tracker

The CMS Silicon Strip Tracker (SST) is structured in two barrels, the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB), made of 4 and 6 layers, respectively. The TIB is closed by two Tracker Inner Disks (TID), made of three wheels on each side, and the TOB is in between two Tracker End Caps (TEC) made of 9 wheels. The SST has a diameter of 2.4 m and length of 5.5 m, with a large active area of $198\,\mathrm{m}^2$. The tracker acceptance covers a region of 2.5 in absolute pseudorapidity ($|\eta| < 2.5$). The entire SST consists of 15 148 modules, each one holding one or two silicon sensors with the corresponding readout electronics. Most of the silicon modules are single-sided, but to obtain tridimensional information to separate tracks, double-sided sensors (contains two back-to-back silicon sensors) are placed in the two innermost layers of TIB and TOB, and also in the two innermost rings of TID and TEC, and in addition in the fifth ring of TEC, as shown in Fig. 3.9, where the strips with double side

sensors are shown in blue and the single side ones in magenta. The sensors in the TIB and TID provide a single hit resolution between 23 μm and 35 μm in the $r-\phi$ direction; on the other hand, the ones in the TOB and TEC provide a hit resolution between 35 μm and 53 μm in the $r-\phi$ direction. The single point hit resolution in the $z-$direction is about 230 μm in the TIB and 530 μm in the TOB.

3.2.4 Calorimeters

In particle physics, the energy of the particles involved in the collision is measured by an experimental apparatus called calorimeter; this name comes from the fact that almost all the energy of the particles is converted to heat. Most particles enter the calorimeter and initiate a particle shower, and its energy deposited in the calorimeter is measured. The energy can be measured entirely using a homogeneous calorimeter, or it may be sampled using a sampling calorimeter. The material that produces the particle shower is different from the material that measures the deposited energy. The calorimeters are often the most practical way to detect and measure neutral particles from an interaction. They are also crucial in calculating the missing energy attributed to particles as neutrinos which rarely interact with matter and escape the detector.

3.2.4.1 Electromagnetic calorimeter

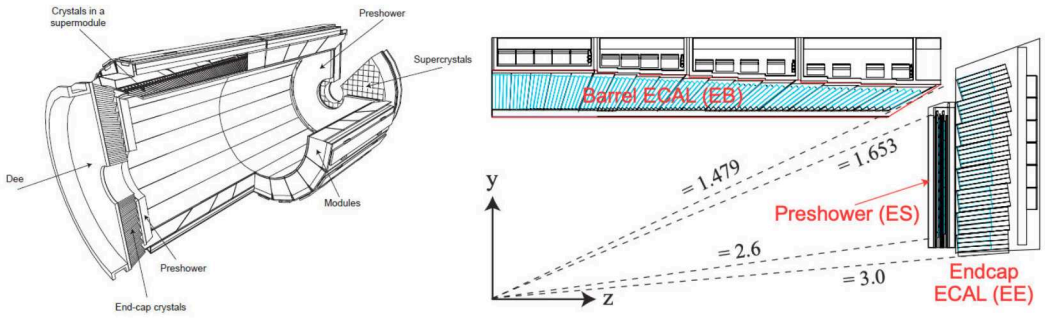
The CMS electromagnetic calorimeter (ECAL) is a homogeneous calorimeter with cylindrical geometry, made in a total of 75 848 lead tungsten ($PbWO_4$) scintillating crystals [92] with a truncated pyramidal shape. The $PbWO_4$ crystals give a relatively fast response, with 80% of the light being emitted within 25 ns, ensuring high granularity, and is a radiation-hard material suitable for the high radiation environment of the LHC collisions. The ECAL, located inside the solenoid magnet and right outside the tracker, is composed of the barrel (EB), two end caps (EE), and the preshower detector (ES), as shown in Fig. 3.10. Its main purpose is to measure the energy and position of photons, electrons, as well as charged hadrons. It has a high resolution and efficient photon identification, which resulted to be crucial in the observation of the Higgs boson via the $\mathrm{H} \to \gamma\gamma$ decay channel [14]. The ECAL barrel energy resolution can be parameterized as a function of the incident electron/photon energy (E in GeV) [93]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E[\,\mathrm{GeV}]}} \oplus \frac{12\%}{E[\,\mathrm{GeV}]} \oplus 0.3\% \tag{3.11}$$

where the first term depends on the event to event fluctuation in lateral shower containment, photo-statistics, and photo-detector gain; the second term represents the noise, which depends on the electronic noise and event pileup (other particles causing signals that overlap in time); and the last term, the constant term which dominates at high energies, depends on the non-uniformity of the longitudinal light collection, leakage of energy from the rear face of the crystal and the accuracy of the detector inter-calibration constants.

Electromagnetic barrel calorimeter

The ECAL barrel covers the central region ($|\eta| < 1.479$) and is located at a radial distance of 1.3 m from the beamline. The EB is made of 36 supermodules of half-barrel length, equipped with avalanche photodiodes (APDs) and readout electronics, each one composed



(a) (b)

Figure 3.10: Layout view of the CMS electromagnetic calorimeter(ECAL), with the arrangement of crystal modules, supermodules, endcaps, and the preshower (a). Schematic view of one-quarter of the CMS ECAL, in the yz plane, showing the barrel ECAL (EB), preshower (ES), and endcap ECAL(EE) regions and its pseudorapidity coverage indicated with dashed lines (b). Figures taken from Ref. [93].

of 4 modules which contain 400 or 500 $PbWO_4$ crystals, according to their η position. In total the EB contains 61 200 crystals with an approximate cross-section of 22×22 mm^2 in the $\eta \times \phi$ plane and a length of 230 mm in the radial direction. The scintillation light produced in the crystals by the interacting particle is converted into an electrical current by the ADPs. An internal gain is added by using higher electric fields that increase the number of charge carriers collected. Since the gain factor is susceptible to the temperature and the applied voltage, the ECAL has a cooling system designed to extract the heat from the readout electronics and keep the temperature of the crystals and photo-diodes constant with a ± 0.05 °C precision, which helps to preserve the energy resolution in the measurements.

Electromagnetic endcap calorimeter

The two electromagnetic endcaps are placed perpendicular to the beam pipe and cover a pseudo-rapidity range of $1.479 < |\eta| < 3.0$ and are located at $|z| = \pm 314$ cm. The endcaps are made of two semicircular aluminum halves called *Dee*'s, each consisting of 3662 tapered $PbWO_4$ crystals, with a front area of 2.68×2.68 cm^2 and length of 22 cm. The crystals in each *Dee* are arranged into 138 standard 5×5 structures called supercrystals and 18 special shaped supercrystals located at the inner and outer radii. Vacuum photodiodes detect the scintillation light from the crystals (VPTs) [94], with an active area of 280 mm^2, which are glued to the rear face of the crystals.

Preshower detector

The preshower detector [95] is a sampling calorimeter of two layers, a lead absorber layer that initiates the electromagnetic shower, and a silicon strips detectors plane placed in the rear face of the absorbers. There are two preshower detectors placed in front of each EE, covering the region $1.653 < |\eta| < 2.6$. The main feature of the ES is to detect neutral pions (π^0) and improve the γ/π^0 discrimination. This detector also identifies electrons against minimum ionizing particles and determines the position of electrons and photons with very high granularity.

3.2.4.2 Hadron calorimeter

The CMS hadron calorimetry system (HCAL) [96] is composed of four subsystems: the HCAL barrel (HB), HCAL endcap (HE), HCAL outer (HO), and HCAL forward (HF). The HB and HE are located between the ECAL and the magnet coil, while the HO is outside the coil and the HF is in the most forward direction close to the beam pipe as it is shown in Fig. 3.11, providing full coverage of $|\eta| < 5$. The HCAL main feature is the hadron jets identification and measurement of its energy. In addition, the signature from particles as neutrinos and exotic particles, characterized by the presence of MET, is identified by using the HCAL subsystems information.

The HB calorimeter [97] is sampling calorimeter which use brass as the absorber and plastic scintillator as active material. The HB covers up to $|\eta| = 1.4$, it is segmented into 72×32 towers in $\phi \times \eta$ corresponding to a size of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The towers are composed by 155 mm thick brass plates interleaved with plastic scintillator plates of 3.7 mm thickness, except for the first scintillator plate which is of 9 mm thickness.

The HE covers the range of $1.3 < |\eta| < 3.0$ overlapping partially with the HB coverage. It is segmented in 72×13 towers in $\phi \times \eta$, with a size of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$ for the more central region($|\eta| < 1.6$) and $\Delta\phi \times \Delta\eta = 0.017 \times 0.017$ for the most forward region ($|\eta| > 1.6$).The composition of the towers layers is similar to the HB towers but using 79 mm thick brass absorbers followed by the plastic scintillators.

The HO is an array of plastic scintillators located outside the magnet in the region $|\eta| < 1.26$, which uses the steel return yoke and the CMS magnet material as absorber [98]. The HO complements the HB and EB calorimeters since the barrel calorimeters do not fully contain the hadronic shower, given the space limitations inside the magnet. The HO improves the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurements at LHC high energies and is helpful for muon identification.

The HF is located 11 m away from the interaction point covering the most forward region $3.0 < |\eta| < 5.0$. It is a Cherenkov-based calorimeter made of a steel absorber and quartz fibers that collect the Cherenkov light, primarily from the calorimeters electromagnetic showers from the calorimeter [99]. The HF can detect both the electromagnetic and hadronic components in the shower and plays a major role in the forward jets identification, $E_{\mathrm{T}}^{\mathrm{miss}}$, and luminosity measurements.

3.2.5 Muon system

The CMS muon system provides accurate identification and measurement of muons. The muons are measured in the inner tracker after passing the coil. They also pass through the calorimeter and deposit only ionization energy, but this information is then complimented using the muon system. Right after passing the calorimeters, the muons enter the muon system, located within the iron yoke. Made by several chambers, the muon system is designed to identify and measure minimum ionizing muons.

The muon system, shown in Fig. 3.12, consists of three different gaseous detectors to identify and measure muons [101]. In the central region, $|\eta| < 1.2$, drift tube (DT) chambers are used in the central region, where neutron-induced background, muon rate, and residual magnetic field are low. On the other hand, covering the endcaps region up to $|\eta| < 2.4$, where the muon rate, neutron-induced background, and residual magnetic field are high, cathode strip chambers (CSC) are used. Finally, in addition to DT and CSC systems, resistive plate chambers (RPC) are used in the central region and up to $|\eta| < 1.6$ in the endcaps region.

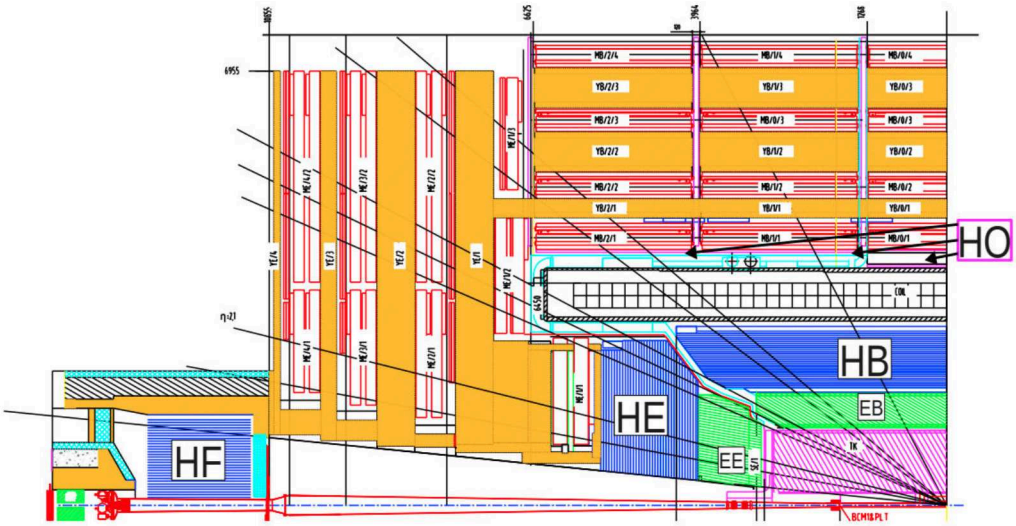


Figure 3.11: Sketch of a quarter of the CMS detector in r-z view, showing the location of the HB, HE, HO, and HF subsystems of the HCAL detector [96].

The RPC chambers provide a fast response with good time resolution, while the DT and CSC chambers provide better position resolution.

3.2.6 Trigger system

The high collision rate ($40MHz$) at the LHC could produce too much data that is impossible to store and study. For this reason, the trigger systems select only the potentially interesting physics events reducing this tremendous rate ($\sim 100Hz$) by performing a fast online selection. In the CMS experiment, the trigger system [102] has two levels or steps for recording the interesting data: the Level-1 trigger (L1) and the high-level trigger (HLT), which achieves this data reduction.

The L1 trigger has a latency of $\sim 3.2\,\mu$s, achieved by fast detector components such as the calorimeters and the muon system, making a simplified event reconstruction. The selection is based on physics objects such as photons, electrons, muons, or jets which are above certain p_{T} threshold or some more global event characteristics as $E_{\mathrm{T}}^{\mathrm{miss}}$ or $H_{\mathrm{T}} = \sum_{jets} p_{\mathrm{T}}$. The L1 trigger reduces the rate to 100 kHz and keeps the high-resolution data of the accepted events quickly acceded by the HLT. The HLT consists of a processor farm of ~ 10000 CPU cores that process the data similarly to offline processing filtering the events based on the physics objects of possible interest for the data analysis. The HLT data processing improves the situation, and the selection is based on offline-quality reconstruction algorithms, which ensure an average processing time of ~ 100 ms per event. The HLT is designed around the concept of *HLT path* implemented as a sequence of increasing complexity steps that both reconstruct and make selections on the physics objects.

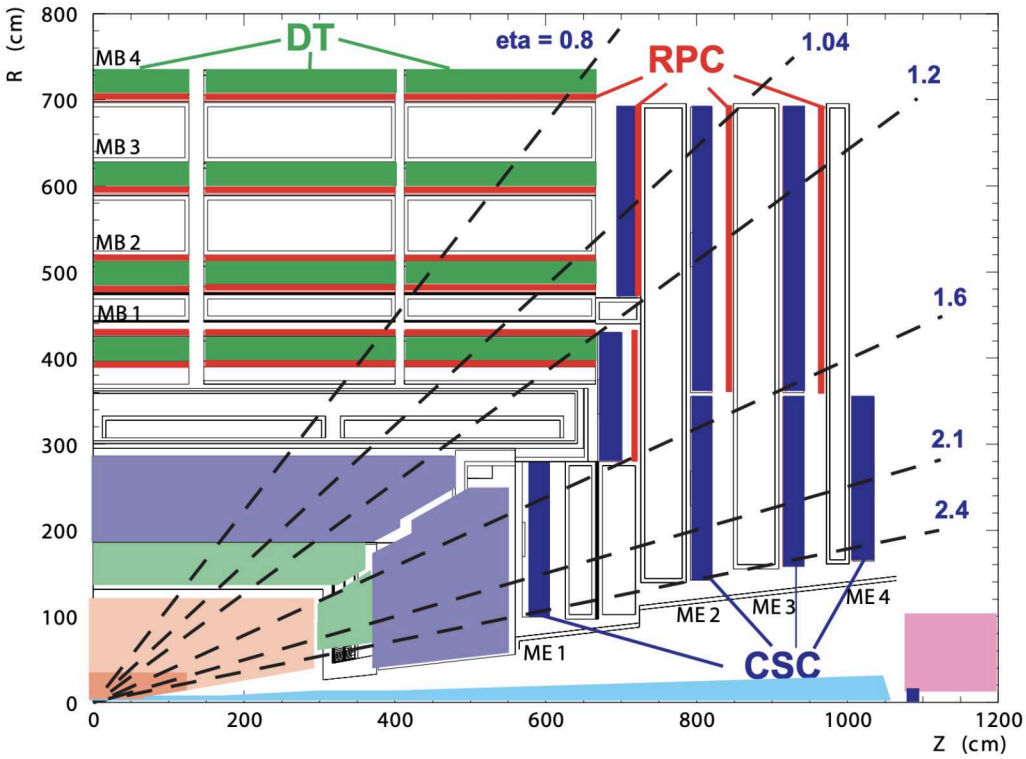


Figure 3.12: Layout of a quarter of the CMS detector in r-z view, showing the DT system located in the central region, the RPC system limited to $|\eta| < 1.6$ in the endcap, and the CSC system (only the inner ring of the ME4 chambers have been deployed) [100].

CHAPTER

4

EVENT RECONSTRUCTION AND SELECTION

Contents

4.1 Particle identification and Particle Flow algorithm 41
4.2 Jet reconstruction and selection 43
4.3 Event selection 45
4.4 Observables and phase space definition 45
4.4.1 Four leading jets p_{T} spectra 46
4.4.2 Multi-differential jet multiplicity 46
4.5 Removing the background 47

Using all the information from sub-detectors and trigger systems described in the previous chapter (in Sec. 3.2) and based on the concept of Particle Flow (PF) [103] the CMS offline event reconstructions are carried out. A complete PF reconstruction algorithm tuned to the CMS detector was developed and has been widely used in physics analysis within the CMS collaboration. In this chapter, I will describe how the PF algorithm is used in the jet reconstruction since jets are the physics objects used in the measurements carried out in this thesis. The event selection will be discussed in Sec. 4.3, and the observables measured in this thesis will be presented and defined at the end of the chapter.

4.1 Particle identification and Particle Flow algorithm

Starting from the interaction point, particles first enter the tracker, in which charged-particle trajectories (*tracks*) and its origins (*vertices*) are reconstructed from the signals (*hits*) of the tracking detector layers. The vertices are classified as primary vertices (PV) located at

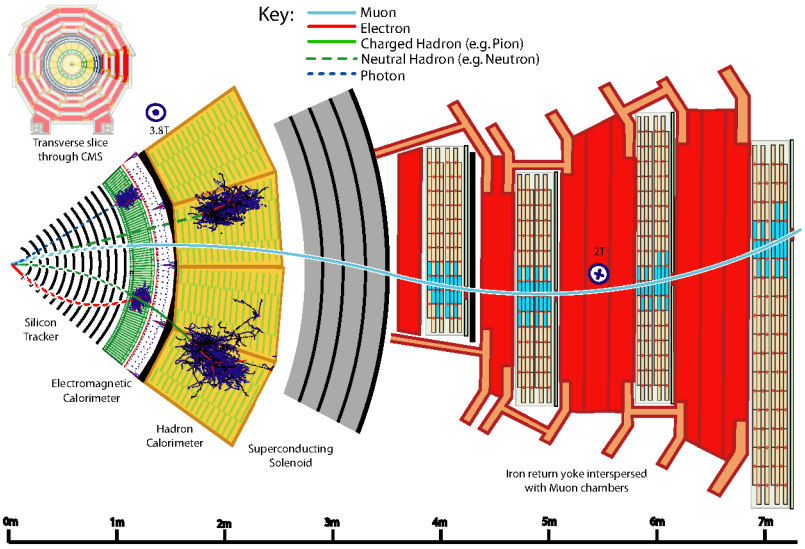


Figure 4.1: Transverse slice of the CMS detector, from the beam interaction region (most left) to the muon chambers (most right). The interaction of different particles is depicted (figure taken from Ref. [103])

the beamline or secondary vertices (SV) if displaced. The effect of bending the trajectories by the intense magnetic field of the solenoid magnet allows the measurement of the momenta of charged particles. Electrons and photons are absorbed into the ECAL, and the corresponding electromagnetic showers are detected as clusters of energy, from which energy and direction of the particles can be measured. Neutral and charged hadrons can also initiate a shower in the ECAL. Still, they are only fully absorbed in the HCAL and using the corresponding clusters, the energy and direction of the hadrons can be determined. Other particles, such as muons and neutrinos, pass the calorimeters with little or no interactions, respectively. While neutrinos escape undetected, muons produce hits in the muon detection chambers located outside the calorimeters. In Fig. 4.1 the signature from muon, electron, charged hadron, neutral hadron, and a photon is depicted.

The information from tracks and clusters must be correlated to identify each final-state particle using particle-flow reconstruction to achieve an improved event description. The CMS detector is designed for PF reconstruction with a large magnetic field to separate neutral and charged particles' energy deposits in the calorimeters (HCAL and ECAL), a fine-grained tracker which provides a pure and efficient charge-particle trajectory reconstruction, and an excellent muon tracking system. The PF reconstruction uses iterative tracking, based on a combinatorial Kalman filter method [104] for pattern recognition. This ensures high efficiency while keeping a low track miss-identification rate. In addition, PF clusters are formed from "cluster seeds", which are identified as calorimeter cells with energy over a certain threshold. The *PF tracks* and *PF clusters* are combined by using a "link algorithm", forming the called *PF blocks*, possibly connected to single particles. Using the information from the *PF blocks*, the algorithm creates the *PF candidates*. For each block, the muon candidates are reconstructed first, and its corresponding PF elements are removed from the *PF block*. The electron reconstruction follows, intending to collect also the energy of bremsstrahlung photons. Energetic and isolated photons are identified in the same step. The block's remaining

elements are subject to the cross-identification of charged hadrons, neutral hadrons, and photons. The PF jets reconstructed by clustering the *PF candidates* are used in this analysis as detector level jets.

4.2 Jet reconstruction and selection

In CMS, the four-vector of the event particles is reconstructed using the particle flow (PF) algorithm [105]. These four vectors are then used as input to the anti-k_T (discussed in Sec. 2.3.6) jet clustering algorithm. The radius resolution parameter of the anti-k_T algorithm was set to $R = 0.4$ (also called AK4 jets), which is the "default" cone size for CMS at LHC RUN II. The jets are considered within the following two classifications:

- Detector level jets: Jets reconstructed with particles coming from PF reconstruction. Particles coming from pileup interactions can produce calorimeter energy depositions and additional tracks that contaminate the kinematics of the reconstructed jets. To reduce the contamination coming from additional proton-proton interactions within the same or neighboring bunch crossings (pileup), the charge hadron subtraction (CHS) technique [106] is used. The CHS removes charged particles that originate from pileup vertices, and the remaining neutral particle energy is then subtracted under the assumption that it is uniformly distributed inside the detector as illustrated in Fig. 4.2. Based on this approach three different jet identification (ID) criteria are defined [107] but the one used in this thesis measurements is the so called Tight Lepton Veto (TLV), with an efficiency of more than 99% for selecting real jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.5$[1]. The TLV jet ID is summarized in Table 4.1.

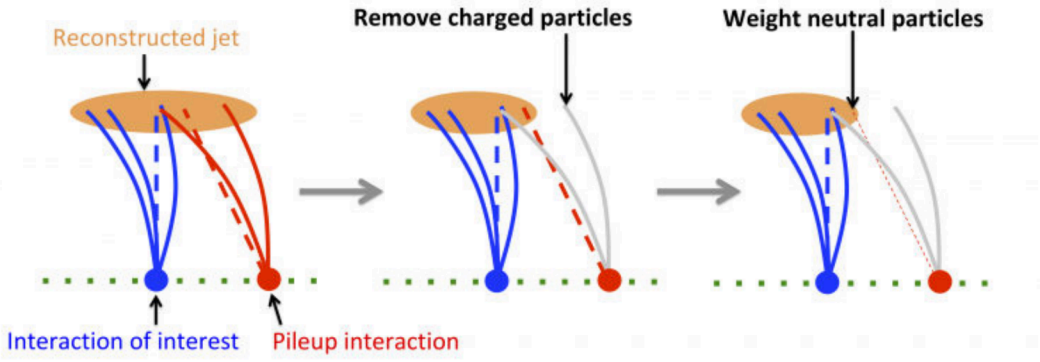


Figure 4.2: Sketch of the CHS technique. The two steps of pileup removal are depicted. Figure taken from Ref. [109]

- Particle level jets: Jets obtained by applying the jet reconstruction algorithm on MC particles produced at the very end of the hadronization stage. The jets clustered from stable ($c\tau > 1cm$) and visible (non-neutrino) final state particles (eg. electrons and muons). Particle level jets are also called generator level jets.

[1] We do not use PU jet ID [108] for pileup treatment since the minimum p_T cut we use is $p_T > 50\,\mathrm{GeV}$ for the extra jets instead the JetID criteria is used.

Jet variables	$\|\eta\|$ range	Tight Lepton Veto
Charged hadron fraction	$\|\eta\| < 2.4$	> 0.0
Charged multiplicity	$\|\eta\| < 2.4$	> 0
Charged em. fraction	$\|\eta\| < 2.4$	< 0.9
Muon fraction	$\|\eta\| < 2.4$	< 0.8
Neutral hadron fraction	$\|\eta\| < 2.7$	< 0.9
Neutral em. fraction	$\|\eta\| < 2.7$	< 0.9
Neutral multiplicity	$2.7 < \|\eta\| < 5$	< 0.9
	$2.7 < \|\eta\| < 3$	> 2
	$3 < \|\eta\| < 5$	> 10

Table 4.1: Jet ID requirements in the different regions of the detector for the Tight Lepton Veto selection [107].

The different levels at which a jet can be reconstructed are depicted in Fig. 4.3. The parton level jet definition was used to compare theoretical predictions with unfolded data. It can be more model-dependent, and one does not include the non-perturbative effects, then particle level (labeled as Hadrons in the figure) is used. Then, when we refer to jets, we will refer to the ones reconstructed at detector level or particle level.

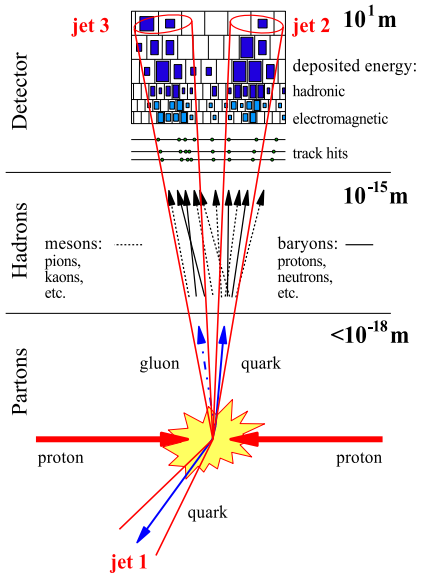


Figure 4.3: Sketch of the jet in different levels of reconstruction. First the partons come from the hard scattering (parton-level), then they hadronize and become stable-particles (particle level), and finaly these particles interact with the tracker and leave energy deposits in the hadronic and electromagnetic calorimeter (detector-level) (Figure taken from Ref. [110]).

4.3 Event selection

This thesis is focused on measurements of QCD jet observables. For this propose each event is selected to have at least one off-line reconstructed vertex [111] with at least one primary vertex (PV). The PV should be located in a cylindrical volume (according to CMS coordinate system presented on Sec. 3.2.1) defined by $|z(PV)| < 24\,\text{cm}$ and $\rho < 2\,\text{cm}$, where $|z(PV)| = 0$ (proton-proton collision point along the beamline) is at the center of the CMS detector and ρ is the radius in the $x - y$ plane. The number of degrees of freedom in the vertex fit must be bigger than four. All detector (particle) level jets with $p_T > 20\,\text{GeV}$ and $|y| < 3.2$ are considered [2] (ordered by decreasing p_T) and the event is selected if it contains at least two jets with $p_{T1}(p_{T2}) > 200(100)\,\text{GeV}$ and $|y_1|(|y_2|) < 2.5$. Additional jets must have $p_T > 50\,\text{GeV}$ and $|y| < 2.5$. In Fig. 4.4 an example of the data events selected in this measurements is shown, there the two leading jets are in nearly back-to-back topology with four extra jets.

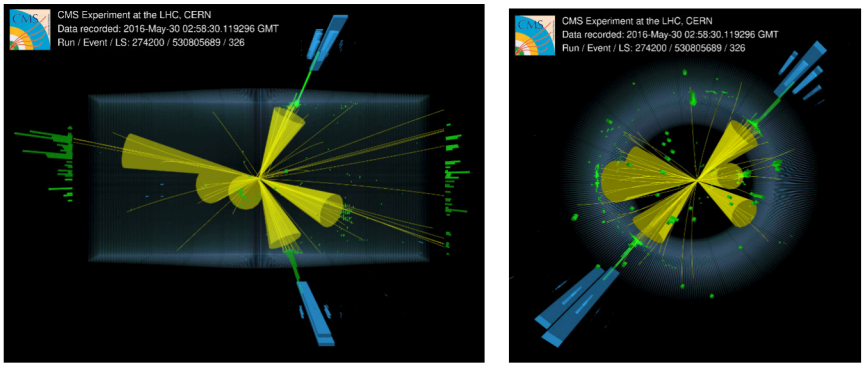


Figure 4.4: This is an event with six jets (yellow cones) with transverse momentum of more than 50 GeV each and the two leading jets of more than 1.5 TeV, produced in proton-proton collisions at a centre-of-mass energy of 13 TeV at the LHC. The yellow lines correspond to trajectories of charged particles in the tracking detectors and the blue and green shaded towers show the energy deposits in the calorimeters. The left figure corresponds to the yz plane and the right one to the azimuthal (xy) plane. Figures taken from Ref. [112]

4.4 Observables and phase space definition

In this thesis, the measurement of the cross section as a function of the jet multiplicity N_{jets}, the azimuthal angle $\Delta\phi_{1,2}$ between the two leading jets, and the transverse momentum of the leading jet p_{T1} in dijet events is presented. In addition, the cross section as a function of p_T for the four leading jets in such events is also measured. This dijet multi-differential cross sections will help us to understand how many jets build the azimuthal decorrelation in dijet events in different regions; of special interest is the nearly back-to-back region ("Sudakov region") where $\Delta\phi_{1,2}$ is close to π ($170^o < \Delta\phi_{1,2} < 180^o$) and resummation (parton shower) effects are dominant. In addition, the measurement of the four leading p_T jets differential cross sections (as a function of each jet p_T) bring information on the kinematics of such events.

[2] In App.B the effect of using $|y| < 5.0$ instead of $|y| < 3.2$ is studied in simulation and in data.

4.4.1 Four leading jets p_{T} spectra

The differential p_{T} cross sections of the four leading jets are measured. The two leading jets (p_{T1} and p_{T2}) are measured up to 2 TeV and the third (p_{T3}) and the forth (p_{T4}) are measured near to the TeV scale. The definitions (at particle level) for the four leading jets p_{T} are:

$$\frac{\mathrm{d}\sigma_{pp\to jj}}{\mathrm{d}p_{T1}},\ \frac{\mathrm{d}\sigma_{pp\to jj}}{\mathrm{d}p_{T2}},\frac{\mathrm{d}\sigma_{pp\to jjj}}{\mathrm{d}p_{T3}},\frac{\mathrm{d}\sigma_{pp\to jjjj}}{\mathrm{d}p_{T4}} \tag{4.1}$$

where $\sigma_{pp\to jj}$,$\sigma_{pp\to jjj}$,$\sigma_{pp\to jjj}$ corresponds to the inclusive dijet, trijet, and four jet cross section in proton-proton collision. This meas that we need at least two jets ($N_{jet} \geq 2$) to measure $\sigma_{pp\to jj}$ as function of p_{T1} and p_{T2}, and at least three jets ($N_{jet} \geq 3$) to measure $\sigma_{pp\to jjj}$ as function of p_{T3} and four jets ($N_{jet} \geq 4$) to measure $\sigma_{pp\to jjjj}$ as function of p_{T4}. This measurement is done taking into account all the correlations among the p_{T} in 2-jet, 3-jet, and 4-jet events, and is revolutionary in QCD jet measurements since before the p_{T} spectra of the jets (eg. in $Z+jets$ [113,114] and in multijet events [115]) have been measured without measuring the effect of such correlations (more details on the correlations treatment will be given in Chapter 6). The p_{T} spectra are measured with the same binning for all the four jets, with the following (particle level)[3] bin edges:

$$\begin{aligned} p_T^{edges}/\,\mathrm{GeV} = [&50,\ 74,\ 100,\ 133,\ 166,\ 200,\ 272,\ 330,\ 395,\ 468,\ 548,\ 638,\\ &737,\ 846,\ 967,\ 1101,\ 1248,\ 1410,\ 1588,\ 1784,\ 2000] \end{aligned} \tag{4.3}$$

4.4.2 Multi-differential jet multiplicity

The measured cross section as function of the jet multiplicity, $\Delta\phi_{1,2}$ and p_{T1} can be defined (at particle level) as follows:

$$\frac{\mathrm{d}\sigma_{dijet}}{\mathrm{d}N^i_{jet}\mathrm{d}p^j_{T1}\mathrm{d}(\Delta\phi^k_{1,2})} \tag{4.4}$$

where i, j, k corresponds to the binning scheme used for N_{jets}, p_{T1}, and $\Delta\phi_{1,2}$ defined in the following:

- $\Delta\phi_{1,2}$ $[0^o, 150^o, 170^o, 180^o]$
- p_{T1} [200,400,800,13000] (GeV)
- N_{jets} $[=2, =3, =4, =5, =6, \geq 7]$

This is the first time in QCD jet measurements that the dijet cross section as a function of the jet multiplicity [115] have been measured differentially in $\Delta\phi_{1,2}$ and p_{T1}. In addition, also for the first time, the correlations among the bins considered were taken into account (more details on the correlations treatment will be given in Chapter 6)).

[3]The particle level bins comes from the detector level binning by merging two neighboring bins to ensure a better unfolding behavior. Then the detector level jet p_{T} edges are:

$$\begin{aligned} p_T^{recedges}/\,\mathrm{GeV} = [&50,\ 64,\ 74,\ 84,\ 100,\ 117,\ 133,\ 153,\ 166,\ 196,\ 200,\ 245,\ 272,\ 300,\\ &330,\ 362,\ 395,\ 430,\ 468,\ 507,\ 548,\ 592,\ 638,\ 686,\ 737,\ 790,\ 846,\\ &905,\ 967,\ 1032,\ 1101,\ 1172,\ 1248, 1327, 1410, 1497\ 1588, 1684, 1784, 1890, 2000] \end{aligned} \tag{4.2}$$

These new measurements were already presented in July 2021 at the European Physical Society conference on high energy physics (EPS-HEP). The CMS Physics Analysis Summary, for the results presented at the EPS-HEP conference, can be found in Ref. [116].

4.5 Removing the background

Since the study will focus on high and low $\Delta\phi_{1,2}$ regions, we study the fraction of missing transverse energy, defined as the ratio between the missing transverse energy and the sum of all calibrated jets and unclustered particles transverse energies in the event (see definitions in Sect. 5.5.1). This quantity, known as MET fraction, is essential for distinguishing between pure QCD jets events and what we call non-QCD events (events with vector bosons or $t\bar{t}$ plus jets) since the contribution from non-QCD events will create a tale in the distribution. Typical contributions are *Z/W + jets* where W and Z bosons decay into leptons and neutrinos. Fig. 4.5 shows that a cut on this quantity becomes essential to describe the low $\Delta\phi_{1,2}$ region. Therefore events with a MET fraction bigger than 0.1 are vetoed in this analysis. This cut was mainly done for removing non-QCD events contributing to exclusive dijet scenarios measured in the jet multiplicity ($N_{jets} = 2$), more details can be found in App. C

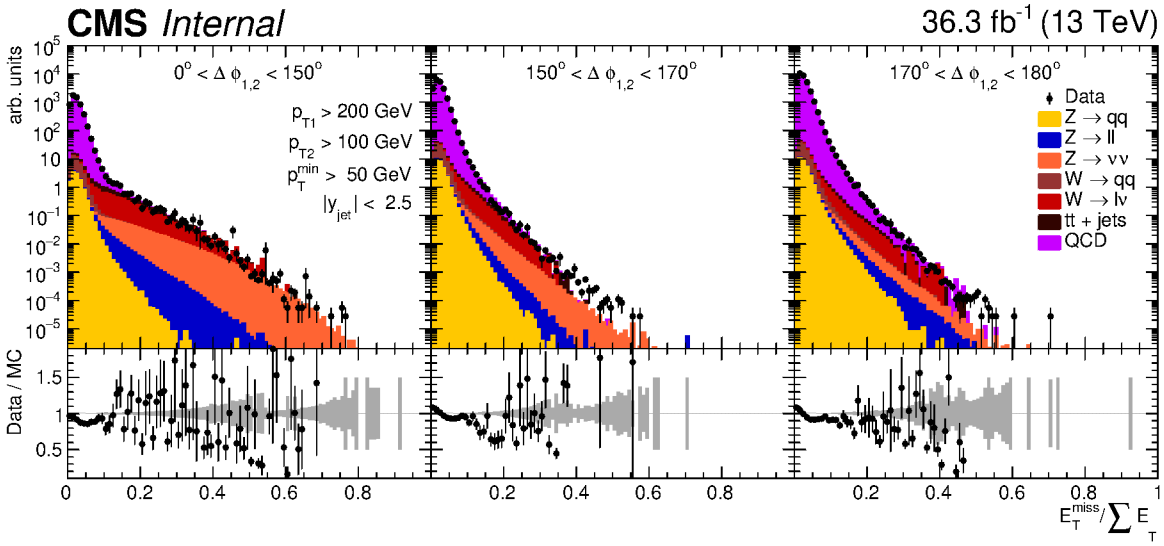


Figure 4.5: MET fraction in bins of $\Delta\phi_{1,2}$.

CHAPTER

5

ANALYSIS AT DETECTOR LEVEL

Contents

5.1 Data and simulated samples 50

5.2 Pileup corrections 51

5.2.1 Measurement of pileup in data 52

5.2.2 Pileup simulation 52

5.2.3 Removal of overweighted events in simulation 53

5.2.4 Procedure of pileup reweighting 56

5.3 Jet Energy Corrections 56

5.3.1 Jet Energy Scale 57

5.3.2 Jet Energy Resolution 58

5.4 Trigger strategy 61

5.4.1 Trigger efficiency determination 62

5.5 Additional corrections 65

5.5.1 Missing transverse energy 65

5.5.2 Hot zones correction 69

5.5.3 Prefiring correction 71

5.6 Comparison data to simulation 72

In this chapter, the analysis at detector level will be described going from the data and MC samples through all the relevant MC and data corrections. At the very end the comparison between data and MC at detector level will be shown for the observables presented in Sec. 4.4.

5.1 Data and simulated samples

In this thesis the data[1] used were recorded with the CMS experiment during Run II in 2016 at a centre-of-mass energy $\sqrt{s}$ of 13 TeV, corresponding to an overall integrated luminosity of 36.3 fb^{-1} for AK4 jets. The data are split into several run periods, as shown in Tab. 5.1. Only *certified* data lumi-sections[2] (LS), *e.g.* with good detector operation conditions, were included, as listed by the so-called golden JSON file[3].

Period	Luminosity [fb^{-1}]	# raw events
B	5.828411734	24,663,381
C	2.617679657	8,868,031
D	4.285851496	14,693,971
E	4.065496038	12,712,080
F	3.135465534	9,033,104
G	7.652808375	21,827,728
H	8.739883629	23,943,253
2016	36.3	115,741,548

Table 5.1: The overview of the data periods used in the analysis.

To correct for the detector-related effects, like the resolution, and unfold the data to the particle-level for comparison with theoretical predictions, several centrally produced MC samples are used, including a simulation of the CMS detector response using GEANT4 package [117]:

- PYTHIA8 (tune CUETP8M1) [118] a Leading-Order (LO) MC generator with $2 \to 2$ parton-parton interaction included in the Matrix Element (ME). The realistic event-structure is achieved by initial- and final-state Parton Showers (PS); the Multi-Parton Interaction (MPI), and hadronization based on Lund String Model [119]. The p_T ordering is applied for the parton shower [120]. This sample uses CMS Underlying Event (UE) Tune PYTHIA8 Monash 1 (CUETP8M1) [121], which is based on the Monash tune [122] but with an extra tuning to the UE-sensitive observables measured by CMS.

- MADGRAPH (tune CUETP8M1) [123] a tree-level ME generator, incorporating $2 \to 2$, $2 \to 3$ and $2 \to 4$ QCD processes in the present set-up. PYTHIA8 consequently simulates the PS, MPI, and hadronization with the same tune as the plain PYTHIA8 sample. The partons from the ME are matched to the parton shower using MLM matching scheme [124].

- HERWIG++ (tune CUETHS1) [125] is a LO MC generator, which compared to the PYTHIA8 is using angular-ordered PS to correctly treat the color coherence effects [126]

[1] /JetHT/Run2016[B-H]-17Jul2018-v1/MINIAOD

[2] smallest fixed time period in data taking ($\sim$ 24 seconds)

[3] /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions16/13TeV/ReReco/Final/Cert_271036-284044_13TeV_23Sep2016ReReco_-Collisions16_JSON.txt

and the cluster model [127] for hadronization. The CMS UE Tune HERWIG++ Set 1 (CUETHS1) is used in this sample.

Since the jet-p_T spectrum is steeply falling (roughly going as $\sim 1/p_T^4$), the event rate for the high-p_T events has to be enhanced compared to the low-p_T region where the cross section is larger. Then for producing samples of "comparable" statistical precision at high and low p_T one can generate unweighted events in hard scale slices or use event reweighting in a flat hard scale interval. The PYTHIA8 and MADGRAPH samples are produced using unweighted events independently in several slices of $\hat{p}_T$ [4] and H_T [5] respectively. In App D the cross sections with the number of events and the corresponding luminosity is shown in Tab. D.1 and in Fig. D.1 the contribution of each slice is plotted as function of the particle level p_{T}. On the other hand, the HERWIG++ sample [6] is generated using event reweighting in a flat $\hat{p}_T$ interval (cross section $\sigma = 1667000000\,\mathrm{pb}$, with 4140205 events, generated with $\hat{p}_T \in [15, 7000]\mathrm{GeV}$). Also a flat PYTHIA8 sample [7] (cross section $\sigma = 2022100000\,\mathrm{pb}$, with 9799552 events, generated with $\hat{p}_T \in [15, 7000]\mathrm{GeV}$) is used to cross check with the sample in slices. All this samples were centrally produced by CMS given its complexity.

Despite there are flat and merged samples for PYTHIA8 for doing cross-check during the analysis of the detector effects, we will mainly only use the merged sample with more events and consequently a smaller statistical error. Also, the MADGRAPH merged sample will be used, and the HERWIG++ flat sample. We will profit from using these tree samples generated with three different models. We will analyze the model dependence when unfolding the data. Since the HERWIG++ flat sample and the PYTHIA8 merged sample have much fewer events (in the phase space of interest: eg. for $N_{jets} = 2$ for $0^\circ < \Delta\phi_{1,2} < 150^\circ$ and $p_{T1} > 800$ the statistical uncertainty from PYTHIA8 sample doubles the one coming from MADGRAPH, this can be seen in Fig. 5.20 at the very end of this chapter in Sec. 5.6) than the MADGRAPH sample. Then, we can not simply address the model dependence by using an envelope created with the nominal sample used and the two others for the variations when calculating an uncertainty. This issue motivates to use of a model reweighing procedure to estimate the model uncertainty as presented in Sec. 6.2.2.

The samples contain jets at hadron (particle level) and detector levels for the hard events; a simulation of the pileup is also included by adding several additional interactions from the minimum bias (MB) sample, which was produced separately using PYTHIA8. The simulation of the pileup will be extensively discussed in Sect. 5.2.

5.2 Pileup corrections

Each collision at the LHC generates highly populated events with more than one proton-proton collision per event, as shown in Fig. 5.1. The pileup (PU) is an unavoidable consequence of high instantaneous luminosity at the LHC since it corresponds to the *number of pp interactions* (originated from primary vertices) per LHC bunch crossing. Therefore, it is essential to include a good simulation of the PU in the MC samples. In this section, general considerations about the PU are given. The two procedures applied in this thesis to correct

[4] /QCD_Pt_*to*_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM

[5] /QCD_HT*to*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM

[6] /QCD_Pt-15to7000_TuneCUETHS1_Flat_13TeV_herwigpp/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM

[7] /QCD_Pt-15to7000_TuneCUETP8M1_Flat_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_magnetOn_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM

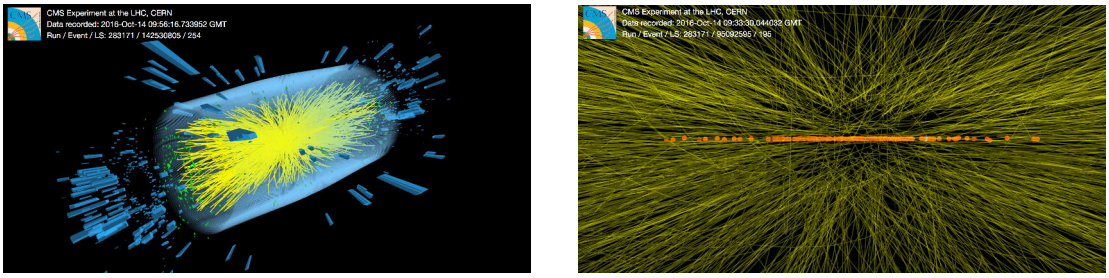


Figure 5.1: (Left) Proton-proton collision at a center-of-mass energy of 13 TeV recorded during a high pileup fill of Run II at the LHC. (Right) The primary vertices of events from isolated bunches with average pileup around 100 are shown as orange dots. Figures taken from Ref. [128]

the PU simulation are described: first, the reduction of overweighted PU events in the MC samples, and second, the reweighting procedure of the PU profile. From the point of view of physics, the contamination from the PU comes from additional tracks or calorimeter clusters from PU interaction which may end within the jet from the main interaction. The PU contamination is mitigated in the PF reconstruction by the CHS (charge hadron subtraction technique as discussed in Sec. 4.2), which removes tracks from a jet if they are associated to another PV. In addition, the Jet Energy Scale (JES) corrections also account for residual PU effects [129], originating mainly from the neutral particles (discussed later in Sect. 5.3).

5.2.1 Measurement of pileup in data

The measurement of the number of pp interactions per LHC bunch crossing (μ) in data is based on exploiting the relation between cross section and instantaneous luminosity (see eq. 3.1) for the total inelastic cross section, or minimum bias cross section $\sigma_{\text{MB}} = 69.2 \pm 3.2$ mb [130] [131]. If a single bunch crossing have an instantaneous luminosity ($\mathcal{L}_i$) then:

$$\mu = \frac{\mathcal{L}_i \cdot \sigma_{\text{MB}}}{f} \tag{5.1}$$

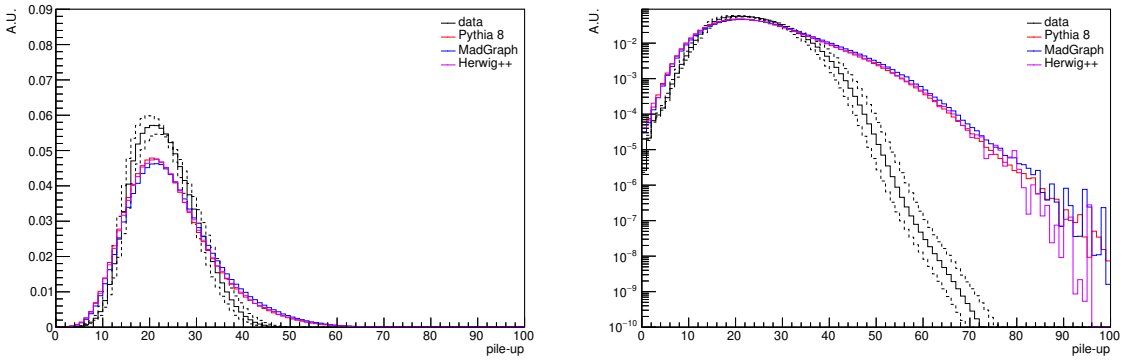
here f is the LHC orbit frequency, needed to convert from a per-time quantity (as $\mathcal{L}_i$) into a per-collision quantity. The uncertainty on the pileup measurement in data is taken as a $\pm 1\sigma$ variation of the nominal value. This is done in practice by varying $\pm 5\%$ the minimum bias cross section.

5.2.2 Pileup simulation

The PU simulation consists in generating QCD events following a Poisson distribution:

$$\mathcal{P}(x, \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda) \tag{5.2}$$

However, over a long-running period, the mean value of PU (λ) is changing, and, consequently, a pure Poisson distribution does not describe the resulting distribution. For this reason, several Poisson distributions with different λ parameters are added to simulate better the data. Afterward, the mean of the Poisson distribution for MC is slightly higher than the expected average PU in data. In other words, the MC samples are produced with slightly



(a) Linear scale.

(b) Logarithmic scale.

Figure 5.2: PU profile in data and simulation before any correction. Here, the data PU measurement uncertainties are shown as black dashed lines.

overestimated scenarios as shown in Fig .5.2. The tale of the MC pileup distributions is larger than the one from the data due to the overestimated mean.

5.2.3 Removal of overweighted events in simulation

Understanding what is the PU made of is the first step to find out possible issues. The PU is made of any possible signal found at CMS:

inelastic = diffractive + non-diffractive

diffractive = single-diff. + double-diff. + central-diff.

This is illustrated in Fig. 5.3, and in Tab. 5.2 the approximate cross-section of each component of the diffractive and inelastic signals are shown. We call minimum bias (MB) signal whatever leaves a significant signal in the HF, excluding the elastic component of the total cross section. The high p_T jets in diffraction are suppressed compared to non-diffractive scenario, since the diffractive proton has large energy which reduces the available energy for the jets. In practice, the PU is simulated using PYTHIA8 with `SoftQCD:inelastic=on`, with cross section $\sigma_{\text{inelastic}} \approx 75$ mb. This value is to be contrasted to the inclusive jet cross section $\sigma_{\text{incl jet}} \approx 1.5 - 2.0$ mb for $p_T > 15\,\text{GeV}$.

To get many events for all p_T values, the PYTHIA8 (MADGRAPH) samples are generated in slices of $\hat{p}_T$ (H_T), as mentioned in Sect. 5.1. However, the PU simulation is performed in addition without $\hat{p}_T(H_T)$ slices. Therefore, when normalizing each slice to the given cross section, some events with high-p_T jets (coming from the PU simulation) will largely dominate the population of specific bins at higher p_T in an unrealistic way.

In summary, the way the PU is simulated causes two issues: *normalization* since the signal is double-counted and *fluctuations* due to the lack of statistics of the inclusive jet signal from the MB sample. The correction to this double problem consists of suppressing the MB sample's high p_T contribution. It is *a priori* not straightforward to find the jets coming from the PU since the information at generator level is available only for the main scattering. As a solution, the following recipe is applied (the same applies for $\hat{p}_T$):

type	cross section
total	$\sim 100\,\text{mb}$
elastic	$\sim 25\,\text{mb}$
inelastic	$\sim 75\,\text{mb}$
$\rightarrow$ single-diffractive	$\sim 15\,\text{mb}$
$\rightarrow$ double-diffractive	$\sim 10\,\text{mb}$
$\rightarrow$ central-diffractive	$\sim 1\,\text{mb}$
$\rightarrow$ non-diffractive	$\sim 50\,\text{mb}$

Table 5.2: Content of the MB signal at 13 TeV as predicted from PYTHIA8. The exact values depend on the MC generator and the UE tune.

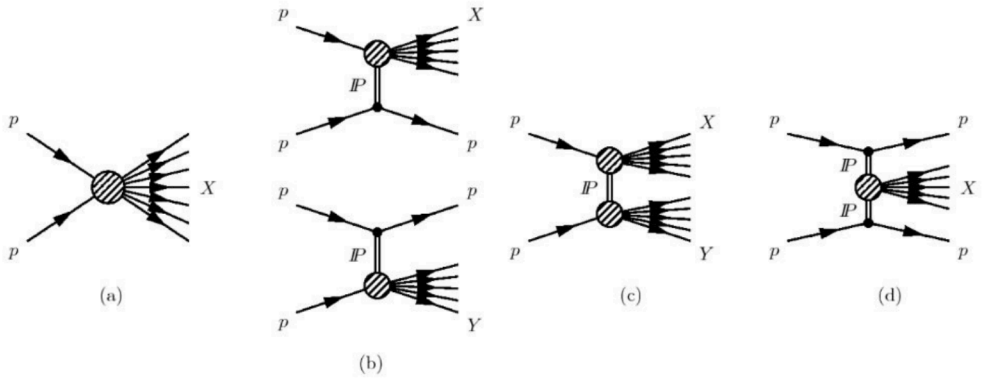


Figure 5.3: Types of interactions in MB signal. From left to right: (a) non-diffractive,(b) single-diffractive, (c) double-diffractive, central-diffractive. Figure taken from Ref. [132].

1. Hard cut: removes events if $\max H_T^{\text{PU}} > H_T^{\text{main}}$. This cut-off is the strongest and significantly affects the H_T spectrum.

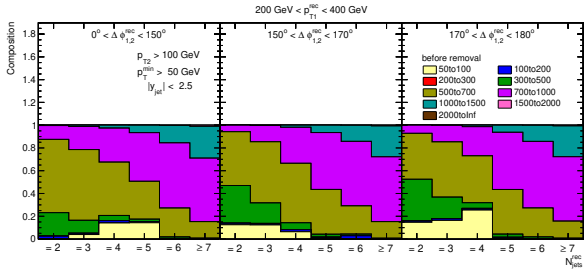
2. Additional removal of residuals: reject contributions beyond a maximal weight allowed for each p_T^{rec} bin. This mainly removes remaining fluctuations at high p_T^{rec}.

3. Renormalization: correct the obtained H_T distribution to the original H_T distribution (for avoiding possible side effects from first step).

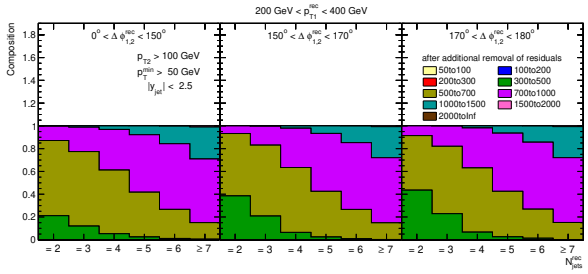
The technical details on the determination of the maximal weight and the two different approaches to correct the flat samples and the samples in slices are discussed in App. E. The effect on the contribution from different H_T slices in the jet multiplicity and p_{T} spectra of the first four leading jets is shown for MADGRAPH in Fig. 5.4-5.5 (the complete set of plots for PYTHIA8 and MADGRAPH is shown in App. F)

In the jet multiplicity, Fig. 5.4a, is clearly showing how lower H_T slices contribute to higher ones, eg. 50to100 (GeV) slice contribute with fluctuations up to 25% in the back-to-back region. Then in Fig. 5.4b after applying the "PU cleaning" procedure this no longer

happens. Fig. 5.4 corresponds to the lower p_{T1} bin ($200\,\text{GeV} < p_{T1} < 400\,\text{GeV}$) where the contamination is higher (the other two bins can be found in App. F).

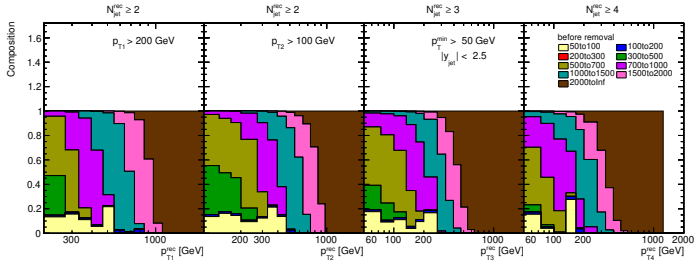


(a) Before cleaning

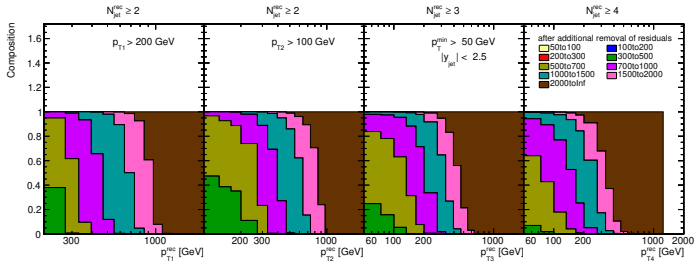


(b) After cleaning

Figure 5.4: Contributions from the different H_T slices to the detector-level distributions of MadGraph sample before and after correction for highly weighted PU events for the jet multiplicity distribution (only the first p_{T1} bin is shown).



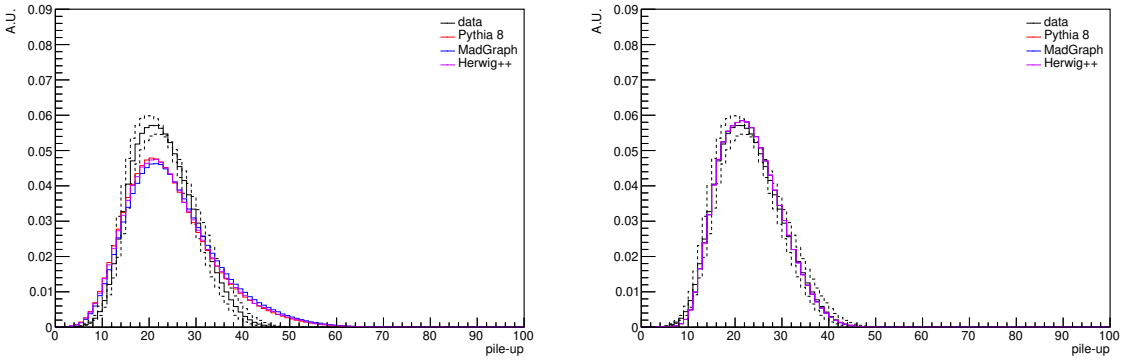
(a) Before cleaning



(b) After cleaning

Figure 5.5: Contributions from the different H_T slices to the detector-level distributions of MadGraph sample before and after correction for highly weighted PU events for the four leading jets p_{T} distributions.

For the four leading jet p_{T} spectra, the contamination from the first H_T slice is visible for



(a) Before reweighting.

(b) After reweighting

Figure 5.6: PU profile in data and simulation. Here, the data PU measurement uncertainties are shown as black dashed lines.

all the four jet spectra up to $\approx 200\,\text{GeV}$ and is around 10-20% as shown in Fig. 5.5a. After the additional removal of residuals, this effect is removed as shown in Fig. 5.5b. Other plots that illustrate how the overall normalization changes after the PU cleaning are shown in App. F. The procedure makes essentially no effect in the particle level spectra, but at detector level, the normalization can change 20-40% depending on the MC sample.

5.2.4 Procedure of pileup reweighting

Reweighting to the observed PU is applied to reproduce the data PU distribution with the MC samples. The distribution on which this procedure is applied is the so-called *PU profile*, which is the probability distribution of the number of interactions in the bunch crossings. Within one lumi-section (LS), the instantaneous luminosity is assumed to be constant, and consequently, the PU profile is simply the Poisson distribution. Over a longer time window (*e.g.* in a fill: time, since injection of proton bunches, in which the luminosity decrease and the beam is dumped), the instantaneous luminosity varies, and consequently, the PU profile distribution is not a plain Poisson anymore.

Since trigger prescales are used (Sec. 5.4), each LS of data is normalized, and a single PU profile corresponding to the data is used. The profile from the simulation samples and data are extracted, and the ratio of both is taken as a correction factor to the profile in the simulation. The absolute normalization does not matter, especially, and the procedure is straightforward. The PU profile is shown in Fig. 5.6 before and after the reweighting procedure.

5.3 Jet Energy Corrections

At CMS, the performance of the jet reconstruction is studied centrally [107, 133]. The detector level jets are corrected for scale and resolution effects of the detector. In this section, the calibration of jets both in data and simulation are discussed. The jet (raw refers to the detector level jets) energies reconstructed at the detector are corrected to particle level jets (see particle and detector level jets definitions in Sec. 4.2), this is the so-called *true* level.

5.3.1 Jet Energy Scale

The purpose of Jet Energy Scale (JES) corrections is to correct the measured to the true energy of jet in the form of a global multiplicative factor to the four-momentum as follows:

$$p_\mu^{\text{true}} = C_{\text{JES}} \times p_\mu^{\text{raw}} \tag{5.3}$$

This correction factor C_{JES} can be factorized into several components, applied in a chain:

$$C_{\text{JEC}} = C_{\text{offset}}(p_T^{\text{raw}}, \eta, A, \rho) \times C_{\text{MC}}(p_T', \eta) \times C_{\text{residual}}(p_T'', \eta) \tag{5.4}$$

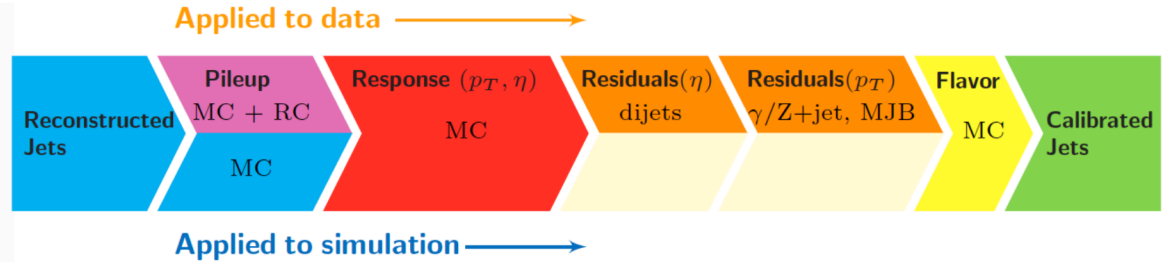


Figure 5.7: Schematic description of the JEC application, here RC stands for *Residual Corrections*. Taken from [133].

The corrections are computed at three different *levels* in sequence as described in eq. 5.4 and illustrated by a diagram in Fig. 5.7, in the following order:

1. The *offset correction* C_{offset} (also `L1Offset` corrections) removes everything that is not related to the pp collision of the hard interaction, *e.g.* pileup and electronic noise ($p_T^{\text{raw}} \to p_T'$). In addition to the η and p_T^{raw} this correction also depends on the jet area A and mean energy deposit per unit of the area of the $\eta\phi$ plane.

2. The *MC calibration* C_{MC} (also `L2L3relative` corrections or MC truth correction) corrects for the main non-uniformities in pseudorapidity and non-linearities in transverse momentum ($p_T' \to p_T''$); typically, the calorimeters have a non-linear response. The correction is determined from the MC sample to have $\langle p_T^{rec}/p_T^{gen} \rangle$ exactly at one. By definition it is identical for MC and Data.

3. The *residual corrections* (also `L2L3residual` corrections) C_{residual} accounts for finer corrections between data and MC ($p_T'' \to p_T^{\text{true}}$) and are applied only to data. The determination of this correction can be factorized into two parts:

   - Relative energy scales are corrected by investigating dijet topologies, where the same energy is "expected" from both jets in opposite directions as illustrated in Fig. 5.8. The probed jet is in the forward direction ($|\eta| > 1.3$), whereas the tagged jet in the central rapidity region ($|\eta| < 1.3$). In practice both jets do not have the same energy, e.g. because of the extra radiation coming from a third jet (Jet 3) as shown in Fig. 5.8. For addressing this the transverse momentum of the third jet is defined as $p_{T3} < \alpha \cdot p_T^{\text{ave}}$ where $p_T^{\text{ave}} = (p_T^{\text{prob}} + p_T^{\text{tag}})/2$. Then the events are studied using p_{T} balance ($p_T^{\text{tag}}/p_T^{textprob}$) (exploding the momentum conservation in the

transverse plane). The p_{T} balance is analyzed for different points of $\alpha < 0.3 \to 0$, extrapolating to $\alpha = 0$ which represent the ideal scenario $p_T^{\mathrm{prob}} = p_T^{\mathrm{tag}}$.

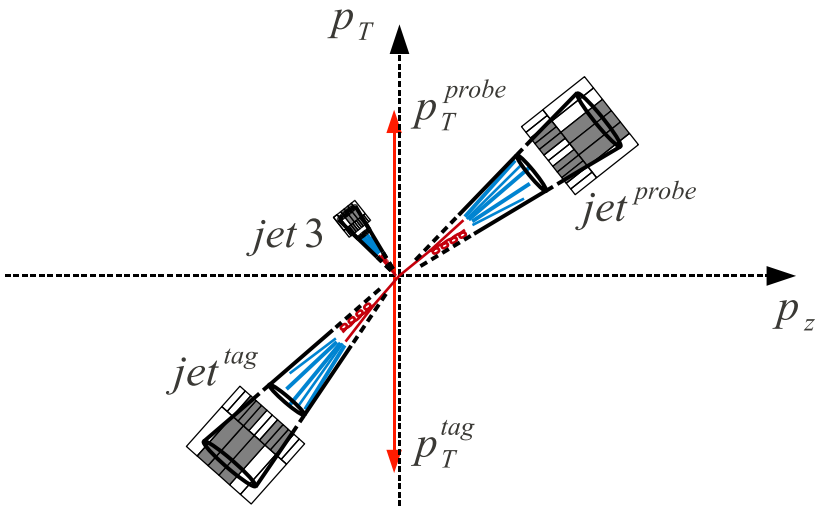


Figure 5.8: Schematic representation of a dijet topology used in the tag and probe method. Taken from [134].

- Absolute energy scales are corrected by investigating Z+jets and γ+jets topologies, where the photon energy is measured in ECAL, and Z energy is reconstructed from decays into electrons or muons in the tracking system. As discussed in the previous item, a similar probe and tag method is used for these events by Z/γ as tagged objects. The energy of the high p_T jets can be calibrated using multijet topologies, where high p_T jet should be balanced by several low-p_T jets, which are well-calibrated from Z + jets and γ + jets events.

At CMS, the JES corrections are provided centrally [129,133], together with an estimation of the associated uncertainties. The last recommended JES corrections (at the time of writing this thesis) provided by the JetMET group are used, namely `Summer16_07Aug2017_V11`.

5.3.2 Jet Energy Resolution

The choice of the binning scheme is related to the Jet Energy Resolution (JER); in QCD measurements at CMS, the binning is standard for all jet analyses to ease comparing measurements. Still, in this particular analysis, we have modified the p_{T} binning to have edges corresponding to the selection (see Sec. 4.3). The resolution on the transverse momentum is usually finer in MC than in data; therefore, the transverse momentum has to be smeared to match the real conditions, which is crucial for well-working unfolding. In the current subsection, we explain the procedure of smearing the transverse momentum spectrum in MC.

JER smearing

In simulation samples, given a jet generated (reconstructed) with a transverse momentum p_T^{gen} (p_T^{rec}), the *resolution* is defined as follows:

$$\Delta = \frac{p_T^{\text{rec}} - p_T^{\text{gen}}}{p_T^{\text{gen}}} \tag{5.5}$$

For a given p_T^{gen}, it is a Gaussian-like curve with a core and two tails. The left tail is more important because of various reconstruction effects (e.g. situations where a jet is reconstructed as two jets, or situations where a pileup jet is considered by mistake). The resolution curve is usually fitted with a (double) Crystal-Ball curve. Sometimes, the term *resolution* refers specifically to the width σ_{JER} of this Gaussian-like curve.

The resolution in data and the smearing scale factors (SF)[8] to correct the simulation to data are measured and released centrally at CMS by JetMET group. In Fig. 5.9 the JER resolution and SF (with their associated uncertainty) are shown. Using the SFs makes sure that the resolutions are similar in data and MC when applying smearing on the spectrum of transverse momentum. Given the resolution in simulation Δ_{MC}, the reconstructed and generated transverse momenta are related by the following formula (which is a consequence of eq. 5.5):

$$p_T^{\text{rec}} = p_T^{\text{gen}} \cdot (1 + \Delta_{\text{MC}} R_G), \tag{5.6}$$

where R_G is a random number distributed according to the Gaussian distribution with variance 1 and mean 0. Two methods exist to correct the resolution [135]:

Scaling method This method assumes that the following matching between reconstructed and generated jets can be done:

- $\delta R < R_{\text{cone}}/2$ where the R_{cone} is the cone size radius of the jet clustering algorithm (here $R_{\text{cone}} = 0.4$), and $\delta R = \sqrt{(\delta y)^2 + (\delta\phi)^2}$ is the angular separation;
- $|\Delta_{\text{MC}}| < 3\sigma_{\text{JER}}$ where σ_{JER} is the measured resolution in data (the goal of this condition is to avoid jets populating the tails of the resolution curves).

Then the resolution obtained from the MC value of p_T^{rec} has to be corrected to the data resolution by using the SFs:

$$\Delta_{\text{data}} = \text{SF} \cdot \Delta_{\text{MC}} \tag{5.7}$$

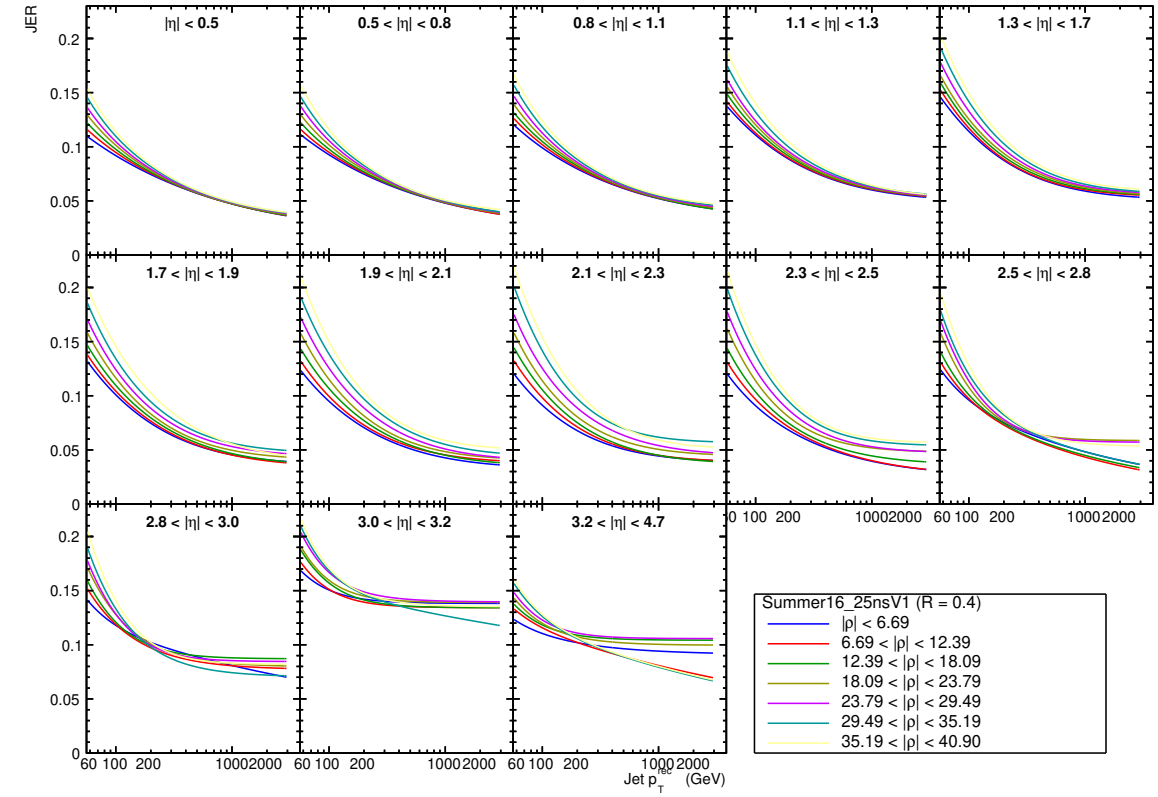
Given this correction to the resolution, the value of p_T^{rec} can be corrected in turn:

$$p_T^{\text{rec}} = p_T^{\text{gen}} \cdot (1 + \Delta_{\text{MC}} R_G) \tag{5.8}$$

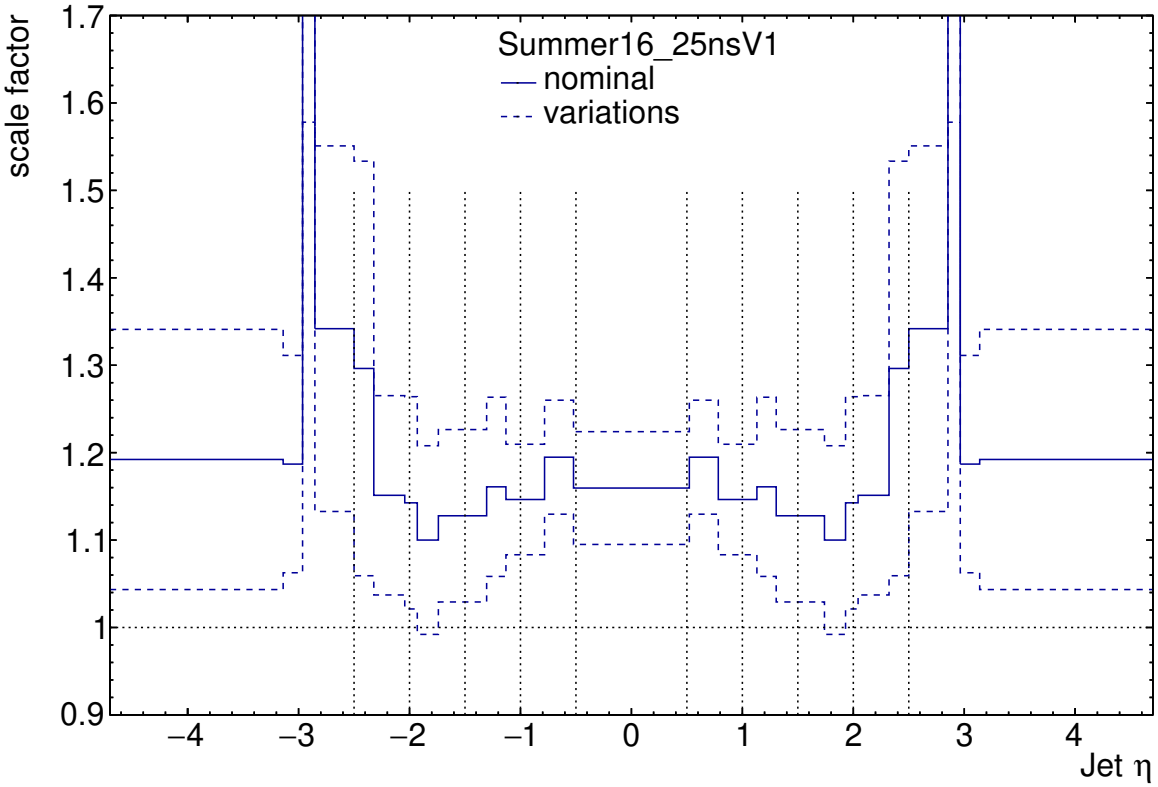
$$\longrightarrow p_T^{\text{rec}} = p_T^{\text{gen}} \cdot (1 + \text{SF} \times \Delta_{\text{MC}} R_G) \tag{5.9}$$

Stochastic method This alternative method is intended to be used if no matching can be done; in this case, one resorts to random numbers. One picks a number from the data resolution according to a centered Gaussian distribution with width equal to the

[8] In this analysis, we use the scale factors with their corresponding uncertainty and the resolutions provided by JetMET, namely `Summer16_25nsV1`.



(a) JetMET resolution (`Summer16_25nsV1`)



(b) JetMET Scale Factors (`Summer16_25nsV1`)

Figure 5.9: Jet energy resolution measured in data (a) and smearing scale factors (b) provided centrally in CMS by JetMET group .

resolution σ_{JER}, *i.e.* one picks $\mathcal{N}(0, \sigma_{\text{JER}})$; then the transverse momentum is smeared accordingly:

$$p_T^{\text{rec}} \longrightarrow p_T^{\text{rec}} \cdot \left(1 + \mathcal{N}(0, \sigma_{\text{JER}}) \cdot \sqrt{\max\left(\text{SF}^2 - 1, 0\right)}\right) \tag{5.10}$$

In practice, a so-called *hybrid method* is applied (as recommended by JetMET in Ref. [135]), according to whether the matched gen-level jet exists or not. This method uses the best from both approaches to smear the jets, in principle, the scaling method should be enough, but if one doesn't have a matching candidate, the jets must be smeared using the stochastic method. Given the matching condition to apply the scaling method, it should be made clear that the jets in the tails of the resolution are to be smeared using the stochastic method.

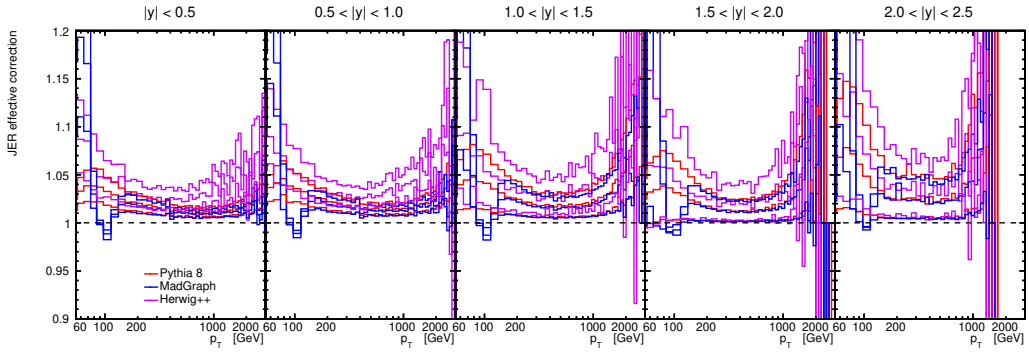


Figure 5.10: Effect of JER smearing in simulation, i.e., the ratio of the detector-level inclusive jet p_T distribution after and before applying the corrections. The central value is plotted for each generator and the JER uncertainty envelope as well.

The effect on the spectrum can be seen in Fig. 5.10, where all simulations give similar results. For PYTHIA8 and MADGRAPH samples (for $p_{\text{T}} > 100\,\text{GeV}$), despite the statistical fluctuations, the uncertainties are $\sim 1\%$ in the most central region ($|\eta| < 0.5$) and increase up to $\sim 3\%$ which could be related to the gap between the central (up to $|\eta|1.0$) and the forward tracker. Whereas for the HERWIG++ flat sample the uncertainty is twice as significant as in PYTHIA8 and MADGRAPH, which could be related to the derivation of the scale factors based on PYTHIA8 (UE tune is the same as for MADGRAPH since it comes from PYTHIA8 see Sec. 5.1).

5.4 Trigger strategy

The *exclusive division method* [136] is used in the triggers combination such that the total cross section is the sum of the cross sections obtained from each trigger:

$$\sigma_{\text{all triggers}} = \sum_{\text{all triggers}} \sigma_{\text{trigger}} \tag{5.11}$$

In this method, each trigger is associated with a different region of the phase space. Since the triggers are defined in terms of p_T of the leading jet of the event (single-jet triggers), the phase space is divided as a function of the leading-jet p_T. However, all subleading jets may

have lower transverse momenta.[9]

The different single-jet triggers are technically denoted as `HLT_JetX_vY`, where X (Y) stands for the p_T threshold in GeV (version of the trigger). In the context of this thesis triggers with X equal to $40, 60, 80, 140, 200, 260, 320, 400$, and $450\,\text{GeV}$ are used. All the used triggers are prescaled (meaning that only the n-th triggered event is stored) except for the last one with $450\,\text{GeV}$ threshold, the prescale value is then used in the event normalization depending on the selected fired trigger. Since from the trigger system a decision is needed very fast, the trigger system has a high-speed algorithm of reconstruction, not so precise as the PF reconstruction; therefore, p_T^{HLT}, corresponding to the HLT reconstruction, and p_T^{PF}, corresponding to the PF reconstruction need to be distinguished.

After this distinction being made, the interval of p_T^{PF} (*e.g.* the phase space) corresponding to each trigger has to be defined. In general, the HLT algorithm is faster. However, the jet p_T reconstruction is less precise, i.e., the resolution is worse. Consequently, for the requirement on p_T^{PF} needs to higher to ensure that the "smeared" p_T^{HLT} is still above the HLT p_T threshold for the given trigger. Such p_T^{PF} for which the trigger is fully efficient is further referred as *turn-on* point T.

To summarize, the trigger is defined (and fired) as a function of $p_T^{\text{HLT}} > Y$ (where the "HLT" indicates that the jet has been reconstructed with fast reconstruction), but is only efficient for jet $p_T^{\text{PF}} > T$ (where PF indicates that the jet has been with reconstructed with the particle flow algorithm, offline and more time-consuming). We will describe the determination of T using different methods in the following paragraphs.

5.4.1 Trigger efficiency determination

Three methods exist to determine the trigger efficiency: the *reference trigger* method, the *trigger emulation* method, and the *Tag & Probe* method. These methods will be disused here, together with the standard method used to measure the trigger efficiency.

Reference trigger method

This method is the easiest in terms of methodology. Given a trigger of a certain p_T threshold known to be fully efficient, one tests another trigger of a higher p_T threshold. The efficiency is therefore obtained according to the following formula:

$$\epsilon = \frac{N(\text{test fired}|\text{ref fired})}{N(\text{ref fired})} \tag{5.12}$$

This method has the drawback of possibly very low statistics of the numerator, especially for the trigger of lowest p_T where a minimum-bias or zero-bias trigger should be used as a reference. The low amount of the test trigger and ref trigger overlap is also often related to the trigger prescales. Therefore, the method has not been used in practice but is only mentioned to motivate the second method.

[9] Alternatively, the phase space can be divided according to the different triggers not only for the leading jet but for all jets.

Trigger emulation method

The second method is an improved version of the reference trigger method, with the difference that the test trigger is emulated instead of directly checked if it has fired. Indeed, in the reference trigger method in the case of the single jet triggers, the statistics are limited due to the prescale of the triggers. By reproducing the conditions in which the test trigger would have fired in case of no prescale was applied, one gets higher statistics.

$$\epsilon = \frac{N(\text{test emulated}|\text{ref fired})}{N(\text{ref fired})} \tag{5.13}$$

However, the turn-on point of the trigger of lowest p_T threshold (`HLT_PFJet_40`) is computed by using the Tag & Probe method.

Tag & Probe method.

The third method allows determining the turn-on of the trigger of lowest p_T and cross-checking the result obtained from the emulation method. The principle of the Tag & Probe method is not restricted to the determination of the trigger thresholds: it is a general method to determine the efficiency of reconstruction of a given object from situations where two such objects are expected in an event. In the present case, it consists in using events with a dijet final state and checking when only one or both should have fired the trigger

First, PF jets (from L1 reconstruction) are matched jets with HLT objects and dijet topologies are defined:

- The dijet final state is defined such that
  1. both leading jets are back-to-back: $\Delta\phi_{1,2} > 2.4$;
  2. and all other jets have significantly lower p_T: $p_{Ti} < 0.3 \times \frac{p_{T1}+p_{T2}}{2} \; \forall \; i > 2$.
- The matching between PF and HLT jets is defined in $\Delta R < 0.5$.

The values of the parameters are motivated from the matching used in the tag and probe method of Sec. 5.3.1 (notice here $\alpha = 0.3$) but with a looser ΔR matching since these jets are still not corrected for JES or JER effects. Then the efficiency is computed as follows:

$$\epsilon = \frac{N(\text{probe}|\text{tag})}{N(\text{tag})} \tag{5.14}$$

where the *probe* jet defines whether the event has fired and the *tag* jet tests the trigger. This method will be used to determine the turn-on point of the first trigger (`HLT_PFJet_40`).

In Fig. 5.11, three different ways to obtain the efficiencies are shown:

1. Using the emulation method as default and the Tag & Probe for the first trigger (continuous black). This is the standard approach we use for extracting the turn-on points.
2. Using the Tag & Probe method for each trigger turn-on (dashed blue).
3. By taking another reference trigger to determine the efficiency of each trigger (dashed red), *e.g.* trigger 60 can be used instead of trigger 80 to assess the efficiency of trigger 140.

Fig. 5.11 shows the consistency between the tested methods with fluctuations smaller than 0.5% among them. The edges of each interval (min p_T^{leading}) or turn on points, summarized in Tab. 5.3, are determined (by using the standard method) if the trigger efficiency is bigger than 99.5% in all the rapidity bins. The only exception is for the lowest trigger, in the first p_{T} bin in the $1.0 < |y| < 1.5$ region, where the efficiency falls to about 99.35% ($\sim 0.15\%$ below). Still, since the statistical uncertainty is significant in this particular bin, we take 99.5% efficiency for this trigger. This particular issue will not affect our measurement since we select the event if the leading jet p_{T} is above 200 GeV, and these triggers are based on the leading jet p_{T}.

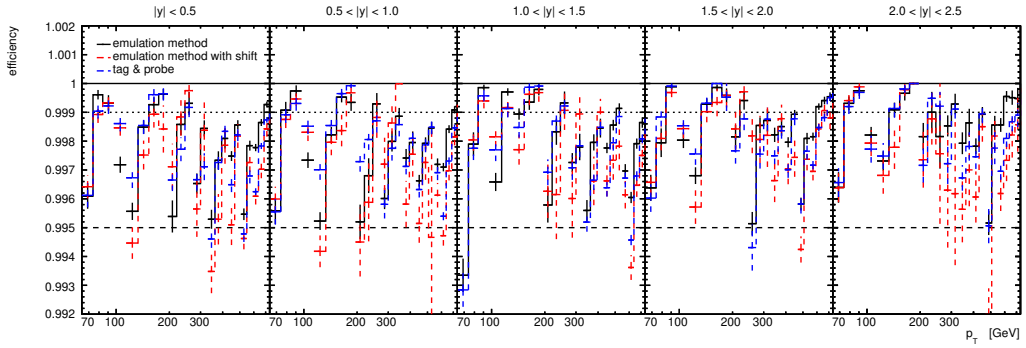


Figure 5.11: Measurement of trigger efficiency for AK4 jets with various methods. In the x-axis we have the leading jet p_{T} in the event and in the y-axis the efficiency, for five different rapidity bins.

Trigger	min p_T^{leading}	$\mathcal{L}/\text{pb}^{-1}$	#entries
HLT_PFJet_40	74	0.26703	1875509
HLT_PFJet_60	97	0.726258	2121935
HLT_PFJet_80	133	2.75889	2691293
HLT_PFJet_140	196	24.1946	4626175
HLT_PFJet_200	272	103.827	4486906
HLT_PFJet_260	362	593.764	5172414
HLT_PFJet_320	430	1772.5	8292398
HLT_PFJet_400	548	5194.33	3279941
HLT_PFJet_450	592	36321.9	44322931

Table 5.3: Summary of the triggers with the corresponding turn-on point and effective luminosities. Here the turn on points are determined by using the standard approach.

The combination of the different triggers in the p_T spectrum is shown in Fig. 5.12, this is done by considering the triggers and the turn on values in Tab. 5.3. The high p_{T} threshold triggers can contribute to lower p_{T} in the inclusive p_{T} spectra, e.g., 450 contributes down to 80 since this is a combination of single jet triggers where the event fires the trigger according to the leading p_{T} jet, but all the lower p_{T} jets are also recorded. In this analysis, the prescale approach has been used, by normalizing each LS with the corresponding prescale of the used

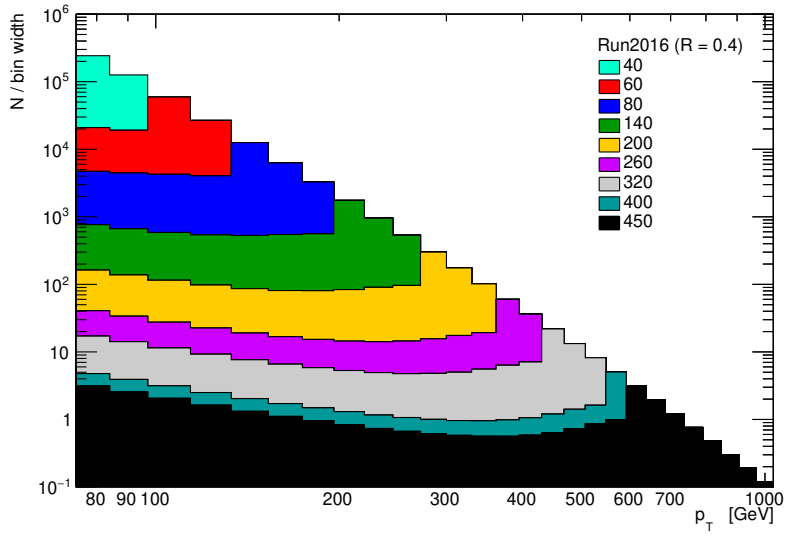


Figure 5.12: Contribution from the different p_T triggers considered in the analysis to the p_T spectrum, integrated over the rapidity, by using the turn on points from the standard method summarized in Tab. 5.3.

trigger, in consequence with the PU profile reweighting method as discussed in Sec. 5.2.4.

5.5 Additional corrections

In this section, other corrections are discussed, showing their effects after the application of the jet energy corrections, proper normalization, and pileup reweighting. However, in practice, the hotzones and MET filter corrections are applied at the very beginning to raw data and MC samples, right before the application of jet energy corrections, for ensuring a better quality of the data. The prefiring correction is instead applied at the end to the MC samples to mimic the prefiring effect in the data.

5.5.1 Missing transverse energy

The *missing transverse energy* (MET) corresponds to the momentum vector imbalance in the perpendicular plane (xy) to the beam axis (z):

$$\vec{\not{E}}_{\mathrm{T}}^{\mathrm{raw}} = -\sum_{\mathrm{i}\in\mathrm{all}} \vec{p}_{T\mathrm{i}} \tag{5.15}$$

where i runs over the different PF reconstructed particles in the event. Non-zero MET can be explained either from detector effects: since some objects can be partly or entirely missed in the event reconstruction; or from physics effects: since some particles are hardly detectable, typically neutrinos coming from the decay of a $W(Z)$ boson.

The raw particle flow MET (PFMET) is systematically different from *true* MET, i.e., the transverse momentum carried by invisible particles, for many reasons, including the non-compensating nature of the calorimeters and detector misalignment. In order to have a better estimate of the *true* MET, *corrections* must be applied.

The raw MET can be separated into two disjoint sets: either clustered as jets or unclustered:

$$\vec{\not{E}}_{\mathrm{T}}^{\mathrm{raw}} = -\sum_{\mathrm{i}\in\mathrm{jets}} \vec{p}_{T\mathrm{i}}^{\,\mathrm{raw}} - \sum_{\mathrm{i}\in\mathrm{uncl}} \vec{p}_{T\mathrm{i}} \tag{5.16}$$

In this analysis, we use the Type-I corrected PFMET (most current recommendation from JetMET group for jet analyses) [137]. This correction consists of propagating the jet energy corrections to the measured MET. The Type-I correction replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC are applied. The Type-I correction replaces the jet raw transverse momentum $\vec{p}_T^{\,\mathrm{raw}}$ with the corrected jet transverse momentum $\vec{p}_T^{\,\mathrm{JEC}}$, it can be written as:

$$\vec{C}_{\mathrm{T}}^{\,\mathrm{Type\text{-}I}} = \sum_{\mathrm{i}\in\mathrm{jets}} \vec{p}_{T\mathrm{i}}^{\,\mathrm{raw}} - \sum_{\mathrm{i}\in\mathrm{jets}} \vec{p}_{T\mathrm{i}}^{\,JEC} \tag{5.17}$$

This Type-I correction factor is a vector term that can be added to the raw MET to produce the corrected Type-I PFMET, as follows:

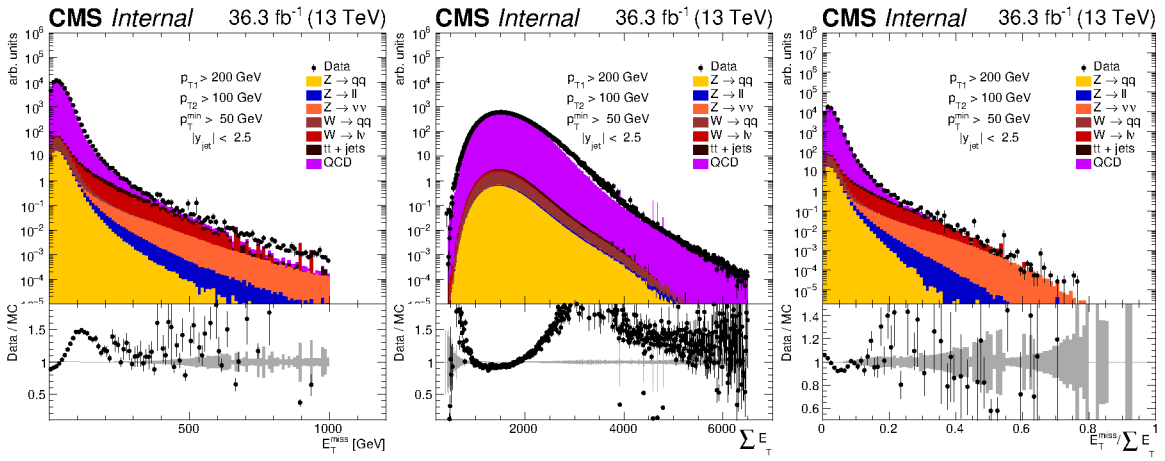
$$\vec{\not{E}}_{\mathrm{T}}^{\mathrm{Type\text{-}I}} = \vec{\not{E}}_{\mathrm{T}}^{\mathrm{raw}} + \vec{C}_{\mathrm{T}}^{\mathrm{Type\text{-}I}} \tag{5.18}$$

$$\vec{\not{E}}_{\mathrm{T}}^{\mathrm{Type\text{-}I}} = -\sum_{\mathrm{i}\in\mathrm{jets}} \vec{p}_{T\mathrm{i}}^{\,JEC} - \sum_{\mathrm{i}\in\mathrm{uncl}} \vec{p}_{T\mathrm{i}} \tag{5.19}$$

Finally, the *missing transverse energy* (MET) used in the analysis is the magnitude of the Type-I corrected PFMET:

$$E_T^{\mathrm{miss}} = |\, \vec{\not{E}}_{\mathrm{T}}^{\mathrm{Type\text{-}I}} \,| \tag{5.20}$$

For the measurements described in this thesis, a cut is applied to the MET fraction to reduce the background (see Sec. 4.3). The MET fraction is defined as the ratio between the E_T^{miss} and the sum of the transverse energy of each jet and unclustered particles in the



(a) Missing transverse energy. (b) Transverse energy sum. (c) MET fraction.

Figure 5.13: Control distributions for E_T^{miss} (a), $\sum E_T$ (b), and MET fraction (c).

event($\sum E_T$), where the $\sum E_T$ is calculated using the calibrated jets p_{T} as follows:

$$\sum E_T = \sum_{\mathrm{i}\in\mathrm{jets}} |\vec{p}_{T\mathrm{i}}^{JEC}| + \sum_{\mathrm{i}\in\mathrm{uncl}} |\vec{p}_{T\mathrm{i}}| \tag{5.21}$$

Control distributions of the missing transverse energy, the transverse energy sum, and the MET fraction are shown in Fig. 5.13 for dijet events (with the selection on Sec. 4.3). These control distributions show a reasonably good agreement between the data and simulated samples, especially for the MET fraction. The MET fraction is used to remove the background from $W/Z + jets$ observed in the tale of the distribution. A more detailed discussion can be found in Sec. 4.5.

MET filters correction

Large values of MET can be caused by interesting physics like invisible particles but in many other cases can be related to detector noise, cosmic rays, or beam-halo particles, which can affect the measurement. The MET coming from by uninteresting causes is called *fake* MET. This subsection discusses the application of selections based on MET or the so-called MET filters, which are a series of cut-offs based on the MET. The goal is to exclude noisy events which can lead to a fake MET topology. based on the material from JetMET in the CMS group found in Ref. [138, 139]. The different filters for 2016 data will be described based on the material from JetMET Physics Object Group (which belongs to CMS) found in Ref. [138,139]. Only the recommended filters by JetMET were applied and will be discussed in the following:

- `goodVertices`: This filter is based on events with a huge contrast between large deposits of energy in the calorimeters and a lack of tracks. In those events, sometimes, the tracking algorithm gives up for some iterations because of many clusters (in the calorimeters). On the other hand, a hard collision didn't happen at the center of the detector but displaced outside the Pixel Tracker in many cases displaced more than 0.75 meters and detected by the forward and central calorimeters. The filter is very effective and uses the division between the good vertices tracking stumps and the HT of all jets in the event with a minimum of 10% allowed.

- `globalSuperTightHalo2016Filter`: Beam halos are machine-induced particles flying parallel to the beam at a large radius (up to 5 m) produced through beam-gas, beam-pipe interactions. This filter exploits the calorimeters information by building halo clusters candidates based on the shape of halo deposits in each subdetector (EE, EB, HB, HE). Suppose a halo candidate from the calorimeters information can be matched to a CSC flat segment in the muon chambers. The halo cluster has a pattern characteristic of halo muons. In that case, the event is beam halo tagged.

- `HBHENoiseFilter`: The HB/HE noise filter is designed to find anomalous HCAL noise not due to electronics but rather to instrumentation issues from Hybrid Photo Diodes (HPDs) and Readout Boxes(RBXs). This filter is based on suppressing the noise by using timing, pulse shape and other readout errors in the HPDs and the RBXs.

- `HBHENoiseIsoFilter`: This filter in addition to the standard `HBHENoiseIsoFilter` helps to reach a nearly perfect efficiency ($\sim 100\%$) for identifying noise from $MET >$

25 GeV. This filter uses HCAL isolation (sum of all neighboring hits per cluster energy), ECAL isolation (energy sum of ECAL hits in the same towers as cluster per cluster energy), and Track isolation (energy sum of tracks pointing to cluster hits or their neighbors per cluster energy). Then it uses dedicated cuts on those variables that quantify the isolation level on the calorimeters and the tracker and eliminates events with overestimated MET in the calorimeters.

- `EcalDeadCellTriggerPrimitiveFilter`: The ECAL dead cell filter uses primitive L1 trigger information to assess the energy loss on masked ECAL crystals corresponding to Very Front End (VFE) or Front End (FE) cards with no link to the data. These masked crystals constitute only about 1% of the ECAL composition. Nevertheless, when those masked crystals are located at VFEs or FEs, a significant amount of energy can be lost in that region leading to fake MET. In summary, a cut-based filter is constructed with the energy loss, the matching of masked ECAL cells (in ΔR) with jets, and the surrounding energy of the masked cells.

- `BadPFMuonFilter`: In both data and simulation, muons with excessive momentum have been found. This muons generated in the PF algorithm can have huge wrong momentum in very exceptional cases. Another issue was related to soft muons absorbing the energy of the very energetic HCAL tower it crosses. Both issues were addressed by constructing three filters: (1) rejecting events with muons with $p_T^{PF} \gg p_T^{reco}$, (2) isolated muons with raw calorimeter energy bigger than certain cut, (2) muons with high $p_T^{PF} > 100\,\text{GeV}$ and not being identified as a global and tracker muon at the same time, or with $|p_T^{tracker}/p_T^{global} - 1| > 0.1$.

- `eeBadScFilter`: The bad EE supercrystal filter was created to remove events with anomalously high energy found by the MET scanned in two EE crystal regions during 2012. The filter is constructed with a cut based on the total super crystal energy and the number of events with more than 1 TeV in those two defective crystals.

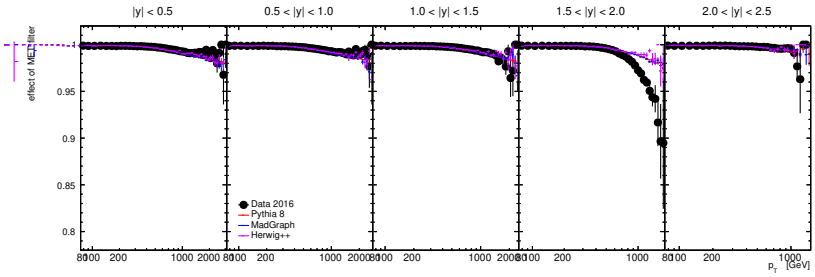


Figure 5.14: Net effect of recommended MET filters at detector level in data and simulation samples. This shows in the y axis the ratio between the inclusive jet cross-section before and after applying the MET filters and in the x axis the p_{T} bins for four rapidity bins.

The MET filters are applied both in the data & simulation[10] samples. The net effect of the recommended MET filters is shown in Fig. 5.14. As we can see, the main effect is taking place at high p_T where larger MET scenarios can contribute, but it is at percent level in most

[10] In the simulation, the MET filters are only defined at the detector level, not at the generator level.

of the absolute rapidity bins. Only in the fourth rapidity bin ($1.5 < |y| < 2.0$) a difference of about 4% coming from the `HBHENoiseFilter` was detected. As shown in Fig. 5.15, the `HBHENoiseFilter` filter effect is the same as when all the filters are applied together. The other remaining filters didn't show significant differences reflected in the net effect. In App.G the detailed study of all the MET filters can be found.

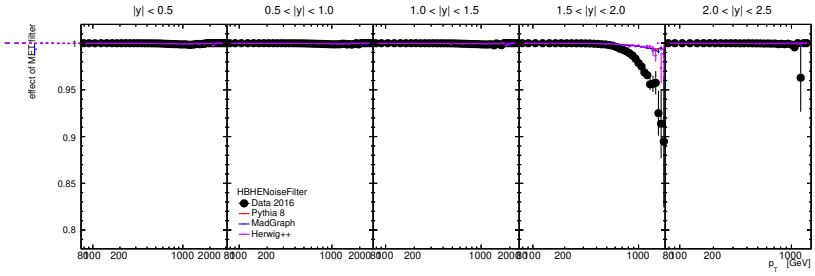


Figure 5.15: Effect of `HBHENoiseIsoFilter` at detector level in data and simulation samples. This shows in the y axis the ratio between the inclusive jet cross section before and after applying the MET filter and in the x axis the p_{T} bins for four rapidity bins.

No explicit uncertainty is considered while applying the MET filters. The effect is not very large but slightly higher in data than in simulation, which indicates that there is an extra noise not included in the simulation (this is to be solved with the MET fraction cut addressed on Sec. 4.3).

5.5.2 Hot zones correction

Since certain detector regions were damaged by radiation, one has to consider excluding them from the measurement. The situation is analogous to the application of the MET filters: Typically, these zones are not well reproduced in the detector-level simulation samples, so some small problematic regions of the phase space are excluded from the detector level selection. However, at the truth level, the simple phase-space definition is kept(without excluding anything), i.e., effectively, the measurement is extrapolated to the whole phase space. As these excluded zones are relatively small, this extrapolation does not introduce any substantial uncertainty.

Constructing the maps

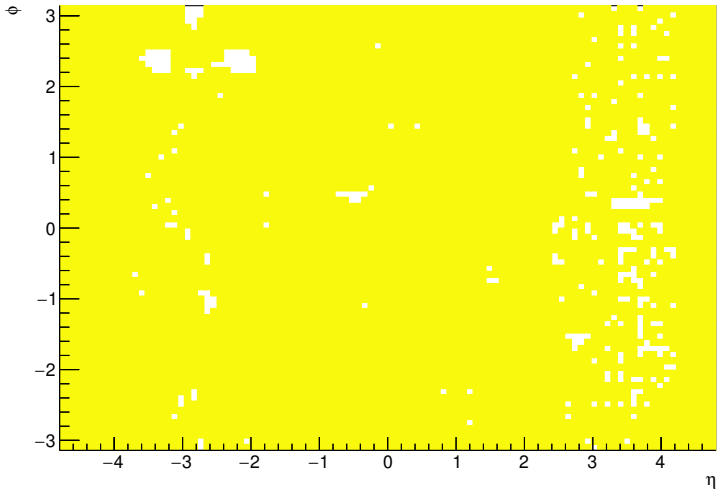
Most of the imperfections in the real calorimeters are simulated; as long as this is the case, the unfolding will correct for these effects. Some hot zones are, however, not properly simulated; therefore, certain regions are removed both in data and simulation at the detector-level. *Hots maps* describe regions where excess or a deficit of jet is observed in data. The maps are provided by the Helsinki group[11] for each era of data taking separately.

Treatment of hot zones

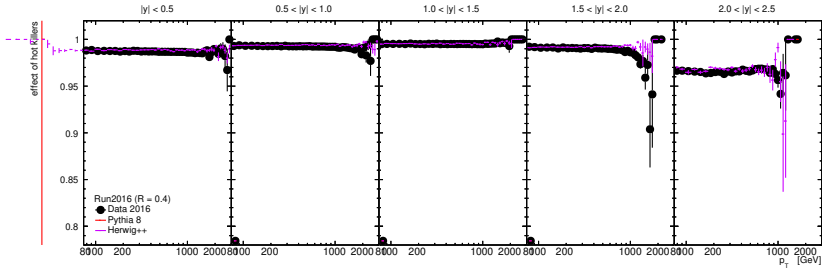
In the current analysis, the conservative approach is used: as soon as a region is *hot* even only in one era, then jets in this region are removed not only for this era but for the whole

[11] `http://hsiikone.web.cern.ch/hsiikone/hotcoldjets/16-LegacyReReco/V1/B/slides.pdf`

year. First, the maps need to be converted to this conservative map, then the reconstructed jet in the hot zones described by the conservative map is removed. The hot map applied in the data, and the simulation is shown in Fig. 5.16a, and the effect on the spectra is shown in Fig. 5.16b. The effect is relatively small and similar in data and Monte Carlo simulation, except in $1.5 < |y| < 2.0$, where it seems that the simulation is missing some effect. The most significant impact is seen in the most forward region ($\sim 4\%$). It is interesting to note that the same region of the detector ($1.5 < |y| < 2.0$), for high p_{T} jets ($p_{\mathrm{T}} > 1000\,\mathrm{GeV}$) showed up differences in data and simulation when applying the MET filters (see Sect. 5.5.1) which could be related to events with high p_{T} jets happening far from the collision point where the pixel information is not available and non-detector material is located (a gap between the forward and central tracker as shown in Fig. 3.9) to correctly identify a PV. Therefore, this could lead to problems in the simulation samples to mimic such situations that happen in data. Nevertheless, this effect is found in a region where the inclusive jet cross-section is several orders of magnitude smaller than in the very central region ($|y| < 0.1$).



(a) Map for hot zones in 2016. Here the hot zones are represented as the white spots, while the zones used in the measurement are represented in yellow. This map was constructed for HLT jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.



(b) Effect of excluding the hot zones on the maps at detector level in data and simulation. This illustrates the inclusive jet cross-section ratio before and after applying the hot killers in Data and simulation.

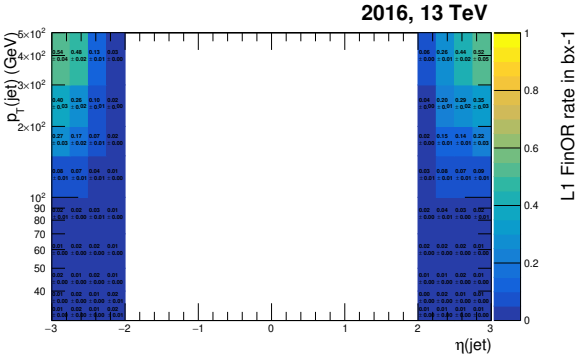
Figure 5.16: Illustration of the treatment of hot regions of the detector in data and simulation.

5.5.3 Prefiring correction

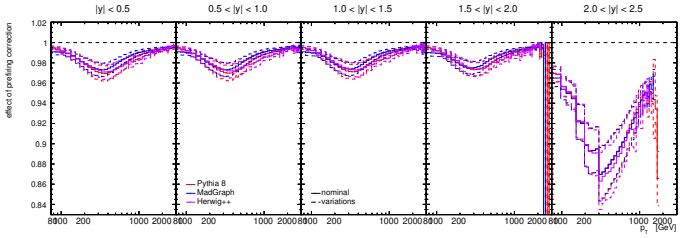
From the end of 2016 (and mainly in 2017), the increasing pileup had a side effect, called *prefiring issue*. Specific jets (and photons) in the forward region ($2.0 < |\eta| < 3.0$) were wrongly considered by the trigger at L1 as belonging to the previous bunch crossing, conflicting with a rule that *no two consecutive bunch crossings can both fire the trigger*. This was happening only at the trigger level, and it is sufficient to apply a weight to compensate for the loss of efficiency.

A centrally produced map, shown in Fig. 5.17a, is provided and used here directly. The probability of prefiring is multiplied by the detector level jets weights directly from the maps as a function of p_T and η of the jets. As recommended, an uncertainty is associated with the procedure, taking the maximum between 20% of the probability of prefiring and the statistical uncertainty of the probability of prefiring.

The effect on the p_{T} spectrum can be seen in Fig. 5.17b. Steps can be seen in the spectrum, corresponding to bin edges in the provided corrections. The fact that the impact of the prefiring corrections gets smaller at high p_T is related to the correlation between the rapidity and the transverse momentum of the jets. Indeed, high-p_T jets are more likely to be close to $|y| = 2.0$ (where the correction is more minor) than to $|y| = 2.5$ (where it is larger).



(a) Probability for a jet to prefire, as provided centrally at CMS.



(b) Effect on p_T spectrum in bins of rapidity.

Figure 5.17: Effect of correction on prefiring issue. Here, the inclusive jet cross-section ratio is shown before and after the correction is applied for data and simulation.

5.6 Comparison data to simulation

After understanding all detector effects in data and simulation, a proper MC to data comparison can be performed. The corrections are applied following a modular workflow, factorizing each correction, as shown in Fig. 5.18. In this section, the measured jet multiplicity and jet p_{T} spectra of the four leading jets (observable's defined in Sec. 4.4) in dijet events are compared to MC predictions at Leading Order (LO) from PYTHIA, HERWIG, and MADGRAPH at detector level. In addition, the measured azimuthal separation between the two leading jets in bins of the leading jet p_{T} is compared to the same LO predictions.

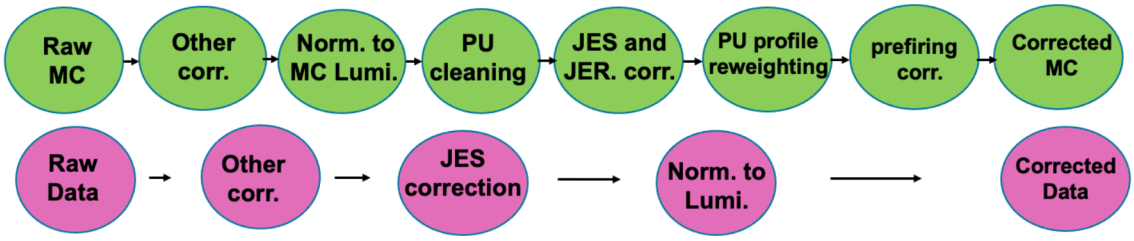


Figure 5.18: Sketch summarizing all the applied corrections to data and simulation samples. Here **Other corr.** refers to MET filters and hot zones corrections. In the MC normalization step (**Norm. to MC Lumi.**) the MC cross section is used to normalize the sample, while the data normalization (**Norm. to Lumi**) uses the measured luminosity for each trigger.

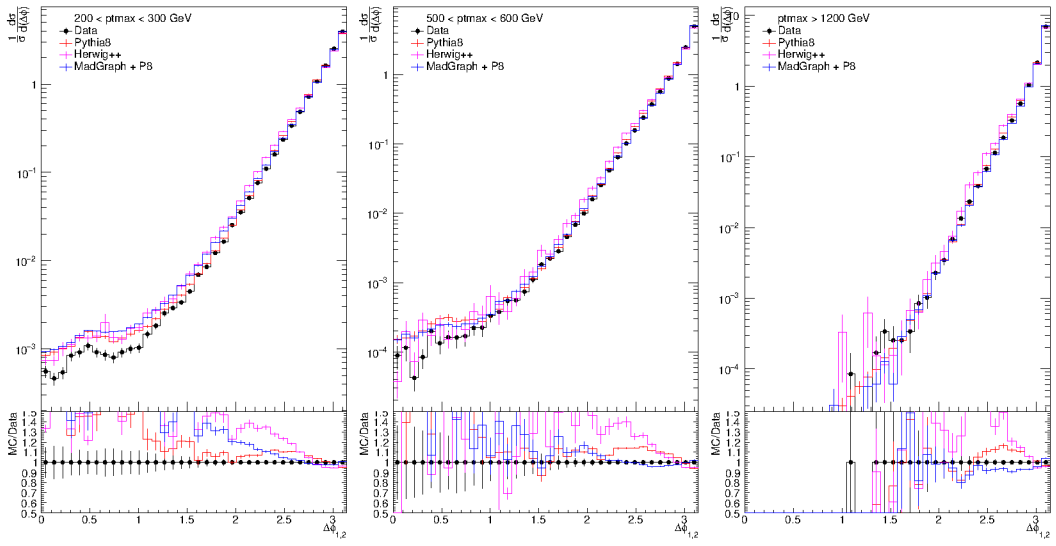


Figure 5.19: Normalized dijet cross section, differential in $\Delta\phi$ between the first two leading jets ($\Delta\phi_{1,2}$), in bins of leading jet p_T (ptmax) (following the selection presented in Sec. 4.3) for the full $\Delta\phi_{1,2}$ region, at detector level. Here the measured data is compared with MADGRAPH, HERWIG++ and PYTHIA8 at detector level.

In Fig. 5.19 the normalized dijet cross-section, differential in the azimuthal angle between the two leading jets,$\Delta\phi_{1,2}$, from MADGRAPH, PYTHIA8 and HERWIG++ LO predictions are compared to data at detector level. In general MADGRAPH and PYTHIA8 describe better the data than HERWIG++ especially in the region from 2.0 to 2.5. Since this is a normalized

cross-section, we can conclude the shape differences between predictions and data. It is important also to notice that in the region from 0 to 1.5 ($\Delta\phi_{1,2} < \pi/2$), the agreement is improving with high leading jet p_{T}, since this region is significantly affected by high MET fraction events (what we call non-QCD events) as observed in Sec. 4.3. We compare pure QCD jet events (no vector boson/top pair plus jets) from the MC simulations.

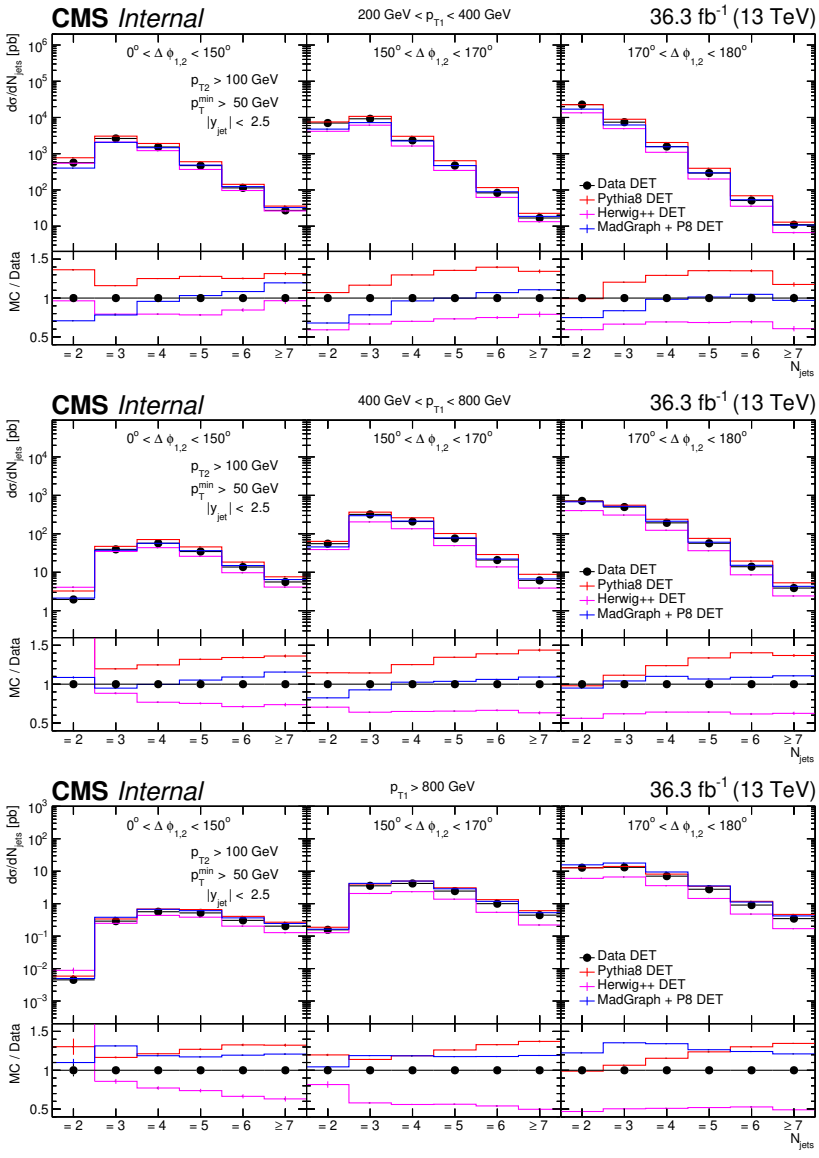
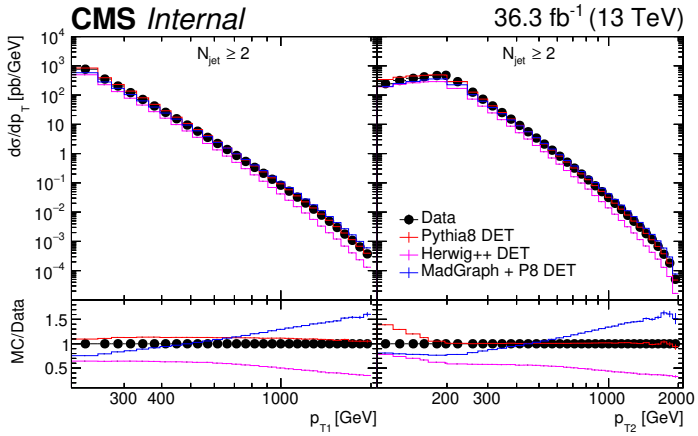


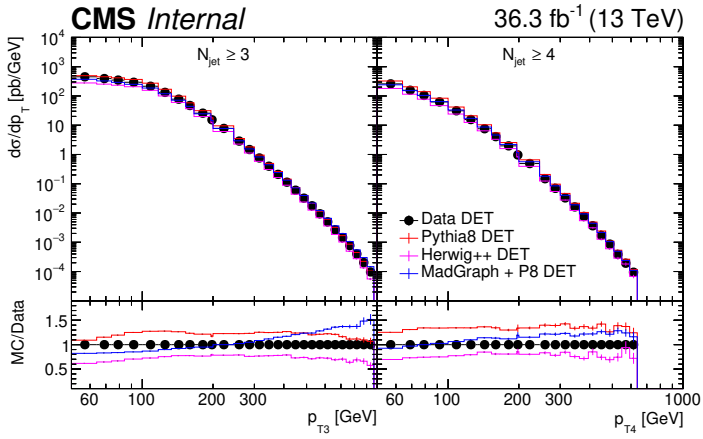
Figure 5.20: Jet multiplicity(N_{jets}) in dijet events in bins of leading jet p_T (p_{T1}) and $\Delta\phi_{1,2}$ at detector level, in the tracker acceptance $|y| < 2.5$. The predictions of the MC are compared to data at detector level.

In Fig. 5.20 we can see the jet multiplicity for different p_{T1} and $\Delta\phi_{1,2}$ bins at detector level. For the first two p_{T1} bins MadGraph describes rather well the data due to the higher LO matrix element calculation (up to 4 partons in the final state). The pythia8 prediction can describe the two jet ($N_{jet} = 2$) events in the back-to-back region ($170^o < \Delta\phi_{1,2} < 180^o$) and intermediate $\Delta\phi_{1,2}$ region ($150^o < \Delta\phi_{1,2} < 170^o$) rather well but fails for higher multiplicities

$N_{jet} > 2$ (in addition to the two leading jets) since these multiplicities can only be produced from the parton shower. HERWIG++ fails in the normalization but nevertheless the shape description is better than with PYTHIA8, becoming more visible in the back-to-back region where the shower plays a major role in the description of multiplicities higher than $N_{jet} = 2$.



(a) Fist two leading jets p_{T} spectrum.



(b) Third and forth leading jets p_{T} spectrum.

Figure 5.21: Transverse momentum distributions of the four leading jets in dijet events at detector level. Here we compare PYTHIA8, MADGRAPH and HERWIG++ to data.

We also studied the p_T of the individual first four jets in the event. In Fig. 5.21 the data and MC comparison is shown at detector level. For the two leading jets (Fig. 5.21a) PYTHIA8 predicts rather well the data, HERWIG++ is off in normalization but shows a similar shape to PYTHIA8, and MADGRAPH overshoots the data at high p_{T}. For the third and fourth leading jets p_{T} distributions, in Fig. 5.21b, MADGRAPH still overshoots the data at high p_{T}, and PYTHIA8 and HERWIG++ show the same shape description of the data but are off by more than 20% in normalization.

CHAPTER

6

ANALYSIS AT PARTICLE LEVEL

Contents

6.1 Unfolding procedure 76

6.1.1 Response matrix construction 77

6.1.2 Migrations in the phase space 80

6.1.3 Unfolding tests 83

6.2 Uncertainties after unfolding 89

6.2.1 Statistical uncertainties 89

6.2.2 Systematic uncertainties 89

6.2.3 Correlations 92

The measured cross sections need to be corrected for smearing and inefficiencies of the detector for comparison with particle level predictions directly. The procedure for obtaining the particle level distribution from the measured distribution is called unfolding. It relies on constructing the so-called response matrix (RM), which maps the true distribution (at particle level) onto the measured one (at detector level). The RM is constructed using the simulated MC samples: PYTHIA8 (CUETP8M1), MADGRAPH (CUETP8M1) showered with PYTHIA8, and HERWIG++ (CUETS1). Given the higher statistics and better matrix element calculation, the MADGRAPH sample will be used as the default for constructing the RM, and HERWIG++ and PYTHIA8 samples will be used to evaluate the effect using different models to build the RM. The CMS detector effects are simulated using GEANT4 [140]. In this chapter, the unfolding procedure will be presented for both jet multiplicity and jet p_{T} measurement, going from the construction of the RM until the treatment of the systematic and statistical uncertainties.

6.1 Unfolding procedure

In this analysis, multidimensional distributions are unfolded using the TUnfold (version 17.9) [141] software package. TUnfold provides all tools for treating systematic and statistical uncertainties coming from MC and data. The data are unfolded by using real Inversion and pseudo Inversion according to the observable measured:

Real Inversion

Mathematically the problem of unfolding can be formulated as follows:

$$\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{y} \tag{6.1}$$

where

- $\mathbf{x}$ is the *truth* vector at generator (particle) level, which represents the distribution that we want to determine;
- $\mathbf{y}$ is the *measurement* vector at detector level;
- $\mathbf{b}$ is the *background* vector at detector level;
- and $\mathbf{A}$ is the *probability matrix* (PM), here an element A_{ij} corresponds to the probability that a quantity with true value in bin i will be measured in bin j. This matrix is constructed from the RM by normalizing to the generator level axis.

Given $\mathbf{y}$, $\mathbf{b}$ and $\mathbf{A}$ the exact solution would be $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$. But this real inversion can lead to instabilities given the MC fluctuations (if the PM is constructed from RM of the MC), especially in the uncertainties. A good estimate of whether real matrix inversion is possible is the so-called condition number of the matrix, the ratio of the highest eigenvalue with the lowest eigenvalue. If the condition number is smaller than ten, then applying matrix inversion is affordable and gives a good result (see also the statistical committee recommendations on Ref. [142]).

Pseudo Inversion

It is formulated as least-square minimization (pseudo Inversion):

$$\chi^2 = \min_{\mathbf{x}} \left[(\mathbf{A}\mathbf{x} + \mathbf{b} - \mathbf{y})^\intercal \mathbf{V}^{-1} (\mathbf{A}\mathbf{x} + \mathbf{b} - \mathbf{y}) \right] \tag{6.2}$$

where $\mathbf{V}$ is the covariance matrix of the detector-level data describing the data statistical uncertainties with correlations. The vector $\mathbf{x}$ must be found such that χ^2 is minimal. This has the following solution [141]:

$$\mathbf{x} = (\mathbf{A}^\intercal \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^\intercal \mathbf{V}^{-1} (\mathbf{y} - \mathbf{b}) \tag{6.3}$$

which reduces to classical Moore-Penrose matrix pseudo-inversion [143] if $\mathbf{V} = \mathbf{I}$. In addition to the vector $\mathbf{x}$ also the truth-level data covariance matrix can be calculated; the explicit formula is provided in Ref. [141]. When the vectors $\mathbf{x}$ and $\mathbf{y}$ have the same size, then eq. 6.2 is equivalent to real Inversion.

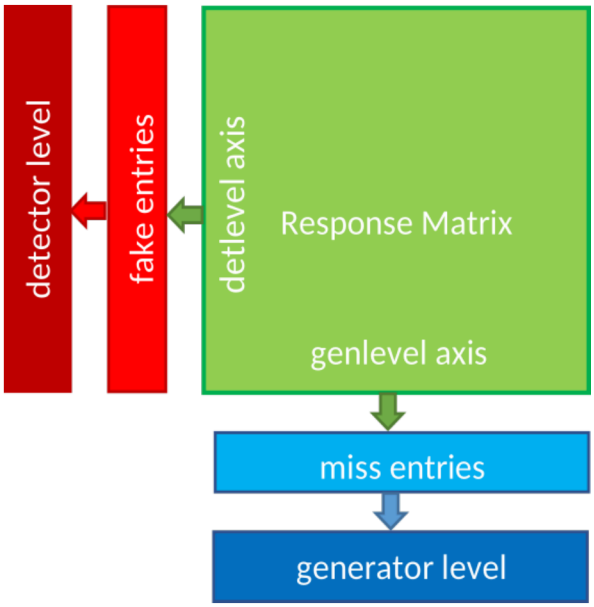


Figure 6.1: Phase space representation with **Response Matrix** (green square), **fake entries** (light red rectangle), **miss entries** (light blue rectangle), **detector level** (dark red rectangle) and **generator level** (dark blue rectangle). The rectangles represent vectors (containing histograms using `TUnfoldBinning` style) and the square is a matrix with each projection corresponding to these vectors as `x` and `y-axis` .

The use of real or pseudo inversion has to do with the measured observable. To measure the four leading jets, p_{T} pseudo inversion is used since one can define the p_{T} binning to have twice the number of bins at detector level than at particle level. On the other hand, when measuring the jet multiplicity in bins of $\Delta\phi_{1,2}$ and p_{T1}, the binning can not be divided since the number of jets is an integer number and real inversion has to be used.

6.1.1 Response matrix construction

For constructing the response matrix it is needed to create a matching strategy according to the observable. This matching strategy will define whether an event enters the RM or the miss (inefficiencies) or fake (background) histograms. The RM, and miss (**miss entries**) and fake (**fake entries**) histograms, are built from the **detector level** and **generator level** entries, as shown in Fig. 6.1. The projection of the RM in the **genlevel axis** (**detlevel axis**) plus the **miss entries** (**fake entries**) give the **generator level** (**detector level**) distribution as illustrated in the figure. In the following the used matching strategies are described for each observable.

Jet multiplicity distribution ($N_j, \Delta\phi_{1,2}, p_{T1}$)

For this observable, one can define two different matching approaches for constructing the RM, miss (inefficiencies) and fake (background):

- **Loose matching:** We require a dijet system at generator and detector level, which fulfill the analysis cuts. Events that have a dijet system at generator (detector) level

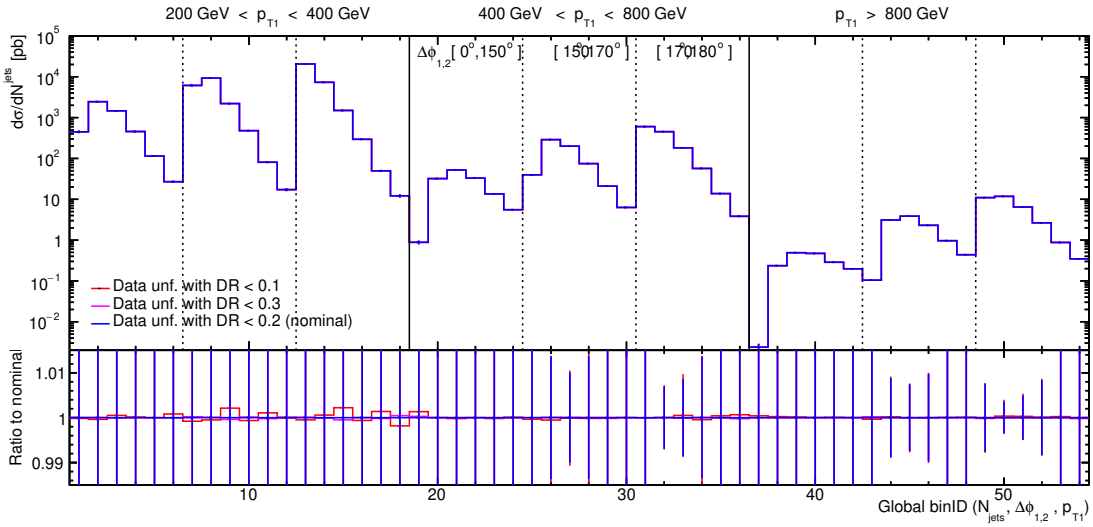


Figure 6.2: Unfolded data using MadGraph RM variating the ΔR parameter for the tight matching

but not at detector (generator) level go into missOUT (fakeOUT) histograms, which means they are outside the phase space or simply don't exist.

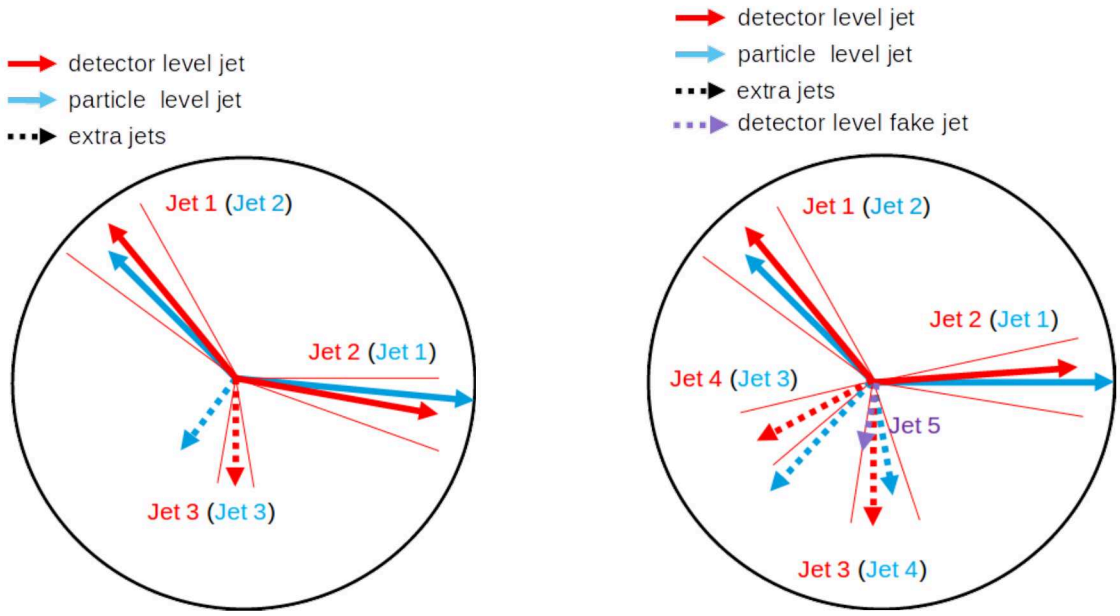
- **Tight matching:** It is required to have the dijet system at detector and generator level with a ΔR matching ($\Delta R < R/2$) between the first two leading jets. The event enters the RM only if both leading and subleading jets exist and are matched in ΔR space (swapping leading and subleading jet is allowed). Dijet events that don't match go to miss and fake histograms simultaneously. Finally, events that have a dijet system at generator (detector) level but not at detector (generator) level go into missOUT (fake-OUT) histograms as in the loose matching.

The tight matching strategy is equivalent to the loose one when using a $\Delta R \gg R/2$. In Fig. 6.2 the effect of variating the ΔR matching parameter in the unfolding is illustrated, and since it is smaller than 0.2% both matching strategies are consistent and in the following $\Delta R < 0.2$ (half of the cone size) is used as default in the unfolding together with the tight matching.

In Fig. 6.3a an example of an event fulfilling the tight matching criteria is depicted. This event would go into the RM as a matched event (in this case, into the diagonal of the matrix). This particular event also shows the swapping of jets in the dijet system due to detector smearing and the JES corrections. The jet swapping effect between leading and subleading jet is given because the jets are ordered in p_{T}, and the JES corrects the jet p_{T} but not the jet axis position. The tight matching will allow better control of the background and inefficiencies by estimating uncertainties varying their contributions by a factor based on the systematic effect of using different RM to perform the unfolding (see model uncertainty discussion on Sect.6.2), therefore it will be used as default in the unfolding.

Four leading jets p_T (p_{Ti}, n_i^{jet})

For this observable, we have defined the following matching approach (a bit different than for the jet multiplicity observable) for constructing the RM, miss(inefficiencies) and



(a) Jet multiplicity matching event.

(b) Four leading jet p_{T} matching jets.

Figure 6.3: Example of an event in the transverse plane (xy plane) considering the tight matching approach. Here the detector level (particle level or generator level) leading and subleading jets are represented as red (blue) arrows and the extra jets are represented with dashed lines (Jet 5 in Fig.6.3b is a detector level jet) as it is shown in the legend, and the thin red lines delimit the cone size of the detector level jets. The length of the arrows corresponds to the p_{T} of the jets (larger arrow represents bigger p_{T} of the jet).

fake(background):

- **Tight matching:** It is required that the dijet system at detector and generator level (following the dijet event selection cuts) exists. Then each jet is matched in ΔR space when $\Delta R < R/2$ (the first four leading jets are used) considering particle (reference jet) and detector level (matching candidates) jets, and in cases with more than one matching candidate the one with the higher p_{T} is chosen. If jets are matched, this becomes an entry in the RM. In contrast, jets that don't match, because of migration outside the phase space for the detector (generator) level, fill missOUT (fakeOUT) histograms. The remaining unmatched jets go to miss and fake histograms.

In Fig. 6.3b an example event in the transverse plane is shown. In this example, we have four jets at particle level and five jets at detector level, which means that the event weight could enter up to four times in the RM matrix. In this case, it enters three times since Jet3 at particle level doesn't match in ΔR with any of the detector level jets in the event while the other particle level jets have a matching jet at detector level. The case in which leading jets are swapped because of detector resolution effects and JES corrections (as explained in the jet multiplicity matching strategy) is also shown. A similar swapping effect is depicted for Jet3 and Jet4. The existence of the fake jet (Jet5) is shown, where Jet4 at particle level is matched in ΔR with Jet3 and Jet5 at detector level, and as said before in the tight matching definition, the jet with the highest p_{T} (Jet3) is chosen as the right matching candidate, this fake jet (Jet5) which comes from detector effects, e.g., a pileup jet not eliminated by the

corrections.

Probability Matrix

The probability matrix (PM) is constructed by normalizing the response matrix to the generator (particle) level axis. The PM will encode the probability of having a generator level event (jet) reconstructed and measured in a different (same) bin at detector level. In these measurements, the MADGRAPH sample has been used to construct the PM (HERWIG++ and PYTHIA8 probability matrices can be found in App. H).

In Fig. 6.4 the MADGRAPH probability matrix is shown for the jet multiplicity measurement. As this measurement of the jet, multiplicity is done in bins of p_{T1} and $\Delta\phi_{1,2}$ the matrix depicted in the figure represents a three-dimensional matrix $(N_j, \Delta\phi_{1,2}, p_{T1})$ as described in the caption. In the figure, migrations among p_{T1} and $\Delta\phi_{1,2}$ bins can be observed since not all the events are reconstructed at the same bin that they were generated.

In Fig. 6.5 the MADGRAPH probability matrix for the four leading jets p_{T} measurement is shown. The figure represents a two-dimensional matrix (p_{Ti}, N_i^{jet}) as described in the figure caption. There, one can identify migrations among the jets by identifying the effects described in the matching strategy (see 6.1.1) where the leading jet can be reconstructed as the second leading jet, and the third leading jet can be reconstructed as the fourth leading jet in the event.

6.1.2 Migrations in the phase space

As a result of the detector effects like smearing, inefficiencies, and limited resolution, there are migrations inside the PM (which define the phase space of the measurement), as well migrations into or out the phase space because of the edges defined in the phase space. The migrations inside the phase space are quantified by the *purity* and *stability* while the migrations outside the phase space can be quantified from *acceptance*, *background* (or fake rate) and *missrate* (or inefficiencies). These quantities are derived from the simulated samples (MC), and therefore they have been studied for all the MC samples available in this analysis.

The *purity* is defined as the fraction of events (jets) that were selected at generator level (*gen*), with a corresponding detector level event (jet) *matched*, and reconstructed in the same bin at detector level (*det*). It can be written as:

$$P_i^{MC} = \frac{N_{matched}^{MC}(E_{det}^{MC} \in bin\ i \wedge E_{gen}^{MC} \in bin\ i)}{N_{matched}^{MC}(E_{rec}^{MC} \in bin\ i)} \tag{6.4}$$

The *stability* is very similar to *purity*, its defined as the fraction of events (jets) selected at detector level (*det*), with a corresponding generator level event (jet), that comes from the same bin at generator level (*gen*). Similar to *purity* it can be written as:

$$S_i^{MC} = \frac{N_{matched}^{MC}(E_{det}^{MC} \in bin\ i \wedge E_{gen}^{MC} \in bin\ i)}{N_{matched}^{MC}(E_{gen}^{MC} \in bin\ i)} \tag{6.5}$$

The *acceptance* corresponds to the fraction of events selected at the generator level for which a corresponding detector level event is found. Using similar notation as for *purity* it can be written as:

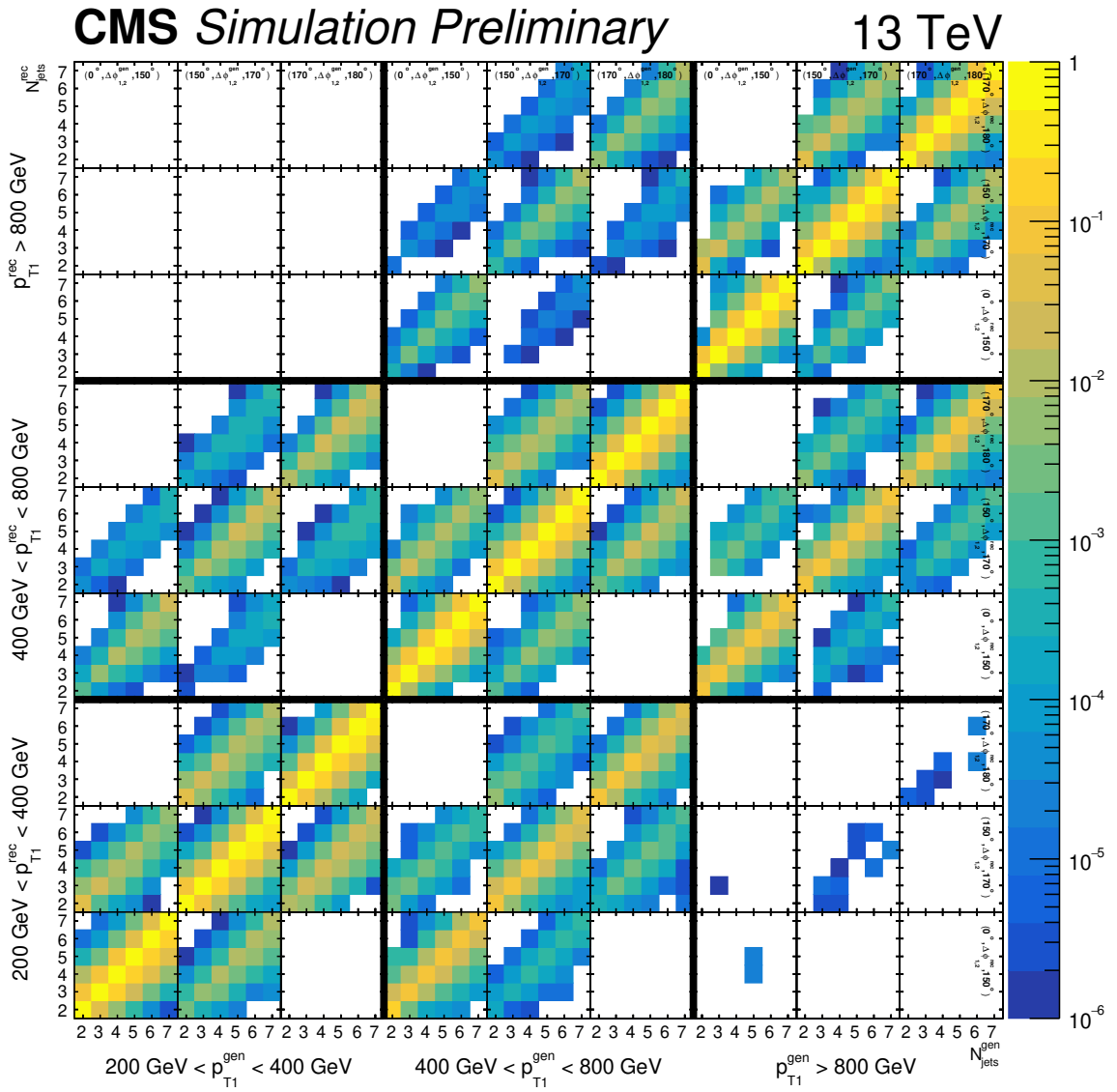


Figure 6.4: Probability matrix (condition number: 2.9) for the jet multiplicity distribution constructed with the MadGraph sample. The global 3×3 sectors (delimited by the thick black lines) corresponds to the p_{T1} bins, indicated by the labels in the x(down) and y(left) axis. Inside these, there are smaller 3×3 sectors corresponding to the $\Delta\phi_{1,2}$ bins, indicated in the uppermost row and rightmost column, the $x(y)$-axis of these $\Delta\phi_{1,2}$ cells corresponds to the jet multiplicity at particle(detector) level. The z-axis covers a range from 10^{-6} to 1 indicating the probability of migrations from the particle level bin to the correspondent detector level bin.

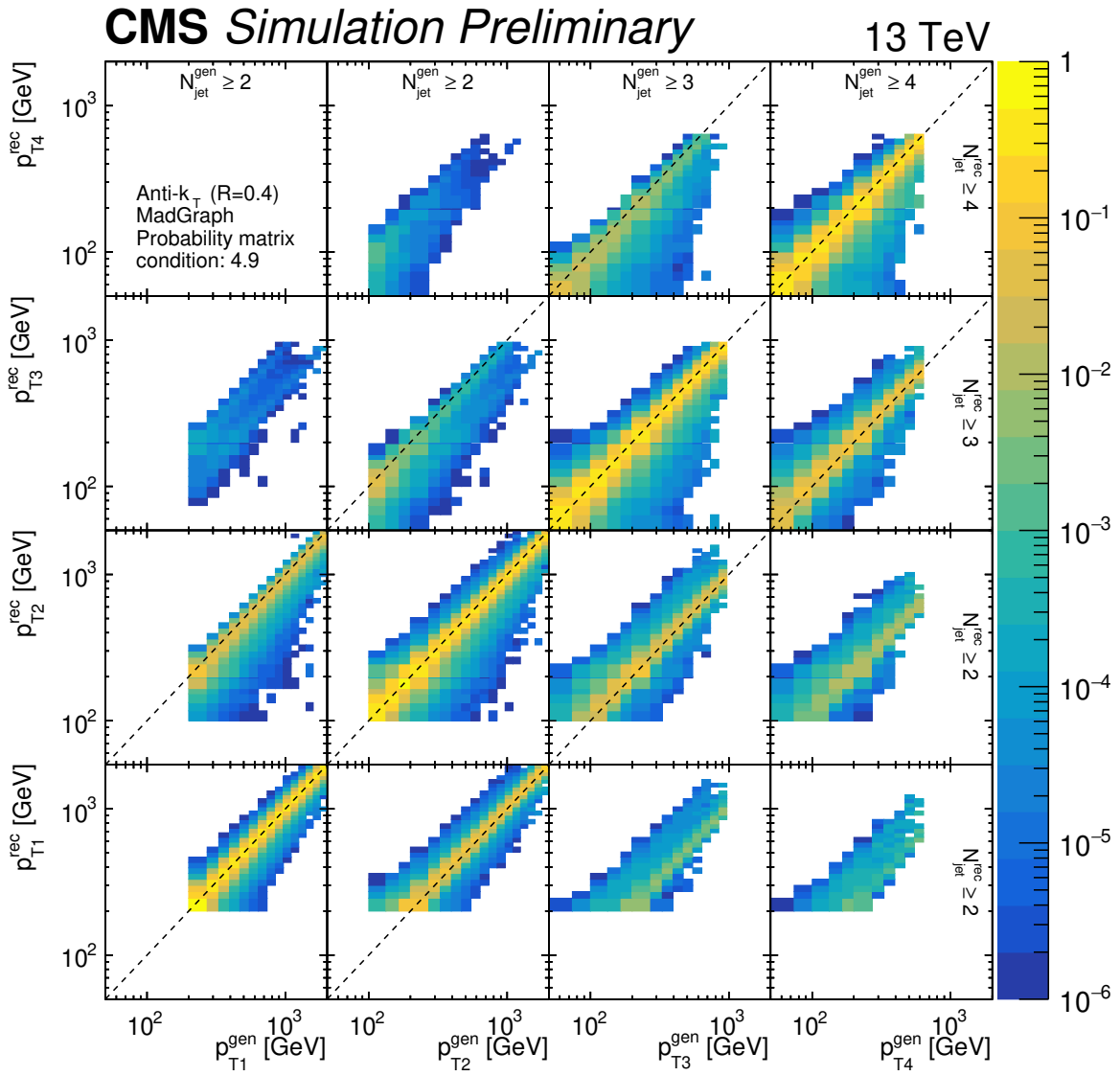


Figure 6.5: Probability matrix (condition number: 4.9) for the p_{T} of the four leading jets constructed with the MadGraph sample. The global 5×5 sectors correspond to the first four jets p_{T} distributions. The x-axis corresponds to the particle(gen) level, and the y-axis corresponds to the detector(rec) level labeled in each axis. The z-axis covers a range from 10^{-6} to 1 indicating the probability of migrations from the particle level bin to the correspondent detector level bin.

$$A_i^{MC} = \frac{N_{matched}^{MC}(E_{gen}^{MC} \in bin\ i)}{N_{all}^{MC}(E_{gen}^{MC} \in bin\ i)} \tag{6.6}$$

The *background* is the fraction of events (jets) that were measured at detector level and did not have a corresponding *gen* level event (jet). It can be written as:

$$B_i^{MC} = 1 - \frac{N_{matched}^{MC}(E_{det}^{MC} \in bin\ i)}{N_{all}^{MC}(E_{det}^{MC} \in bin\ i)} \tag{6.7}$$

Similar to *background* the *missrate* corresponds to the fraction of events at generator level that were not measured at detector level. It can be written as:

$$M_i^{MC} = 1 - \frac{N_{matched}^{MC}(E_{gen}^{MC} \in bin\ i)}{N_{all}^{MC}(E_{gen}^{MC} \in bin\ i)} \tag{6.8}$$

In Fig. 6.6 all these migration quantities are plotted for the jet multiplicity measurement for all the studied MCs. Since migrations happen due to the detector effects, and the detector simulation is the same for all samples, these migrations should be similar for the different models. This is the case for MadGraph and pythia8, but already for herwig++, we can see differences of the order of 5 to 8%, especially in the miss and fake rates, but also in purity and stability. These differences hint at the model dependence since the detector simulation is the same for all the samples. We can also notice that purity and stability are above 0.50 (50 %), which is acceptable and safe for performing the unfolding.

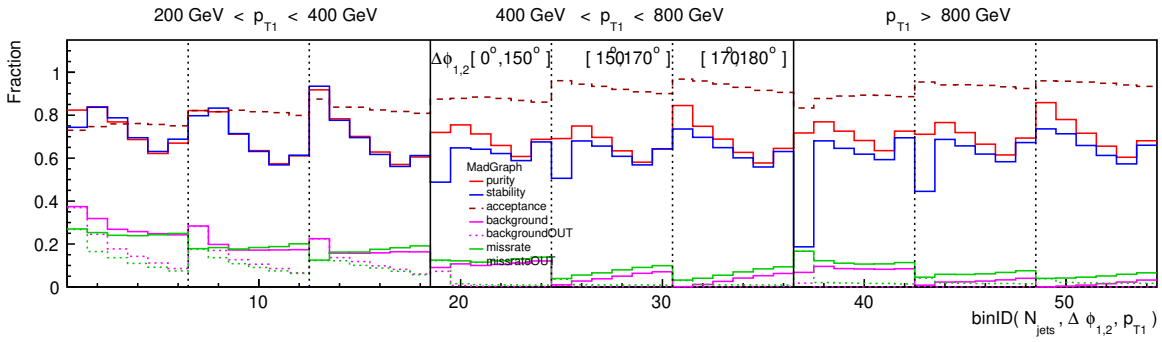
In Fig. 6.7 migrations are shown for MadGraph, herwig++ and pythia8 for the four leading jets p_{T} measurement. We observe again differences in the miss and fake rates, which could be related to the usage of different MC models in the RM construction. It is important to mention that the purity and stability are mostly larger than 0.5 (50%).

6.1.3 Unfolding tests

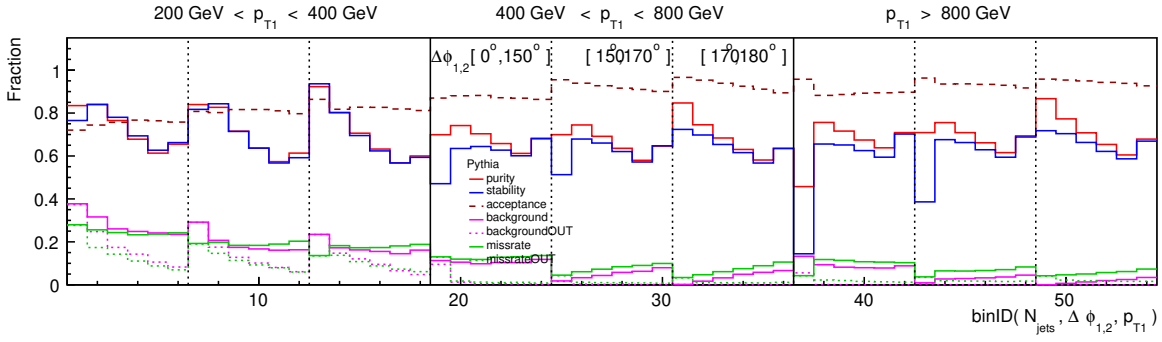
Different unfolding tests are performed to validate the procedure, determine a potential model dependence, or detect issues in the treatment of migrations and phase space definition. They are described in the following:

Closure test

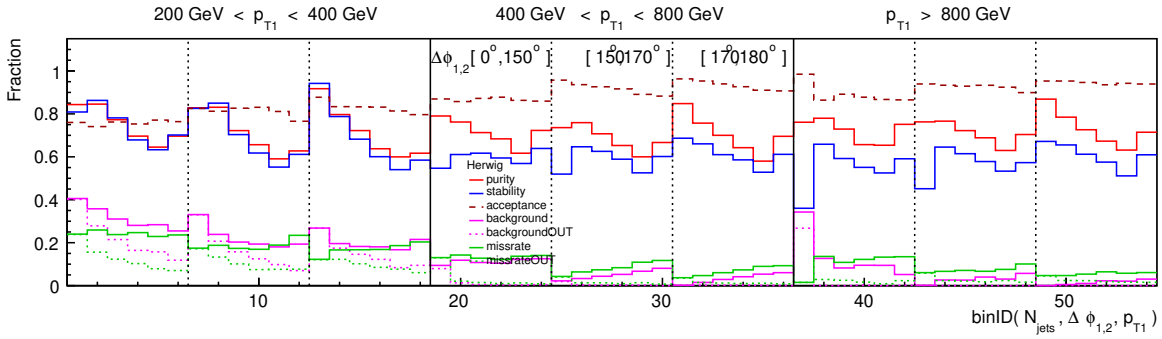
This test aims to ensure that the unfolding procedure works properly. The basic idea is to use two statistical independent samples, one as pseudo data (the smaller one) and the other as MC (to construct the RM) and perform the unfolding, or simply unfold the same sample as pseudo data with its correspondent RM. The unfolding procedure is reasonable if the correspondent unfolded distribution divided by the pseudo data (MC sample) particle level is one, at least when using the same MC sample. Finally, if different MC samples are used as pseudodata, and we unfold with a specific sample, (MadGraph in this analysis), this can give us an idea of the size of the model dependence in the studied observable.



(a) MadGraph

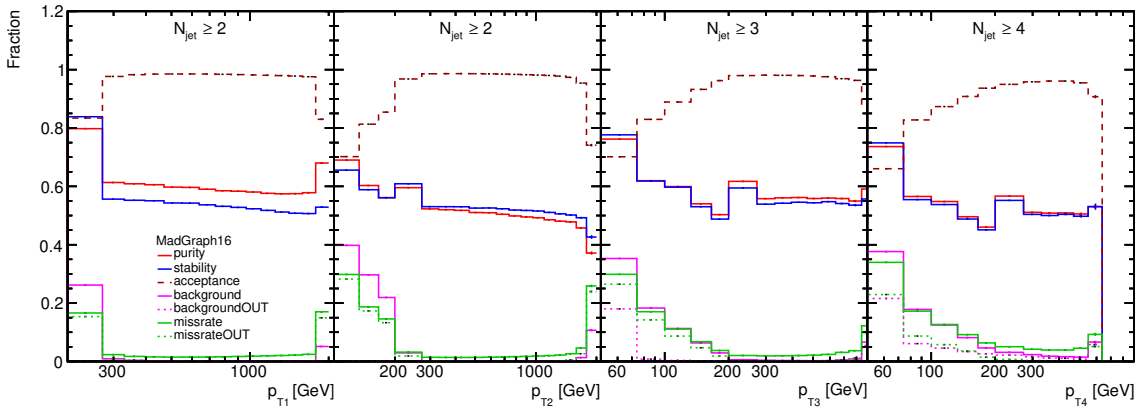


(b) Pythia8

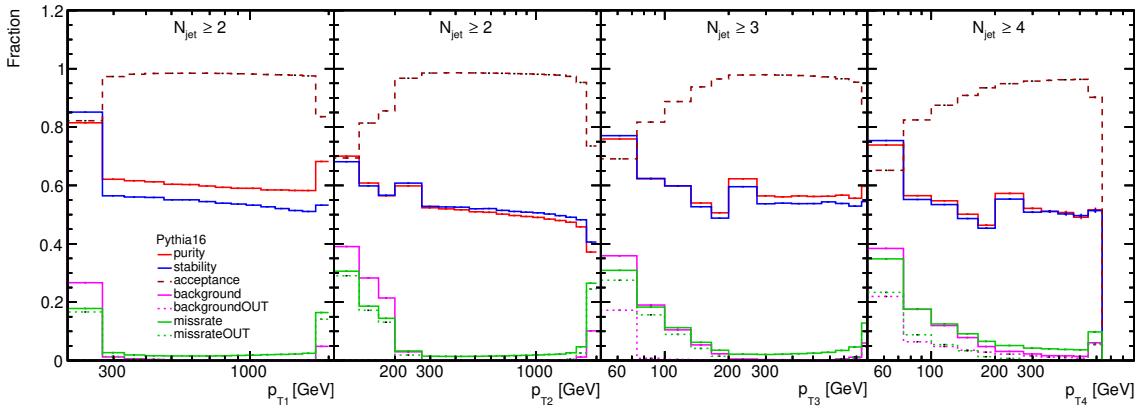


(c) Herwig++

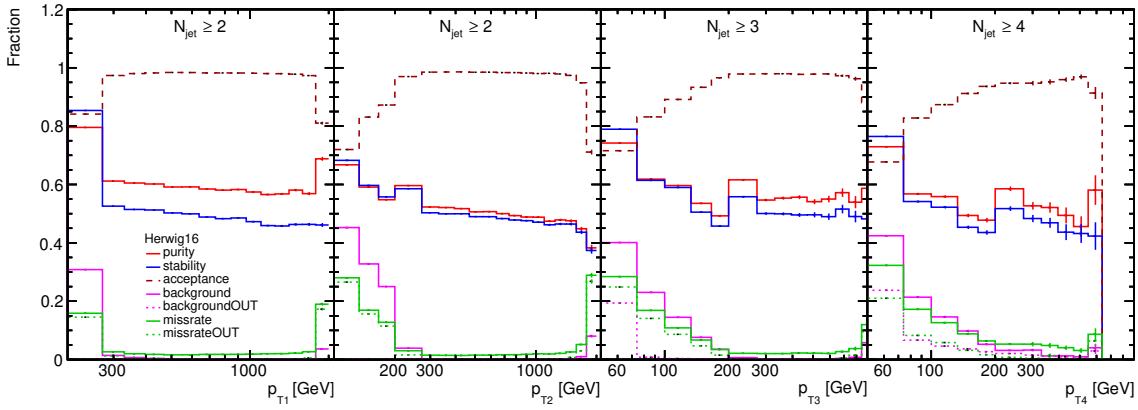
Figure 6.6: Distributions of purity, stability, acceptance background, and missrate (here backgroundOUT and missrateOUT are part of background and missrate respectively) for the jet multiplicity distribution are shown for different MC's. Here the x-axis corresponds to the binID for each jet multiplicity measured and the y-axis to the fraction (since all these quantities are defined from 0 to 1). The three $\Delta\phi_{1,2}$ bins (separated with vertical dashed lines) for each corresponding p_{T1} bin (separated with vertical solid lines) are shown.



(a) MadGraph



(b) pythia8



(c) herwig++

Figure 6.7: Distributions of purity, stability, acceptance background and missrate (here backgroundOUT and missrateOUT are a part of background and missrate respectively) for the p_T of the four leading jets are shown for different MC's : MadGraph, pythia8, and herwig++. On the y-axis the fraction and on the x-axis the corresponding p_T for each jet is shown.

Bottom line test

The ratio of MC to data before (detector level) and after unfolding (particle level) is performed and then compared. This ratio is expected to be the same or at least very similar since, in the ratio, the detector effects are expected to cancel. Also, an essential sanity check (recommended by the CMS statistics committee) is to compute the χ^2 of agreement between data($\mathbf{y}$) and MC($\mathbf{y}'$) distributions at detector (particle) level, defined as:

$$\chi^2 = (\mathbf{y} - \mathbf{y}')^{\intercal}\, \mathbf{V}_y^{-1}\, (\mathbf{y} - \mathbf{y}') \tag{6.9}$$

Here $\mathbf{V}_y^{-1}$ corresponds to the covariance matrix of the data uncertainties at detector(particle) level. Notice that the limited MC statistical uncertainty contributes to the statistical uncertainty at particle level after unfolding. Following the recommendations from the statistics committee, the χ^2 at particle level(data + MC stat.) must not exceed the detector level value for the test to be successful.

Folding and backfolding tests

The folding(backfolding) test consists in using the PM, the miss, and fake entries to fold(backfold) the detector level distribution from the generator level (unfolded data) distribution. The success of the folding test ensures the right construction of the PM matrix and the miss and fake entries. The backfolding test is important to check the unfolding procedure. Introducing the inefficiencies (miss entries $\mathbf{m}$) in eq. (6.1), we can formulate the folding (backfolding) test as follows :

$$\mathbf{A}(\mathbf{x}^{\mathbf{gen(unf)}} - \mathbf{m}) + \mathbf{b} = \mathbf{y}^{\mathbf{folded(backfolded)}} \tag{6.10}$$

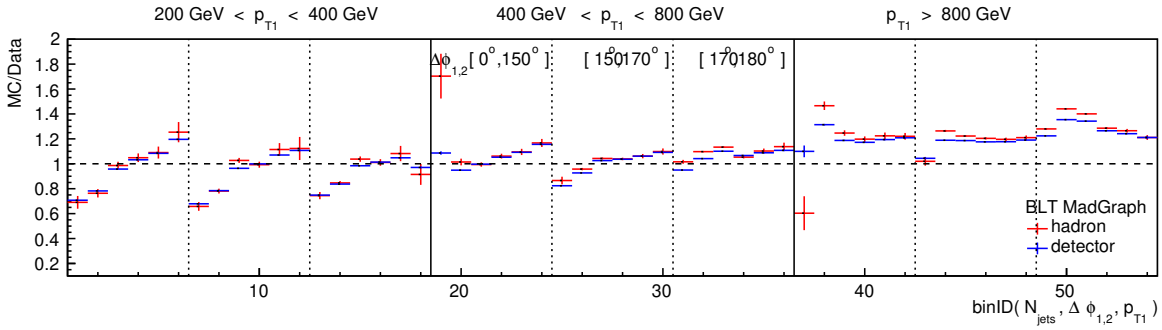
In Fig. 6.8 all unfolding tests described above are shown for the jet multiplicity observable. The tests were performed with MadGraph as default MC. herwig++ and pythia8 MCs were used in the CT as pseudodata, giving a hint on the approximate size of the model uncertainty. In Table. 6.1 the corresponding bottom line test χ^2 values are summarized, and

Table 6.1: Bottom line test χ^2 values for the jet multiplicity distribution ($N_j, \Delta\phi_{1,2}, p_{T1}$) unfolded with real inversion using herwig++, pythia8 and MadGraph response matrices to unfold the data.

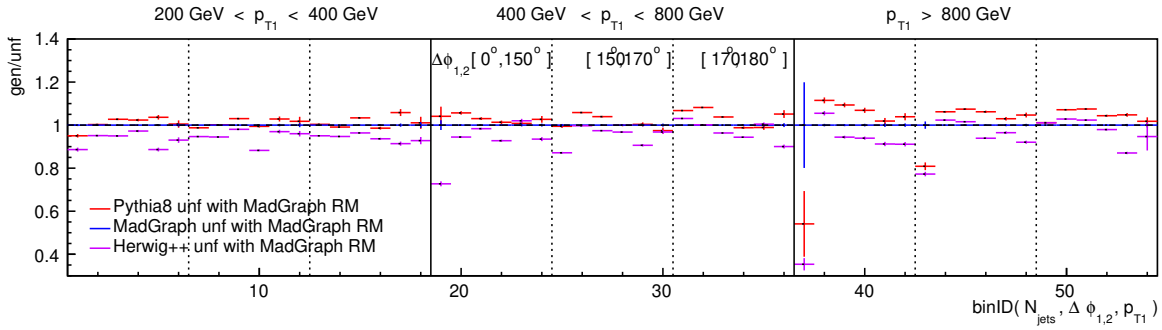
	herwig++	pythia8	MadGraph
detector level	920966	150183	208155
particle level (data stat. only)	965835	162146	210711
particle level (data + MC stat.)	185801	94085	172736

Table 6.2: Bottom line test χ^2 values for using HERWIG++, PYTHIA8 and MADGRAPH RM's for unfolding the data for the four leading jets p_{T} (p_{Ti}, n_i) observable with pseudo inversion.

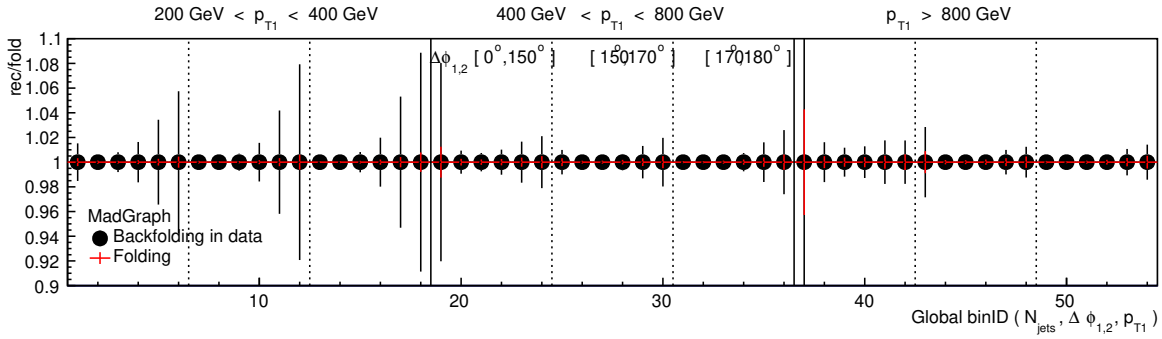
	HERWIG++	PYTHIA8	MADGRAPH
detector level	3676970	277176	534201
particle level (data stat. only)	4271410	295463	515316
particle level (data + MC stat.)	1013850	172894	424168



(a) Bottom line test

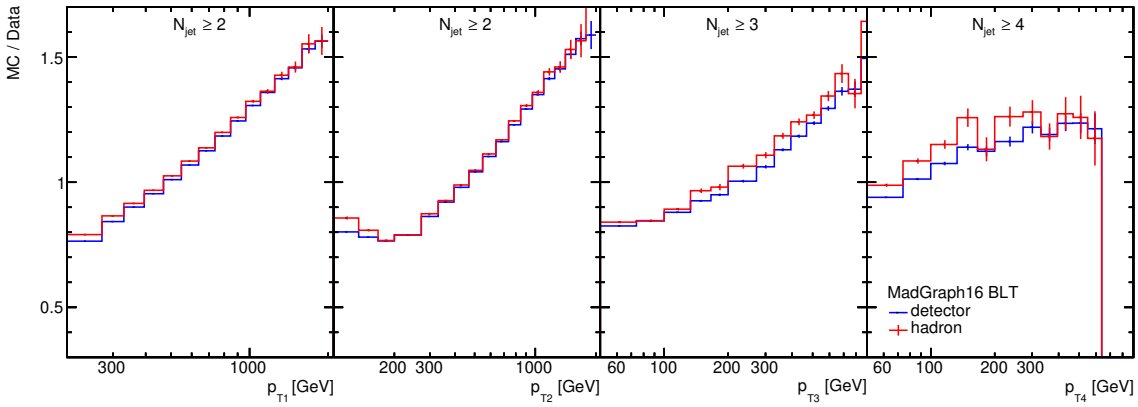


(b) Closure test

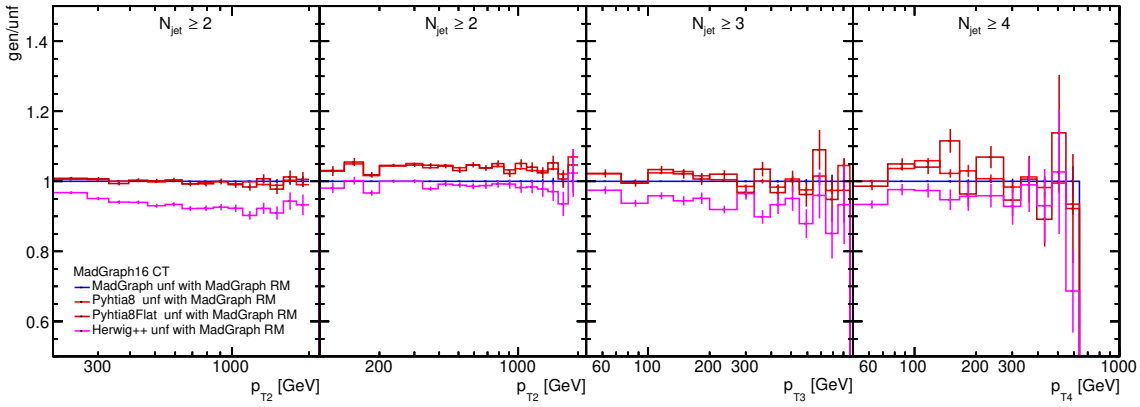


(c) Folding and backfolding

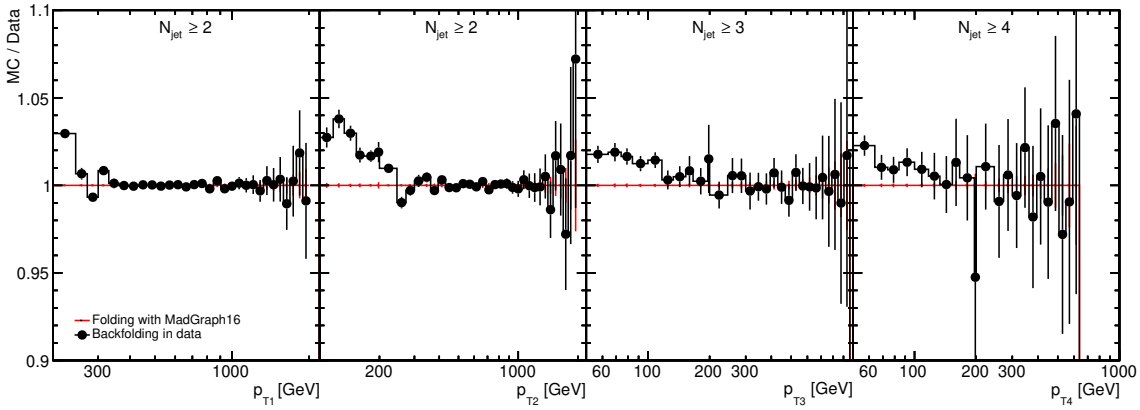
Figure 6.8: Here bottom line test (a), closure test (b) , folding and back-folding tests (c) are shown. MADGRAPH is the default MC to perform the unfolding. The x-axis label and other details are as in Fig. 6.6.



(a) Bottom line test



(b) Closure test



(c) Folding and backfolding

Figure 6.9: Bottom line test (a), closure test (b) , folding and backfolding tests (c) are shown. MadGraph is the default MC to perform the unfolding.

for all samples, the test is successful. The backfolding test gives a perfect agreement since we are using real inversion.

In Fig. 6.9 all the unfolding tests are shown for the transverse momenta of the four leading jets. The bottom line test shows reasonable agreement between particle and detector level MC/data ratios as shown in Fig. 6.9a. The χ^2 values on Tab. 6.2 for all three studied MC samples are higher for detector level than for particle level (data + MC stat.) as expected for a successful test. Since pseudo inversion is used (and not real inversion) for this observable, the backfolding test is good enough. The folding test shows the closure of the phase space and the good treatment of the migrations.

6.2 Uncertainties after unfolding

This section will describe the treatment of statistical and systematical uncertainties.

6.2.1 Statistical uncertainties

The statistical uncertainty from data and simulation (inefficiencies, background and RM limited statistics) are considered in the total covariance matrix as input to the unfolding (by using `TUnfold::Density::SetInput()`). The covariance matrix of the measurement in data after unfolding is obtained using `TUnfoldDensity::GetEmatrixInput()`. For the simulation the contribution from the RM is calculated using `TUnfoldDensity::GetEmatrixSysUncorr()`, for the background by using `TUnfoldDensity::GetEmatrixSysBackgroundUncorr()`, and for the inefficiencies the calculation its done by hand after unfolding. All these covariance matrices (from data and MC statistics) are added to get the final covariance matrix. The statistical uncertainty on data after unfolding increases as shown in Fig.6.10 for the jet multiplicity and the jet p_{T} measurements.

6.2.2 Systematic uncertainties

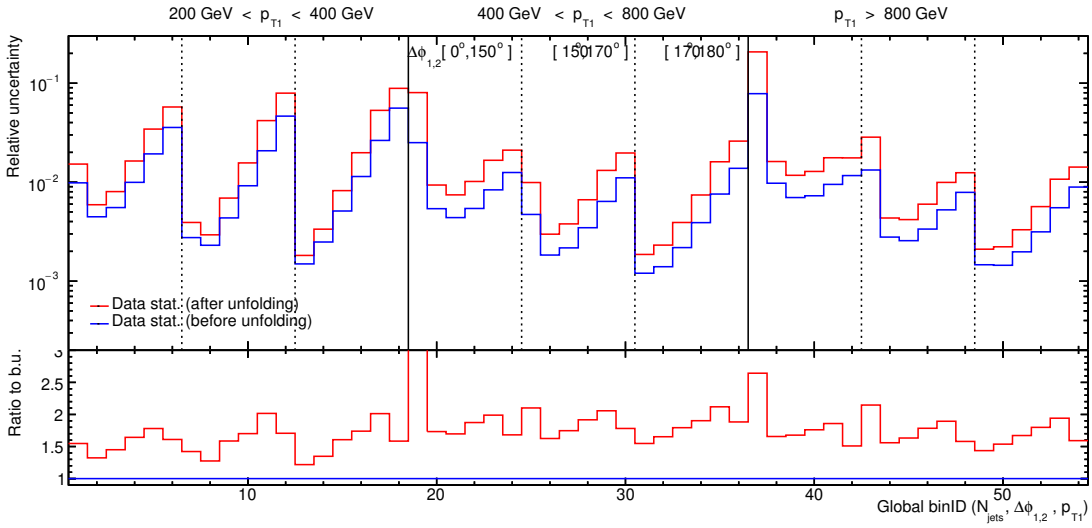
The systematic uncertainties considered in this analysis come from the jet calibrations, pileup treatment, luminosity, and model dependence, as described in the following.

Jet energy scale corrections (JES)

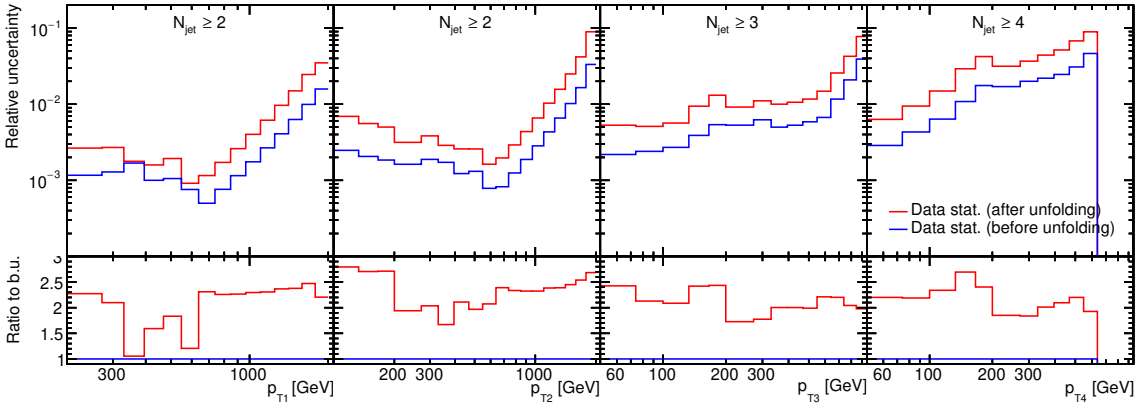
The variations of the JES corrections are applied in the data (following the JetMET recommendations). The unfolding is done for each variation (using the nominal RM matrix from MC), the difference with the nominal value is calculated and added in quadrature for calculating the uncertainty.

Jet energy resolution (JER)

For the JER, the systematic effects change the migrations in the RMs. Therefore we use one RM for each variation, repeating the unfolding for each one, and finally, the uncertainties are obtained using `TUnfoldDensity::GetDeltaSysSource()`.



(a) Jet multiplicity



(b) Four leading jets p_{T}

Figure 6.10: Data statistical uncertainty before and after unfolding.

Other

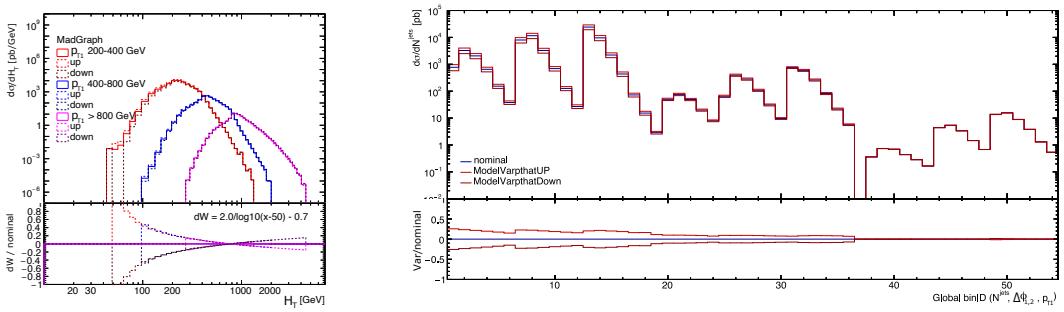
This is composed of the pileup, model dependence, luminosity, and prefiring uncertainties (added in quadrature) presented in the following:

Pileup reweighting

The pileup reweighting uncertainty is estimated by varying up and down the minimum bias cross section by ±5% and repeating the unfolding for the up and down variations. A final uncertainty of about 1% goes into the measurements.

Model dependence

The model dependence is addressed by applying a hard scale reweighting as shown in Fig. 6.11 and repeating the unfolding for each variation. This is done instead of simply estimating the model uncertainty from the differences between unfolded data with HERWIG++ and PYTHIA8 compared to MADGRAPH because of the significant statistical fluctuations, e.g., more than



(a) Hard scale distribution for each leading jet p_{T1}.

(b) Effect on jet multiplicity at detector level (before unfolding).

Figure 6.11: Hard scale (H_T) reweighting on MadGraph sample. Ont the left, the event weight is varied (up and down) according to: $dW = 2.0/log10(x-50) - 0.7$, with x corresponding to the hard scale. On the right, the effect on the jet multiplicity at detector level is shown.

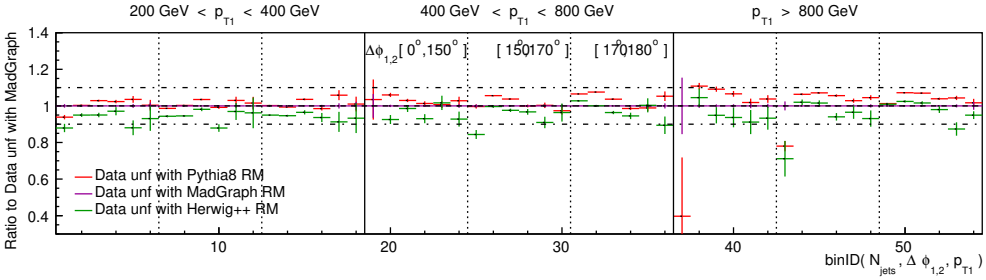


Figure 6.12: Model dependence caused by unfolding data with different MC samples.

20% statistical uncertainty in bin number 37 in Fig. 6.12 corresponding to $N_{jets} = 2$ in the low $0^\circ < \Delta\phi_{1,2} < 150^\circ$ region at $p_{T1} > 800\,\text{GeV}$ after unfolding, was observed coming from using the PYTHIA8 and HERWIG++ RM (HERWIG++ is not within the plot y-axis but the statistical uncertainty is even larger than the one from PYTHIA8). Since the background and inefficiencies are different for different MC models (see Sec. 6.1), and considering the effect of unfolding the data with the different models (see Fig. 6.12), we also add the uncertainties coming from a 15% variation of background (fake) and inefficiencies(miss). Finally, the model uncertainty is composed by adding in quadrature the effect of the hard scale reweighting, background, and inefficiency uncertainties. In Fig. 6.13 the estimated model uncertainty is shown, and it is in reasonable agreement with the effect of unfolding the data with different MC models (PYTHIA8 and HERWIG++).

Luminosity

The luminosity uncertainty is added after unfolding since it is an uncertainty on the absolute scale of the measurement. Here the value of 1.2% for 2016 data is used [130].

Prefiring

The prefiring uncertainty is determined using the prefiring weight variation(Sec.5.5.3), which corresponds to the maximum between 20% of the probability of prefiring and the statistical uncertainty of the probability of prefiring.

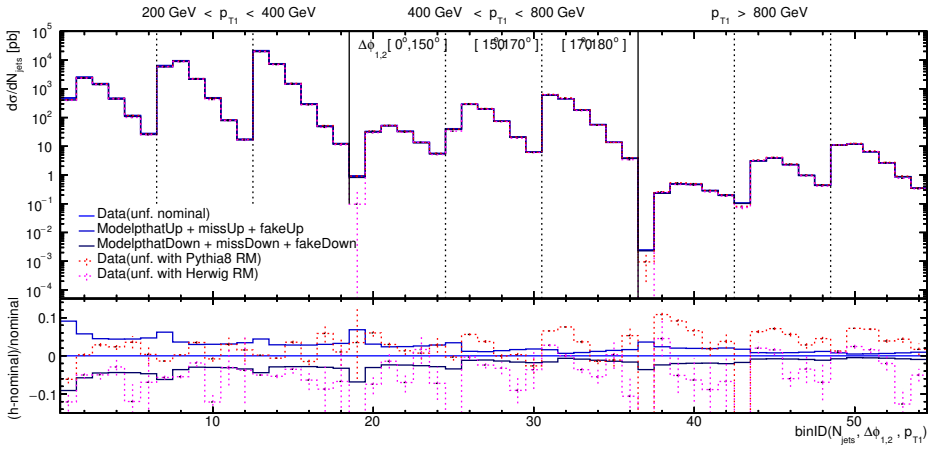


Figure 6.13: Model dependence caused by unfolding data with different simulated samples compared to the estimated model uncertainty.

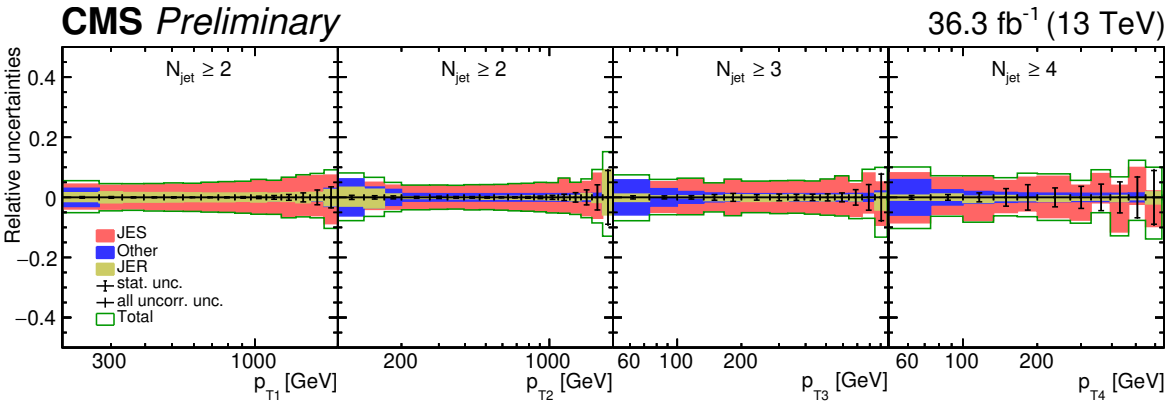


Figure 6.14: Four leading jets p_{T} relative uncertainties. The blue band includes pileup, luminosity, prefiring and model uncertainties added in quadrature.

In Figures 6.14-6.15 the relative uncertainties for p_{T} of the four leading jets and jet multiplicity (in bins of p_{T1} and $\Delta\phi_{1,2}$) measurements are shown. The dominant uncertainty is the JES (red band) for these observables. The total (relative) uncertainty is about $\pm 5-15\%$ for the four leading jet p_{T} measurement and about $\pm 5-20\%$ for the jet multiplicity measurement.

6.2.3 Correlations

The new multi-differential measurements presented in this thesis were done taking into account the correlations among all the different bins. The correlation matrices include data and MC limited statistics contributions of the MC sample used for the unfolding (from the RM as well as from miss and fake). Although figures are shown with correlation and anticorrelation coefficients, since neighboring bins have an opposite sign in these coefficients, we will refer to them as (absolute) correlations in the following discussion. Correlations in multi-differential measurements are crucial for adequately unfolding the data since they measure how strong the migrations already seen in the probability matrices are. The study of

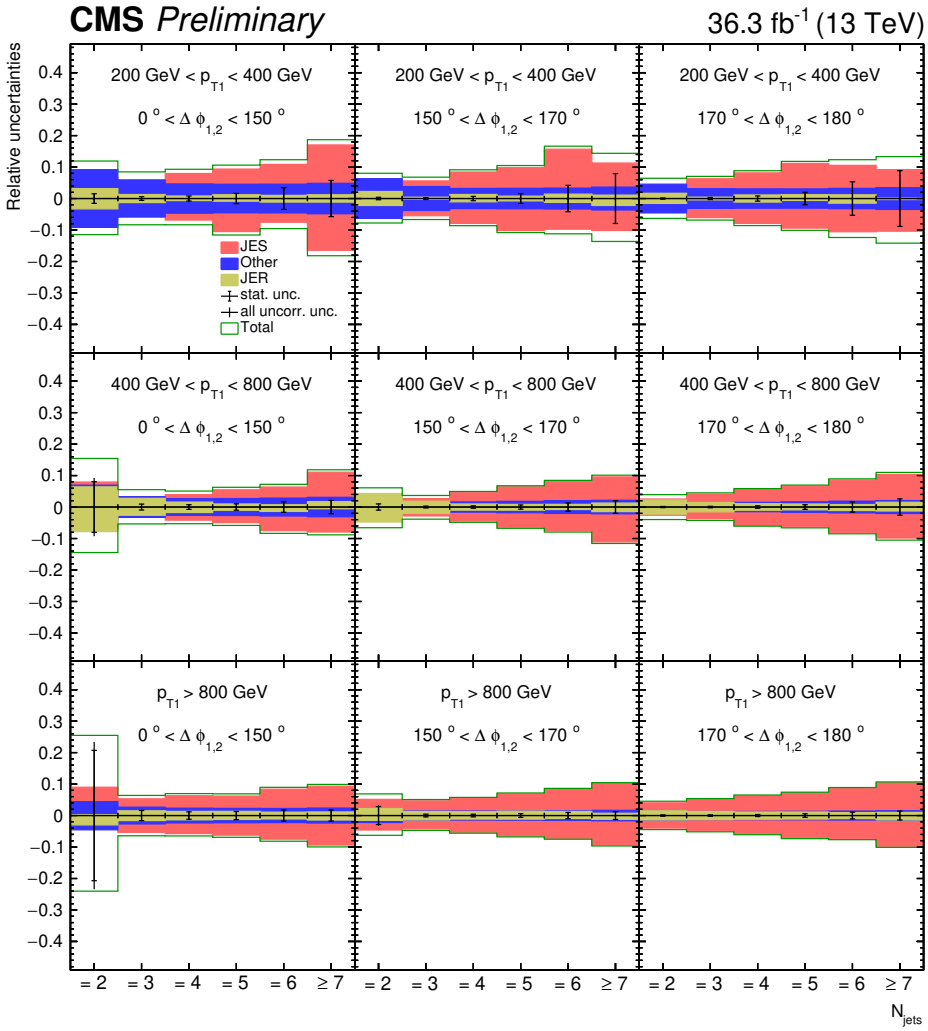


Figure 6.15: Relative uncertainties for the jet multiplicity in bins of the leading jet p_{T} (p_{T1}) and the azimuthal angle between the leading jets ($\Delta\phi_{1,2}$). The blue band includes pileup, luminosity, prefiring and model uncertainties added in quadrature.

correlations can lead us to understand the migration effects deeply in the measurements. Even though they are a consequence of how the PM is constructed (matching strategy used), the final results (measured multi-differential cross sections) should not depend on the treatment of the correlations. More specifically, they should be independent of the matching strategy used. The total uncertainties with the associated correlation matrix can be used to do a proper PDF fitting using this data. Also for doing a χ^2 comparison data to MC, for instance, to compare how different models describe the data.

In Fig. 6.16 the (total) correlation matrix at particle level is shown for the jet multiplicity measurement in bins of p_{T1} and $\Delta\phi_{1,2}$. From there, it can be concluded that the low p_{T1} bins show a correlation of about 30%, while the $\Delta\phi_{1,2}$ measured bins are correlated by less than 10% and consecutive jet multiplicity bins are correlated about $10-40$%. For the measurement of the four leading jets p_{T} the (total) correlation matrix is shown in Fig. 6.17. There one can find correlations of about $20-40$% between the two leading jets and almost less than 20% for the 3rd and 4th leading jets. An interesting fact here is that the two hard leading jets

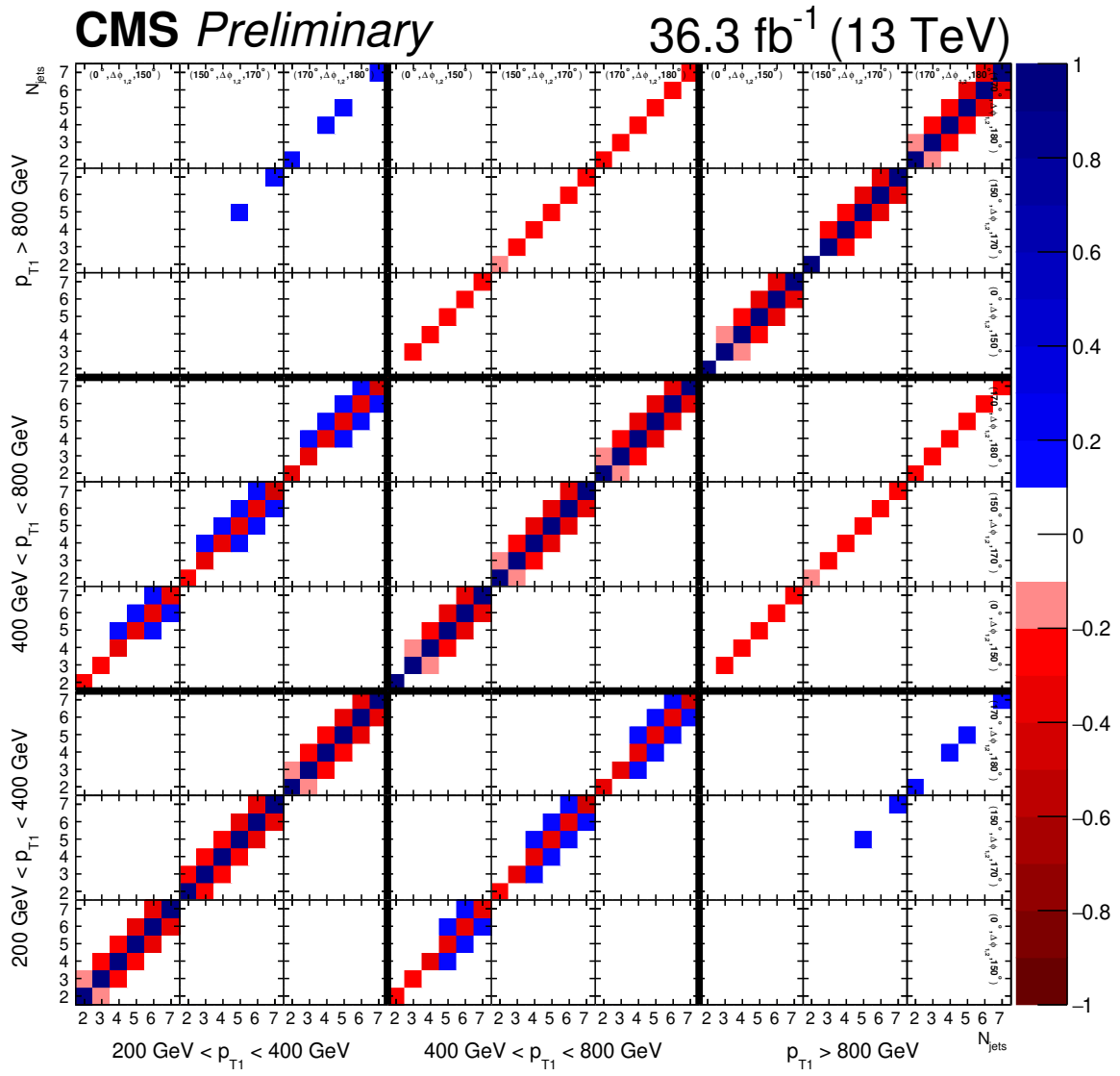


Figure 6.16: Correlation matrix at particle level for the jet multiplicity distribution. It contains contributions from the data recorded in 2016 and the limited statistics from the MadGraph sample. The global 3×3 sectors (delimited by the thick black lines) corresponds to the p_{T1} bins, indicated by the labels in the x(down) and y(left) axis. Inside these, smaller 3×3 structures corresponding to the $\Delta\phi_{1,2}$ bins, indicated in the uppermost row and rightmost column, the x and y-axis of these $\Delta\phi_{1,2}$ cells corresponds to the jet multiplicity. The z-axis covers a range from -1 to 1, indicating the correlations in blue shades and anti-correlation in red shades, the values between -0.1 and 0.1 are represented in white.

are not correlated (less than 10%) with the softer 3rd and 4th and only correlations of about $20-30\%$ are found at the low edges of the phase space, e.g., for p_{T} $50-70\,\mathrm{GeV}$ between the leading and four leading jets. For the p_{T} bins in the same jet (e.g., the leading jet p_{T1}), correlations of about $20-30\%$ are found between at least three neighboring bins.

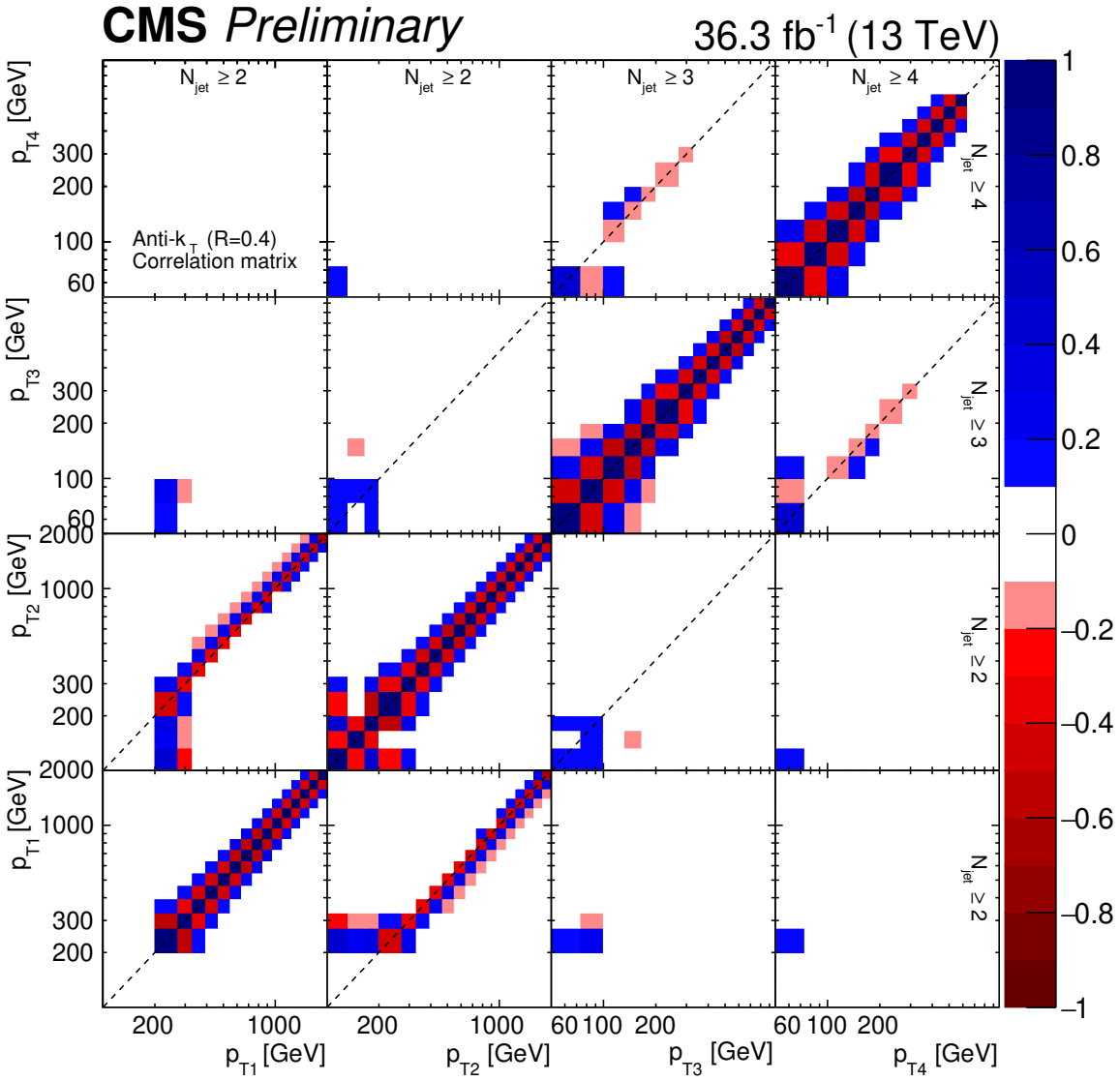


Figure 6.17: Correlation matrix for the particle level p_{T} of the four leading jets. It contains contributions from the data recorded in 2016 and from the limited statistics from the MadGraph sample. Each of the 4×4 sectors corresponds to one of the p_{T} spectra measured, indicated by the x and y-axis labels. The z-axis covers a range from -1 to 1, indicating the correlations in blue shades and anti-correlation in red shades, the values between -0.1 and 0.1 are represented in white.

CHAPTER

7

THEORETICAL PREDICTIONS

Contents

7.1 Comparison with previous jet measurements 98
7.1.1 Inclusive jets 98
7.1.2 Azimuthal correlations 99
7.1.3 Azimuthal correlations in back-to-back region 100
7.2 The role of Multiparton Interactions 100
7.3 Study of MG5_aMC+CA3 (jj) NLO different contributions . . 101
7.4 Dijet and three-jet predictions from MG5_aMC+CA3 NLO . . 103
7.5 Calculations at fixed Next-to-Leading Order 105

The measured differential cross sections will be compared, in the results and discussion chapter, to theoretical predictions at leading order (LO) in the strong coupling (same predictions as presented in Sec. 5.1 at particle level) and at next-to-leading order (NLO). In this chapter, the discussions will be focused on the detailed study of the following NLO predictions:

- MG5_aMC [123] (labeled MG5_aMC+Py8 (jj) NLO) interfaced with PYTHIA8, with MEs computed at NLO for the process $pp \to jj$. The NNPDF 3.0 NLO PDF [144] is used and $\alpha_s(m_Z)$ is set to 0.118. The CUETP8M1 tune is used for PYTHIA8. This calculation is used to study the effect of multi-parton interactions. The factorization and renormalization scales μ_f, μ_r are set to half the sum over the scalar transverse momenta of all produced particles/partons $1/2 \sum_i H_{T\,i}$ (as is the default in MG5_aMCatNLO).

- MG5_aMC [123] (labelled MG5_aMC+CA3 (jj) NLO) interfaced with CASCADE3 [42] with MEs computed at NLO for the process $pp \to jj$. The HERWIG6 subtraction terms in MCatNLO are used, as they are closest to the needs for applying PB TMDs parton

densities, as described in Ref. [42]. The NLO PB TMD set 2 [39] with $\alpha_s(m_Z) = 0.118$ is used. The collinear parton density corresponding to PB set 2 is used for the ME calculation. This parton density leads to cross sections which are 10 – 20 % smaller than the corresponding ones obtained, for example, with NNPDF, which is because PB parton densities are obtained from a fit to HERA data only and do not include any other data set from pp or $p\overline{p}$ scattering. The inclusion of the transverse momentum k_T and initial state parton shower is performed with CASCADE3 [42], final state radiation (since not constrained by TMDs) as well as hadronization is performed with PYTHIA6 [145]. The initial state parton shower follows the PB TMD distribution. In contrast to ordinary parton showers, it does not change the kinematics of the hard process after a k_T from the TMD is added to the initial state partons. Multiparton interactions are not simulated in this approach.

- MG5_aMC [123] interfaced with CASCADE3 (labeled MG5_aMC+CA3 (jjj) NLO), with MEs computed at NLO for the process $pp \to jjj$. The same PB TMD distribution and parton shower, as for MG5_aMC+CA3 (jj) NLO is applied.

In Sec. 7.1 the NLO jj predictions interfaced with CASCADE3 and PYTHIA8 will be compared to previous jet measurements to obtain important information which will help us to understand the predictions. In addition, in Sec. 7.2, the role of multi parton interactions in the MG5_aMC+PY8 (jj) NLO sample will be studied for the measured observables in this thesis. Afterward (in Sec. 7.3) the MG5_aMC+CA3 (jj) NLO sample is studied in detail for azimuthal correlations and the measurements in this thesis. Then, in Sec. 7.4, the MG5_aMC+CA3 (jj) NLO and the corresponding jjj prediction are compared for the observables measured in this thesis. Finally (in Sec. 7.5) fixed order NLO from MG5_aMC (jj) is compared to the LHE level (NLO with the HERWIG6 subtraction) and to MG5_aMC+CA3 (jj) NLO for the observables of this thesis. In all the NLO predictions shown in the following sections, the shaded bands represent the scale variation, and the statistical uncertainty is represented as vertical lines. The scale uncertainty has been estimated by using the 7-point variation of the central factorization and renormalization scales (μ_F, μ_R) by a factor of two (up and down avoiding the extreme variations), choosing the maximum and minimum values of such variations for constructing the envelope. The total experimental uncertainty (statistical plus systematic measured uncertainties) is shown as vertical bar for the data.

7.1 Comparison with previous jet measurements

This section shows comparisons of the NLO calculations with previous jet measurements, on inclusive jets, and on azimuthal correlations, which are all relevant for discussing the results in this analysis. The calculations are performed with MG5_aMC+PY8 (jj) NLO and with MG5_aMC+CA3 (jj) NLO and analyzed with the corresponding RIVET plugins.

7.1.1 Inclusive jets

In Fig. 7.1 we show a comparison of predictions from MG5_aMC+PY8 (jj) NLO and MG5_aMC+CA3 (jj) NLO with the inclusive jet measurements at 13 TeV, measured in 2015 [146]. The PB set parton density is obtained from HERA data alone, with no pp

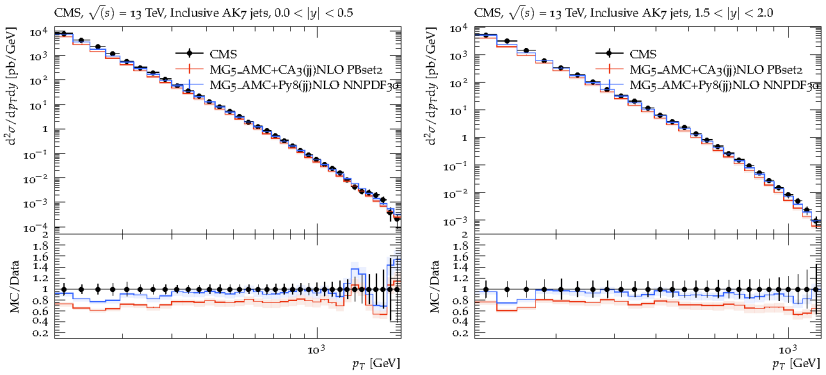


Figure 7.1: Comparison of the inclusive jet measurement with predictions from MG5_aMC+Py8 (jj) NLO) and with MG5_aMC+CA3 (jj) NLO.

or $p\bar{p}$ measurements are included. From the comparison with inclusive jet measurements at 13 TeV, it turns out that this PDF results in predictions that are about 20% too low compared to data as well as compared to predictions obtained with the NNPDF parton density. However, to have a consistent collinear PDF with the TMD PDF, we also kept using PB set2 for the calculation in MG5_aMCatNLO. This comparison will be later used to motivate a normalization to the measured inclusive dijet cross section for the predictions at NLO (the same will be done for the LO predictions) in the discussion of the final results on Sec. 8.

7.1.2 Azimuthal correlations

In Fig. 7.2 a comparison of predictions from MG5_aMC+Py8 (jj) NLO and MG5_aMC+CA3 (jj) NLO with measurements of azimuthal correlations at 13 TeV [6] is shown. Since the distributions are normalized, the agreement with PB set2 is good. In these distributions, one can see the different behavior of the parton shower in Pythia8 and the TMD shower following the PB-TMDs in Cascade3.

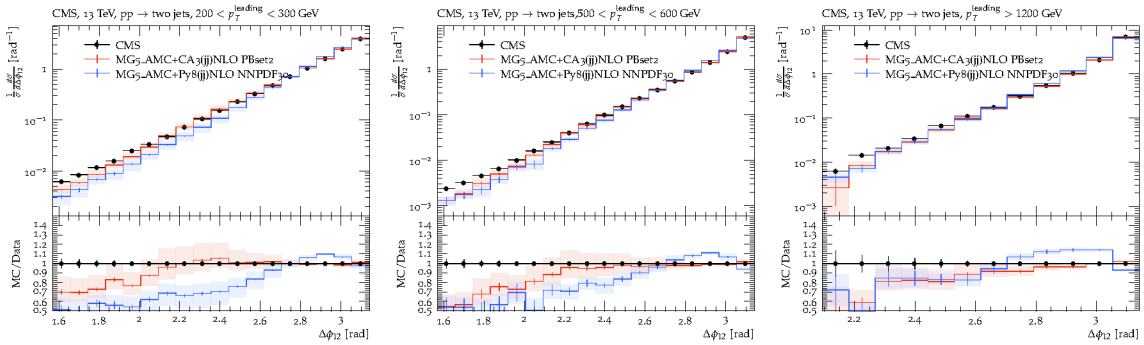


Figure 7.2: Comparison of azimuthal correlations in dijet events with predictions from MG5_aMC+Py8 (jj) NLO) and with MG5_aMC+CA3 (jj) NLO.

7.1.3 Azimuthal correlations in back-to-back region

In Fig. 7.3 we show a comparison of predictions from MG5_aMC+Py8 (jj) NLO and MG5_aMC+CA3 (jj) NLO with measurements of azimuthal correlations in the back-to-back region at 13 TeV [7]. The back-to-back region is especially sensitive to soft gluon

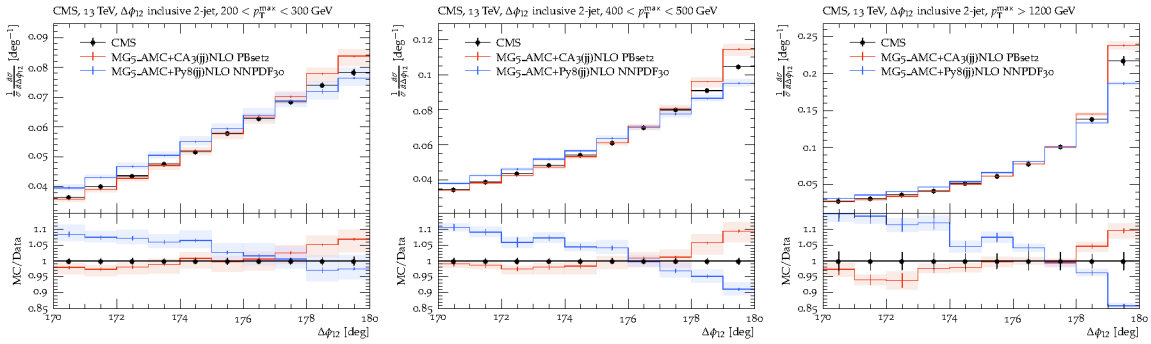


Figure 7.3: Comparison of azimuthal correlations in the back-to-back region in dijet events with predictions from MG5_aMC+Py8 (jj) NLO) and with MG5_aMC+CA3 (jj) NLO.

resummation effects, and therefore, the predictions from MG5_aMC+CA3 (jj) NLO with PB-TMDs is of particular interest. In this phase space region, the PB-TMD shower approach describes the shape of the distribution better than PYTHIA8. In Ref. [47] a more detailed study from the efforts of the PB-TMD group can be found.

7.2 The role of Multiparton Interactions

To study the role of multiparton interactions for the measured observables in this thesis, we use the MG5_aMC+Py8 (jj) NLO prediction with and without the simulation of multiparton interactions (using CUETP8M1 tune [121]). In Fig. 7.4 and Fig. 7.5, we show a comparison of the jet multiplicity and the p_T spectra of the four leading jets. From these comparisons, we can conclude that the effect of MPI is small and within the scale uncertainty.

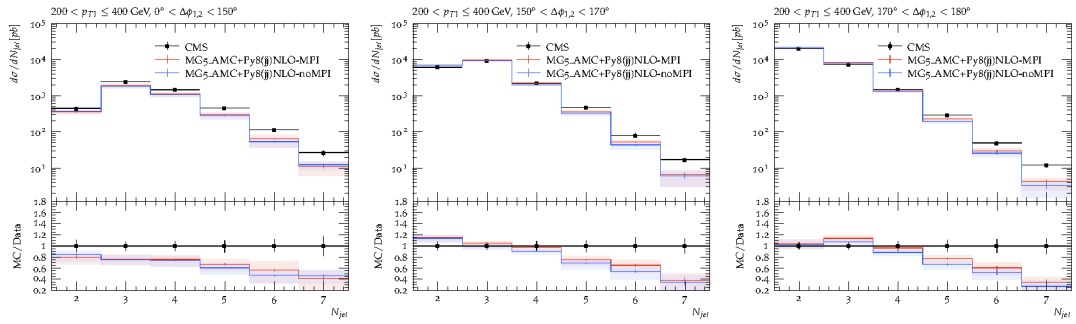


Figure 7.4: Comparison of the jet multiplicity in dijet events with predictions from MG5_aMC+CA3 (jj) NLO.

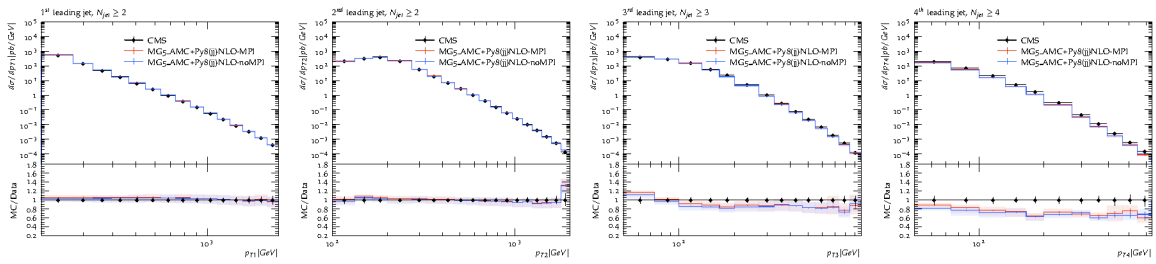


Figure 7.5: Comparison of the cross section of the jet transverse momentum dijet events with predictions from MG5_aMC+CA3 (jj) NLO.

7.3 Study of MG5_aMC+CA3 (jj) NLO different contributions

This section shows how the contribution from PB-TMDs, initial state TMD shower, and final state shower build the final distribution. The prediction of the individual pieces is nonphysical because soft and collinear contributions are subtracted from the NLO matrix element in the MC@NLO procedure, which are then added back by the parton shower and TMDs. However, it is still interesting to see how significant the different contributions are. We show the contributions from pure Les Houches Event [53] level (LHE) after transverse momenta are included via TMDs (TMD), and the effect of initial state TMD shower (ISR), as well as the impact of the conventional final state shower (FSR). In the following, we will concentrate on the azimuthal correlations and predictions for the measurements presented in this thesis.

Azimuthal correlations

In Fig. 7.6 predictions from MG5_aMC+CA3 (jj) NLO are compared with measurements of azimuthal correlations at 13 TeV from CMS [6]. We show the contributions from pure LHE level after transverse momenta are included via TMDs, the effect of initial state TMD shower, and the impact of the conventional final state shower. The distribution from LHE is constructed from fixed NLO with the corresponding shower subtraction (HERWIG6 is used for CASCADE3), which is an nonphysical distribution. In the figure shown as CAS-LHE, one can notice that most of the events are in the back-to-back bin. Later, when the TMD

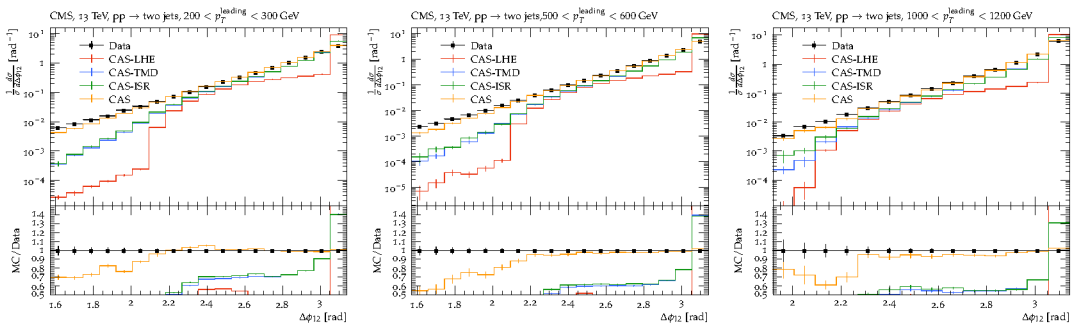


Figure 7.6: Comparison of azimuthal correlations in dijet events with predictions from MG5_aMC+CA3 (jj) NLO.

is applied (CAS-TMD), the effect of the intrinsic k_{T} from the TMD parton density creates (extra) soft radiation that creates a decorrelation in $\Delta\phi_{1,2}$. The ISR from CASCADE3 (CAS-ISR), which follows the TMD, makes a small contribution compared to CAS-TMD in the back-to-back region, only in the $\Delta\phi_{1,2} < 2.8$ one finds a difference of about a few percent. Then finally, when adding the FSR (CAS), which is angular ordered (using PYTHIA6 FSR by vetoing increasing angles), gives a pretty nice description of the non-perturbative region (around $2.5 < \Delta\phi_{1,2} < \pi$). In the end, hadronization is added, but this doesn't affect the distribution (the effect is smaller than the statistical uncertainty).

Jet Multiplicities and jet transverse momenta

In Fig. 7.7 we show how the jet multiplicities are influenced by the inclusion of TMD, initial state TMD shower, and final state shower. The jet multiplicity already brings more information into the game because it is more exclusive than the azimuthal correlation measurement. Starting from CAS-LHE one notice that at LHE level up to 3-jets are created, and this comes from the NLO correction (real emission). Then when adding the TMD, the effect is minimal. And then it comes the PB-TMD CAS-ISR that follows the TMD. In the previous example of azimuthal correlations, we couldn't see a significant contribution after the TMD was applied, but now we see that it creates up to 5-jets. This is a nice phase space

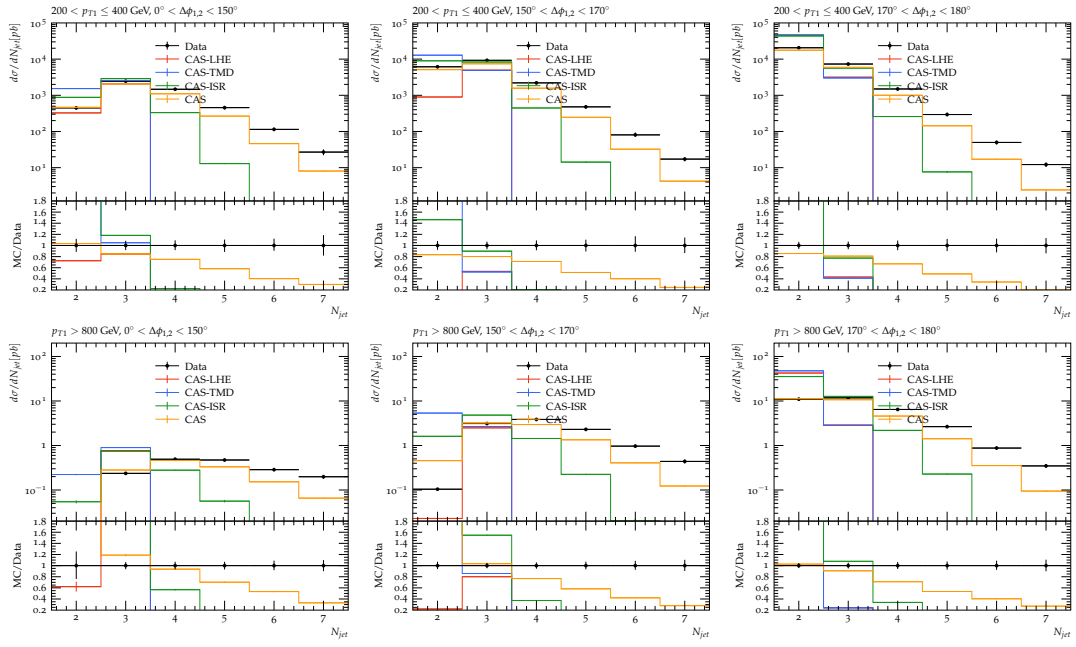


Figure 7.7: Comparison of the measured jet multiplicity in dijet events with predictions from MG5_aMC+CA3 (jj) NLO. Here all the different levels of the calculation are shown applied one after the other, from pure LHE up to FSR with hadronization at the very end.

region (and observable) to see the TMD initial state shower impact from CASCADE3. Then when FSR is added (CAS), it simulates up to 7-jets and more jets (since the last bin in the jet multiplicity is inclusive $N_{jet} \geq 7$). There are two essential regions: the low $\Delta\phi_{1,2}$ and the back-to-back region (both at low and high p_{T1}). The ME LHE level dominates the first two bins in the first one, and the shower generates the rest. In the back-to-back region, it happens the opposite. The calculation at the LHE level is off by more than 80% since this is the

resummation region, and the parton shower generates the additional jets. The back-to-back region at low p_{T1} for CAS is found to be around 20% below the data in accordance = with what was found for inclusive jets in Sec. 7.1.1.

For the p_{T} spectrum in Fig. 9.1, we already can notice different features than for the jet multiplicities. For example, we are more sensitive to the overall normalization since the full prediction (CAS) is around 20% below the data . This is consistent with what we saw for the inclusive jet spectra. It is what motivates us to normalize the predictions when doing the final comparison (in Ch. 8). In the azimuthal correlations and jet multiplicities, we saw that the FSR adds up to the ISR, and for two jet scenarios in the jet multiplicity CAS-ISR contributed more than FSR. This nice observation will be addressed here in the following by studying the p_{T} spectra of the jets. In this case for the leading (p_{T1}) and second (p_{T2}) leading p_{T} jets we see that the TMD and the TMD shower (ISR) overshoots the FSR (CAS) contribution, and even the data. Then for the third leading jet p_{T} spectra (p_{T3}) where soft and hard radiation (from the ME real emission) contributes similarly, e.g., if one compares LHE with TMD and ISR and FSR effects in the ratio, there the ISR still overshoots the FSR contribution, but it's below the data.

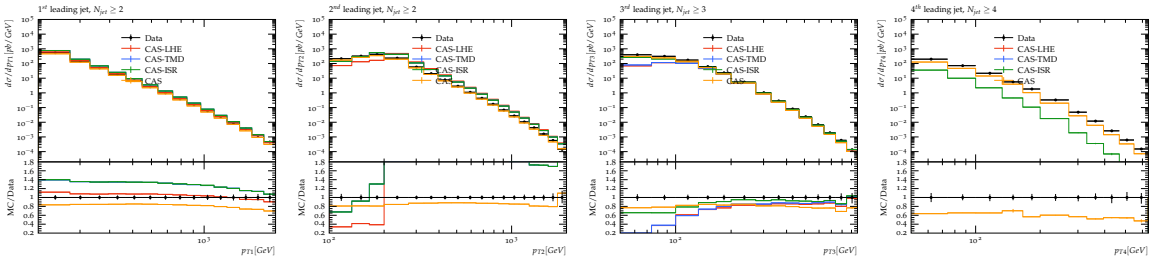


Figure 7.8: Comparison of the cross section of the jet transverse momentum dijet events with predictions from MG5_aMC+CA3 (jj) NLO.Here all the different levels of the calculation are shown applied one after the other, from pure LHE up to FSR with hadronization at the very end.

For the four leading jets, the ISR overshoots the FSR. This is because, for this observable at NLO, we are very sensitive to ISR and FSR effects, especially on the 4th jet distribution, which comes essentially from the shower. Only up to three jets are generated at ME. In principle, ISR and FSR should have similar contributions, leading to different distributions when the jets are clustered.

7.4 Dijet and three-jet predictions from MG5_aMC+CA3 NLO

We have calculated dijet and three-jet matrix elements within MG5_aMCatNLO and passed them through Cascade3 . In Fig. 7.9 and Fig. 7.10 the 2-jet and 3-jet NLO predictions are compared to the measured jet multiplicity and jet p_{T} spectra.

From Fig. 7.9 the effect of higher order correction can be seen, since MG5_aMC+CA3 (jjj) NLO prediction contributes more to higher jet multiplicity ($N_{jet} > 2$) than the MG5_aMC+CA3 (jjj) NLO prediction. This can also be notice when looking at Fig. 7.10 where the MG5_aMC+CA3 (jj) NLO prediction describes the two leading jets better than the MG5_aMC+CA3 (jjj) NLO prediction which describes better the 3rd and 4th jets p_{T} spectra.

Since the MG5_aMC+CA3 (jjj) NLO prediction generates three up to 3-jets (hard jets)

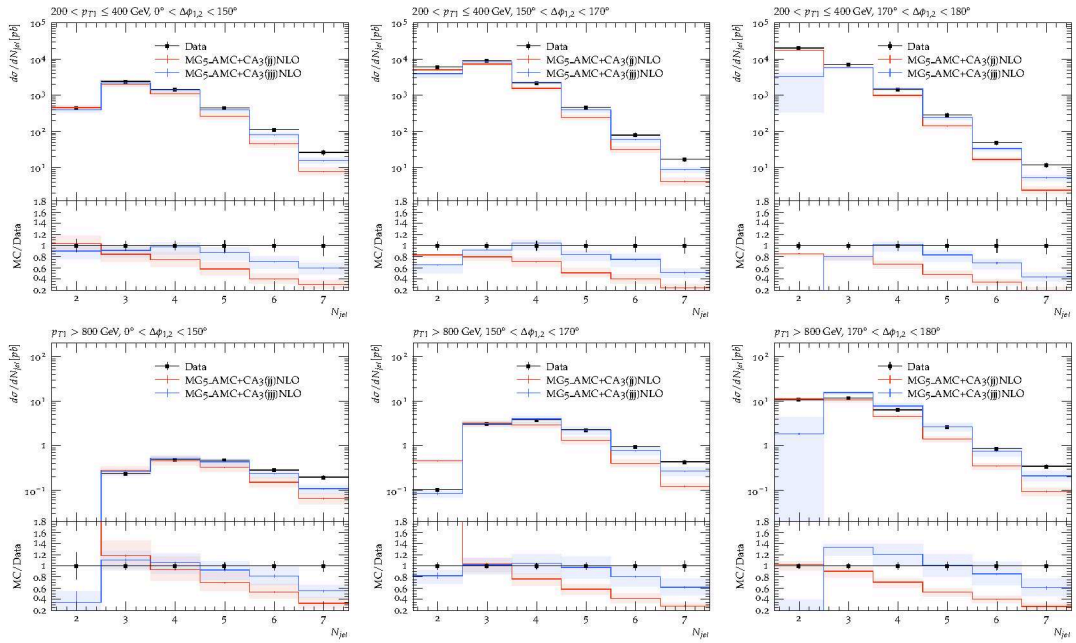


Figure 7.9: Comparison of the jet multiplicity in dijet events with predictions from MG5_aMC+CA3 NLO for two jets and three jet at ME level.

plus the real correction, then one needs at least a fifth jet as hard as the real correction to create such (back-to-back) scenario and the parton shower can not generate such scenario.

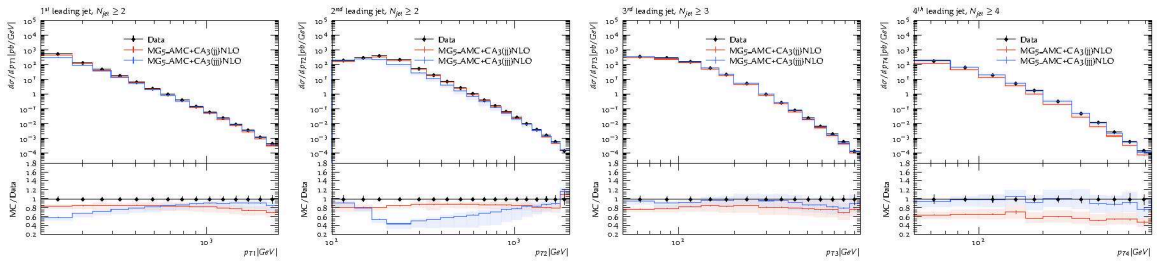


Figure 7.10: Comparison of the four leading p_{T} jets spectrum in dijet events with predictions from MG5_aMC+CA3 NLO for two (jj) and trhee (jjj) jets at ME level.

This reasons explain why the p_{T} spectra in Fig. 7.10 for two leading jets p_{T} (p_{T1} and p_{T2}) the MG5_aMC+CA3 (jjj) NLO prediction doesn't describe the shape of the distributions at $p_{T1}, p_{T2} < 1\,\mathrm{TeV}$ then above this the two leading jets p_{T} can not be decorrelated anymore from the contributions of the real emission and the showers.

Then from the study of these distributions, we conclude that it makes no sense to use MG5_aMC+CA3 (jjj) NLO to describe 2-jets in the back-to-back scenario and for describing the two leading jet p_{T} spectra. This led us to use in the MG5_aMC+CA3 (jjj) NLO prediction for comparison with the three and four leading jet p_{T} measurement and do not use this prediction for the $N_{jet} = 2$ bin for comparison with the jet multiplicity distribution in the future in the results discussion.

7.5 Calculations at fixed Next-to-Leading Order

MG5_aMCatNLO offers the possibility of performing fixed NLO (fNLO) calculations without shower subtraction and storing the results in the form of LHE files, which can be read and processed by CASCADE3 . For the fixed order calculation, CASCADE3 does not change the event record, nor adds parton shower or hadronization, but only produces HEPMC [147] files with the appropriate weight class to be processed in RIVET [148] (grouping event and counter-event with the same event number to guarantee proper calculation of statistical uncertainties). In the following, we compare the calculation of jj at fixed NLO with the results of the matrix element calculation (at nonphysical LHE level) and with the jj jet at NLO supplemented with PB-TMDs, parton shower, and hadronization.

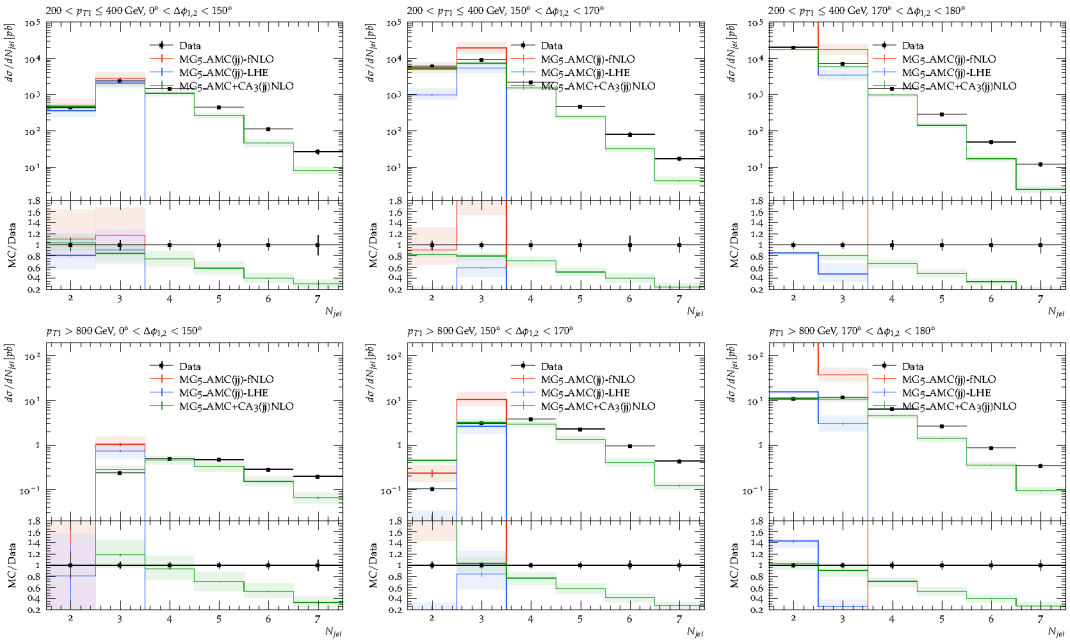


Figure 7.11: Comparison of the jet multiplicity in dijet events with predictions from MG5_aMC (jj) fNLO, MG5_aMC (jj LHE) NLO and MG5_aMC+CA3 (jj) NLO.

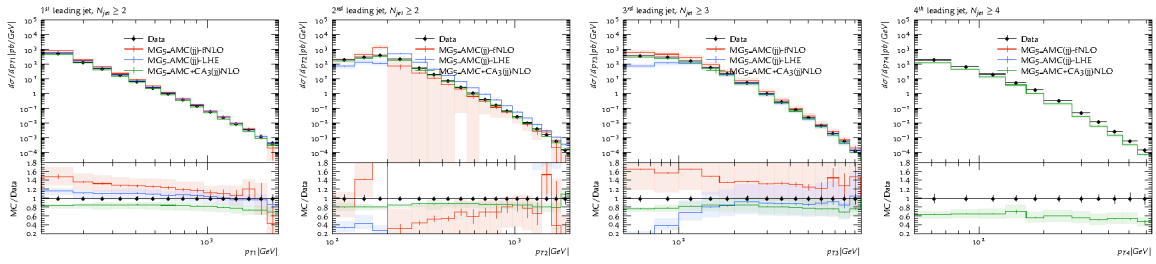


Figure 7.12: Comparison of the cross section of the jet transverse momentum dijet events with predictions from MG5_aMC (jj) fNLO, MG5_aMC (jj LHE) NLO and MG5_aMC+CA3 (jj) NLO.

For the jet multiplicity in Fig. 7.11, we can see that the fixed NLO (MG5_aMC-fNLO) prediction and the LHE level are pretty similar in the low $\Delta\phi_{1,2}$ ($0^\circ, 150^\circ$) region, and this means that this region is dominated by NLO ME level. Then in the other two $\Delta\phi_{1,2}$ regions, especially in the back-to-back region, the parton shower subtraction makes LHE level closer

to data and the full prediction MG5_aMC+CA3 (jj) NLO after ISR and FSR are added. The fixed NLO and the LHE (nonphysical distribution) can only generate up to three hard jets.

On the other hand, for the jet p_{T} spectra in Fig. 7.12 the fixed NLO prediction cannot describe any of the distributions as well as the full prediction, and the scale uncertainty is larger than for the LHE or the whole prediction. This high scale uncertainty comes from the fact that fixed NLO can only describe more inclusive distributions like the inclusive jet p_{T} spectra.

We have learned the importance of the NLO+PS approach in this section since fixed NLO is not helpful when describing exclusive observables like those measured in this work.

CHAPTER

8

RESULTS AND DISCUSSION

Contents

8.1 Jet multiplicity distribution 107

8.2 Transverse momenta of the four leading p_{T} jets 110

In this chapter we compare the unfolded measurements with LO and NLO predictions interfaced with parton shower. Motivated by the comparison to the inclusive jet data (see Chapter 7), all the predictions are normalized to the measured inclusive dijet cross section (σ_{dijet}). The measured inclusive dijet cross section is obtained by summing all the bins of the multi-differential cross section, as function of the jet multiplicity (N_{jets}), azimuthal angle $\Delta\phi_{1,2}$ and leading jet transverse momenta p_{T1}:

$$\sigma_{dijet} = \sum_{i,j,k} \frac{\mathrm{d}\sigma_{dijet}}{\mathrm{d}N^{i}_{jet}\mathrm{d}p^{j}_{T1}\mathrm{d}(\Delta\phi^{k}_{1,2})} \tag{8.1}$$

where i, j, k represent the bins of the measurement as presented in Sec 4.4.2. We use the calculated central value in eq. 8.1 to derive the normalization factors ($N_f = \sigma^{MC}_{dijet}/\sigma_{dijet}$) for each MC prediction.

8.1 Jet multiplicity distribution

The cross section as a function of the jet multiplicity N_{jets}, the azimuthal angle between the two leading p_{T} jets $\Delta\phi_{1,2}$ and the leading jet p_{T1} have been measured. The measurement is presented in three bins of $\Delta\phi_{1,2}$ (p_{T1}), which we will call low $\Delta\phi_{1,2}$ (p_{T1}), medium $\Delta\phi_{1,2}$ (p_{T1}) and back-to-back region (higher p_{T1}) respectively when increasing in $\Delta\phi_{1,2}$ (p_{T1}). In total this measurement have nine bins of $\Delta\phi_{1,2}$ and three different p_{T1} regions, each one divided in bins of the exclusive jet multiplicity up to seven jets (inclusive for this last bin)

as shown in Fig. 8.1. Jets in all the events are first listed in p_{T} within $|y| < 3.2$ and for $p_T > 50\,\mathrm{GeV}$, then the events are required to have $p_{T1} > 200\,\mathrm{GeV}$ and $p_{T2} > 100\,\mathrm{GeV}$ within $|y_{jet}| < 2.5$, and extra jets with $p_T^{min} > 50\,\mathrm{GeV}$ within $|y_{jet}| < 2.5$, as shown in Fig. 8.1 top left legend.

We first discuss the features of the measurement alone. As a measure for the jet multiplicity we can compare the production rate for 3 jets with the one for 7 jets. In the region of low p_{T1} ($200 < p_{T1} < 400\,\mathrm{GeV}$) a large number of additional jets is observed at low $\Delta\phi_{1,2}$ ($0^\circ < \Delta\phi_{1,2} < 150^\circ$), the production rate for 3 and 7 jets changes by two orders of magnitude. In the large $\Delta\phi_{1,2}$ region ($170^\circ < \Delta\phi_{1,2} < 180^\circ$), where the leading jets are nearly to back-to-back, the production rate for 3 and 7 jets changes by three orders of magnitude. It is interesting to note, that even in the back-to-back region a large number of additional jets is observed. In the region of large p_{T1} ($p_{T1} > 800\,\mathrm{GeV}$) we observe that the rate of additional jets at low $\Delta\phi_{1,2}$ is essentially constant, the rate between 3 and 7 jets changes only weakly, indicating that many jets participate in the compensation of the $\Delta\phi_{1,2}$ decorrelation. In the large $\Delta\phi_{1,2}$ region ($170^\circ < \Delta\phi_{1,2} < 180^\circ$) the rate between 3 and 7 jets changes by less than 2 orders of magnitude, in very contrast to the low p_{T1} region. Over the whole range in $\Delta\phi_{1,2}$ more than three or four additional jets at large p_{T1} are needed to describe the data. This observable shows in detail how many jets are needed to build the azimuthal dijet decorrelation in different $\Delta\phi_{1,2}$ regions at different scales (for three different p_{T1} regions).

In the back-to-back region one expects to be more sensitive to the extra jet radiation, since the dijet system would share the p_{T} even between both jets ($p_{T1} \approx p_{T2}$) in most cases and when $p_{T1} \neq p_{T2}$ the "soft" extra jets compensate the dijet system p_{T} so that $\Delta\phi_{1,2} \approx \pi$. On the other hand for lower $\Delta\phi_{1,2}$ the situation is similar just that the scale of the extra-jets can variate since we don't have anymore a fix value of $\Delta\phi_{1,2}$ but a long range which then needs not only soft but more hard jets at similar scales (similar p_{T}) of the lading jets to create $0^\circ < \Delta\phi_{1,2} < 150^\circ$ ($150^\circ < \Delta\phi_{1,2} < 170^\circ$), and also in this region where the dijet system is not in back-to-back topology there is also tiny MET effects since we allowed up to 10% MET fraction in the events.

In Fig. 8.1 predictions from the LO $2 \to 2$ generators PYTHIA8 and HERWIG++ including parton showering and multiparton interactions are shown. The shape of the predictions coming from PYTHIA8 is different to what is observed in the measurement, especially in the region of large $\Delta\phi_{1,2}$. The shape of the prediction from HERWIG++ agrees rather well with the measurement, especially in the large $\Delta\phi_{1,2}$ region. The difference between PYTHIA8 and HERWIG++ in jet multiplicity is coming from the different treatment of the parton shower. The back-to-back region becomes of great interest to test the parton showers and from this first look into LO MC simulations the angular ordered parton shower from HERWIG++, which treats better the correlation effects, describes better the shape of the data than PYTHIA8 where the p_{T} ordered shower over estimates the extra jets ($N_{jets} > 2$) up to 40%. On the other hand in the small $\Delta\phi_{1,2}$ region at low jet multiplicities of $N_{jet} < 6$ the dominant effect is associated to ME contribution and since this is LO in both predictions the normalization and the shape disagrees with the measurement by about 20%. In addition, the prediction from MadGraph+Py8 with up to four non-collinear high-p_{T} partons, supplemented with parton showering and multiparton interactions (from PYTHIA8) is shown. MadGraph+Py8 agrees rather well in shape with the measurement at the same level as PYTHIA8, at least in

at low p_{T1} since they both use the same shower. The normalization from MadGraph+Py8 (25% below the data) is better than from Herwig++ (63% below the data).

The calculations with NLO matrix elements matched with parton shower compared to the measurements are shown in Fig. 8.2. The uncertainty bands of the predictions comes from the variation of the factorization and renormalization scales by a factor of two (using 7-point variations avoiding the largest scale differences). The normalization of MG5_aMC+Py8 (jj) NLO calculation is in reasonable agreement with the measured cross section even for three jets. For higher jet multiplicities the prediction falls below the measurement. The prediction of MG5_aMC+CA3 (jj) NLO is about 20% smaller than the measurement for two and three jets, which is coming from the PB set2 fitted only to HERA data, corrected here by the normalization factor. Apart from the normalization, the MG5_aMC+CA3 (jj) NLO predicts a smaller cross section for more than three jets than observed in the measurement. The MG5_aMC+CA3 (jjj) NLO calculation (using the same normalization factor as for MG5_aMC+CA3 (jj)) gives a rather good description of the three and four jet cross section, while the higher jet multiplicities ($N_{jets} > 6$) are still underestimated.

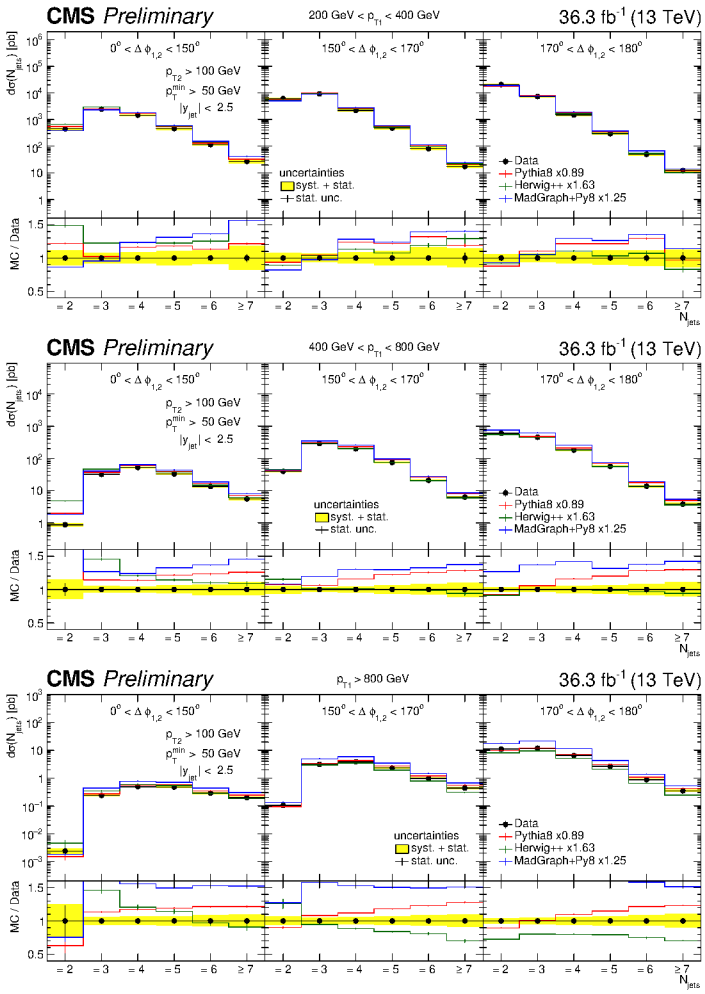


Figure 8.1: Differential cross section as a function of the exclusive jet multiplicity (inclusive for seven jets) in bins of p_{T1} and $\Delta\phi_{1,2}$. The data is compared to LO predictions normalized to the measured inclusive dijet cross section using the scaling factors shown in the legend.

Both MG5_aMC+CA3 (jj) NLO and MG5_aMC+Py8 (jj) NLO in the back-to-back have different shapes as seen in the ratio in Fig. 8.2. The $N_{jet} = 2$ bin is lower than the $N_{jets} = 3, 4$ bins for MG5_aMC+Py8 (jj) NLO while for MG5_aMC+CA3 (jj) NLO this doesn't happen. This feature is related to the PYTHIA8 shower and we saw it already at LO and happens again at NLO. Therefore back-to-back region was found of great interest to shower comparisons at both LO and NLO.

8.2 Transverse momenta of the four leading p_{T} jets

The measured differential jet cross section as a function of the jet transverse momentum, p_{T}, for the four leading p_{T} jets in dijet events is shown in Fig. 8.3. The transverse momentum of the measured jets reaches the TeV scale. We observe that the shape of p_{T} spectrum for the third and fourth leading jets is similar to the one of the two leading jets, however, the

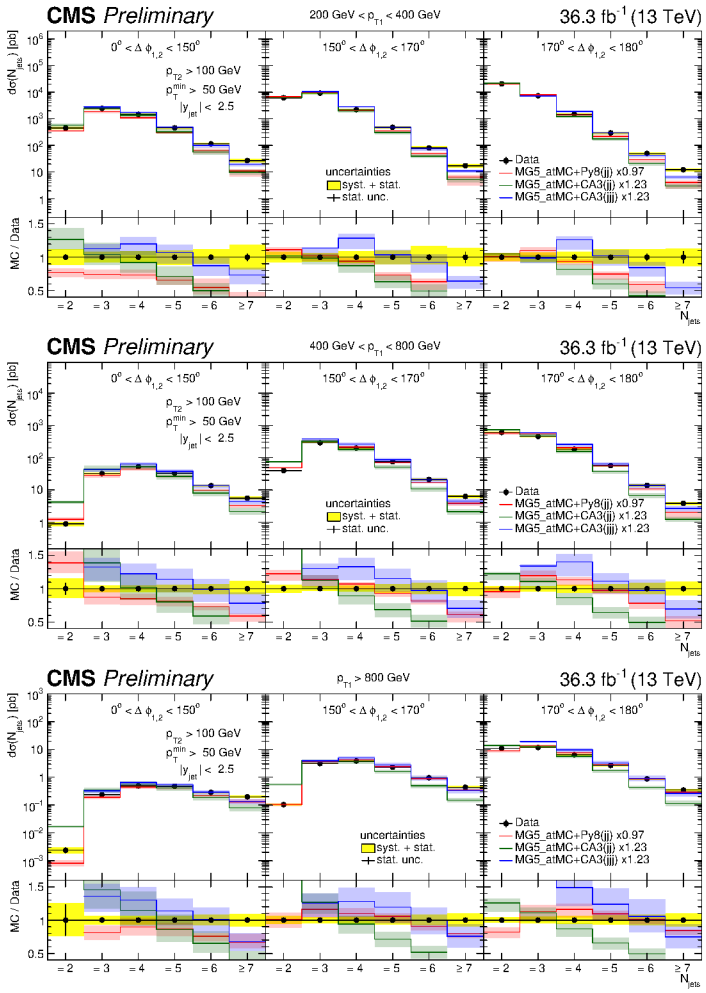


Figure 8.2: Differential cross section as function of the exclusive jet multiplicity (inclusive for 7 jets) in bins of p_{T1} and $\Delta\phi_{1,2}$. The data is compared to NLO dijet predictions MG5_aMC+Py8 (jj) and MG5_aMC+CA3 (jj) as well as the NLO three-jet prediction of MG5_aMC+CA3 (jjj). The predictions are normalized to the measured inclusive dijet cross section using the scaling factors shown in the legend.

cross section is different. The turn on form 100 GeV to 200 GeV for the second leading jet is a consequence of the higher minimum p_{T} cut (200 GeV) applied for the leading jet, in that region the soft radiation compensates the p_{T} difference between the leading jets together with the hard 3rd and 4th jets. In addition, unnormalized PYTHIA8 LO and MG5_aMC+Py8 (jj) NLO (with scale uncertainty as a red band) predictions are shown. Leading order PYTHIA8 overestimates the jets p_{T} except for the second leading jet for $p_T^{jet} > 200$ GeV where the prediction is within data uncertainties. The MG5_aMC+Py8 (jj) NLO predicts within uncertainties the two leading jets p_{T} but for the second and third jets it falls bellow the data.

In Fig. 8.4 the measured differential cross section as a function of the p_{T} for the four leading jets is compared to LO predictions (using the same normalization factors as in Fig. 8.1). The PYTHIA8 prediction describes rather well the shape of the four leading jets p_{T} measured. The leading jet p_{T} (p_{T1}) is nicely described in shape by PYTHIA8 within experimental uncertainties, and also the second leading jet for $p_{T2} > 200$ GeV. The shape the third and fourth jet distributions is not well described, and PYTHIA8 overestimates the p_{T4} up to 40 %. The prediction from HERWIG is not in agreement in shape with the measurements, specially at high p_{T} for the leading and subleading jets where it underestimate the data up to 50%. The prediction from MadGraph+Py8 gives a significantly different shape of the p_{T} spectrum for

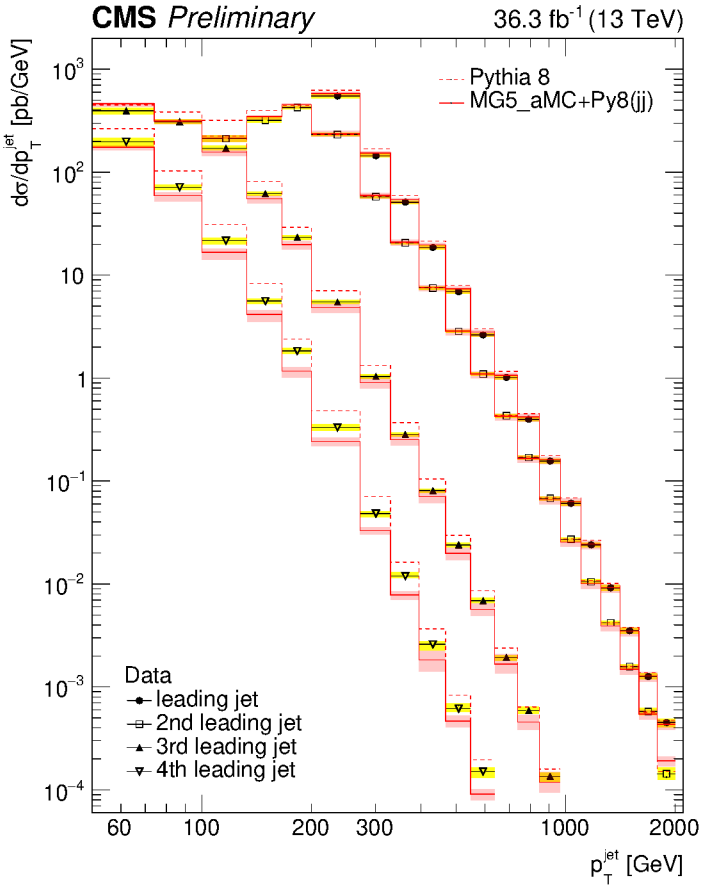
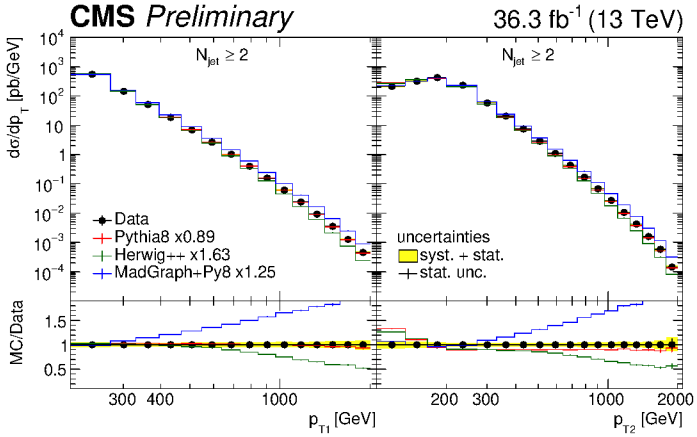


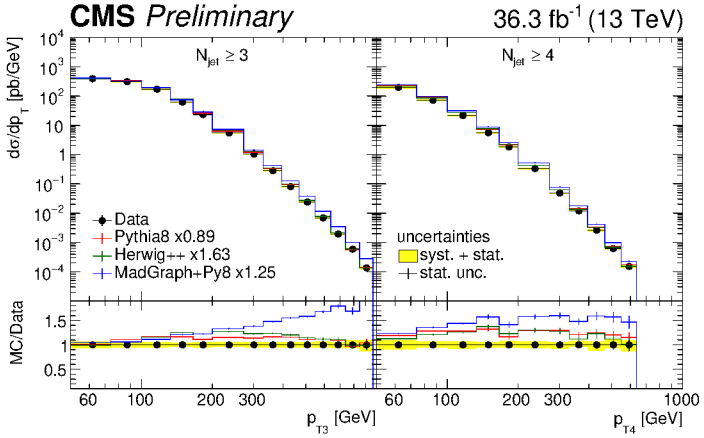
Figure 8.3: Transverse momenta of the measured four leading jets, here the yellow band represents the total experimental uncertainty. Data is compared to LO (PYTHIA8) and NLO (MG5_aMC+Py8) predictions. The red band in the NLO prediction represents the scale uncertainty.

the first three jets. Nevertheless for the second leading jet p_{T} the region $100 < p_{T2} < 200\,\mathrm{GeV}$ the MadGraph+Py8 prediction gives a better description than pythia8 and herwig++ since it is a merged prediction and in this region the contribution from ME higher orders creates such scenario, it is not a matter of how many jets are generated in this region but how the p_{T} is distributed and there contributions from hard ME jets becomes relevant. For the fourth leading jet the p_{T} have similar shape description as pythia8 and herwig++.

The predictions obtained with NLO matrix elements are shown in Fig. 8.5 using the same normalization factors as in Fig. 8.2. MG5_aMC+Py8 (jj) describes the normalization and the shape of the first two jets.For the third jet the shape is different, it is over the data about 15 % for the first bin and falls under the data for the rest of the p_{T} bins ($p_{T3} > 100\,\mathrm{GeV}$) about 20%. For p_{T4} MG5_aMC+Py8 (jj) falls below the data but the shape is reasonable. MG5_aMC+CA3 (jj) predicts the shapes of the first three leading jets quite well and within uncertainties, it improves the third jet shape prediction compared to MG5_aMC+Py8 (jj). The prediction of MG5_aMC+CA3 (jjj) describes the third and fourth jets within uncertainties (predictions for the first and second jets are meaningless for MG5_aMC+CA3 (jjj) and therefore not shown).



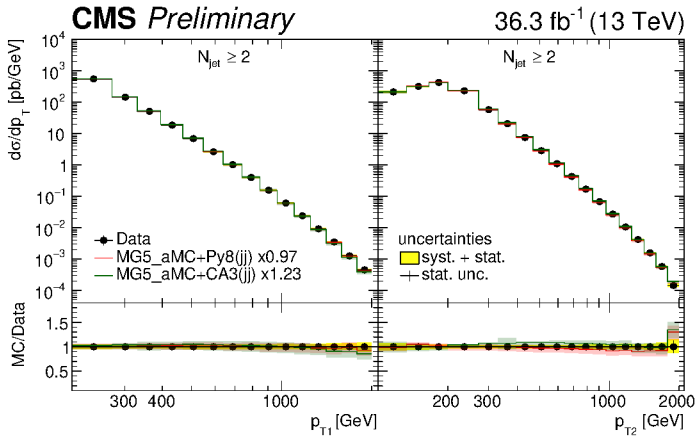
(a) Leading and second leading jet p_{T} distributions.



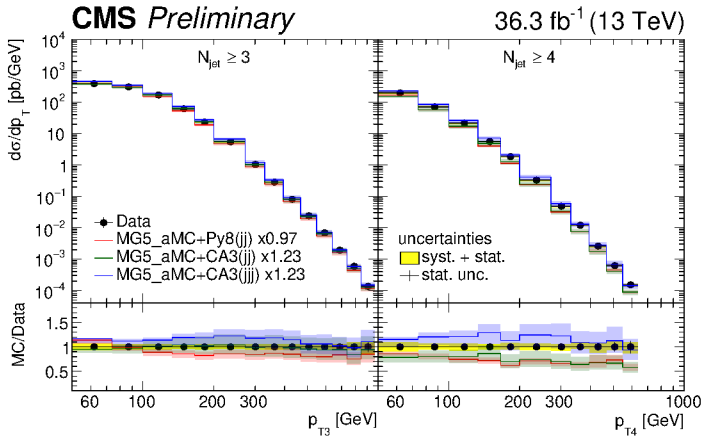
(b) Third and fourth leading jet p_{T} distributions.

Figure 8.4: Transverse momenta of the four leading jets compared to LO predictions normalized to the inclusive dijet data cross section using the scaling factors shown in the legend.

The LO+PS predictions were not able to describe both jet multiplicity and p_{T} measurements. The description is improved for the NLO+PS predictions, but the lack of higher-order contributions can still be seen in the p_{T} and jet multiplicity measurements. The standard parton shower approach from PYTHIA8 in the MG5_aMC+PY8 (jj) prediction gives a similar shape description as PB-TMD, and TMD shower for the jet p_{T} spectra but slightly different for the jet multiplicity. Nevertheless, the PB-TMD approach does not depend on tuning like PYTHIA8 [44]. Also the kinematics in the TMD shower is determined by energy-momentum conservation, not being the case in traditional parton shower like PYTHIA8 where it gives kinematic shifts in the longitudinal momentum distributions [41]. The PB method is less parameter-dependent than PYTHIA8 and other standard approaches without losing the prediction power.



(a) Leading and second leading jet p_{T} distributions.



(b) Third and fourth leading jet p_{T} distributions.

Figure 8.5: Transverse momenta of the four leading jets compared to NLO predictions normalized to the inclusive dijet data cross section by using the scaling factors shown in the legend.

CHAPTER

9

CONCLUSIONS

Summary

In this thesis, new multi-differential jet measurements in dijet events are presented. The data were collected from pp collisions at a center-of-mass energy of 13 TeV, with the CMS detector corresponding to an integrated luminosity of $36.3\,\text{fb}^{-1}$ during 2016. The measurements are performed (at particle level) first selecting all jets within $|y_{jet}| < 3.2$ and $p_T^{jet} > 20\,\text{GeV}$, this jets are then ordered in decreasing p_T and events in which a dijet system contains a jet with $p_\text{T} > 200\,\text{GeV}$ and a subleading jet with $p_\text{T} > 100\,\text{GeV}$, and (possible) extra jets with $p_\text{T} > 50\,\text{GeV}$ within $|y| < 2.5$ are selected.

For the first time, the cross-section as a function of the jet multiplicity, the leading jet p_T and the azimuthal angle between the two leading jets $\Delta\phi_{1,2}$ is measured. In addition, the p_T differential cross-section of the four leading p_T jets is measured up to the TeV scale. As a function of the jet multiplicity, the cross section showed contributions of up to seven jets in the three different regions of $\Delta\phi_{1,2}$ measured at different p_{T1}. Of especial interest was that up to seven jets are measurable in the back-to-back region both at low and high p_{T1}. The measurement of the four leading jets p_T is an important test for calculations including parton showers. These multi-differential jet measurements are presented with their corresponding (total) correlation matrix, which allowed the identification and quantification of relatively unavoidable migration effects between the two leading jets, the third and forth, and between the p_{T1} bins in the jet multiplicity measurement.

The LO multijet MadGraph+Py8 and pythia8 simulations reasonably describe the measured multiplicity distribution, nonetheless in the back-to-back region herwig++ provides a better shape description. Nevertheless, pythia8 was found to be 11% above the data (from the normalization factor), giving the best LO total (dijet) cross section prediction, even better than the LO (MLM) merged MadGraph+Py8 (25% below the inclusive dijet cross section). The measured differential cross section as a function of the transverse momen-

tum of the four leading p_{T} jets is not well described (in the entire phase space) by any of the LO predictions. For the two leading jets, PYTHIA8 gave a better shape prediction than HERWIG++ (MADGRAPH+PY8), which at high p_{T} underestimated (overestimated) the shape of the measurement. For the third and fourth leading jet, HERWIG++ and PYTHIA8 predictions were comparable in shape, and only MADGRAPH+PY8 showed again overestimation for the third jet p_{T}.

The description of the lower jet multiplicity cross section obtained with NLO dijet calculations supplemented with conventional parton shower or with PB-TMDs and TMD parton shower are rather good. The higher jet multiplicities are not described with either parton shower approach since there we still need the contribution from Matrix Elements like the one used in the MG5_aMC+CA3 (jjj) NLO. The low $\Delta\phi_{1,2}$ region and the two leading jet p_{T} helped us to understand ME effects. In contrast, the high jet multiplicity in the back-to-back region and the third and fourth jets p_{T} were more sensitive to the parton showers. The measurements presented here allowed very stringent tests of theoretical predictions in the perturbative high p_{T} and high jet multiplicity regions.

The measurements presented in this thesis constitute a benchmark for Standard Model multijet cross section calculations. In addition to previous $\Delta\phi_{1,2}$ measurements in bins of p_{T1}, these measurements bring new insights on how many jets build the $\Delta\phi_{1,2}$ decorrelation at different scales. The back-to-back region was of special interest since more than seven jets can build the decorrelation and when going to NLO the parton shower is not enough and we need higher order 3-jet NLO+PS for describing such scenario.

Outlook and future measurements

The presented measurements will be of particular importance when testing multijet merged QCD predictions to describe the measurements better since they use higher-order corrections together with patron showers. For Run-III at the LHC more data can be collected and going more differential will be great for exploring other new corners of the phase space never measured before. In the future, will be of great interest to measure the dijet system transverse momentum ($p^T_{1,2}$) in bins of $\Delta\phi_{1,2}$, since in the back-to-back region we could further understand the soft radiation effects, and at low $\Delta\phi_{1,2}$ we would be sensitive to hard radiation effects. In addition, a measurement of the cross section as function of the four leading jets p_{T} in bins of $\Delta\phi_{1,2}$ would be of interest.

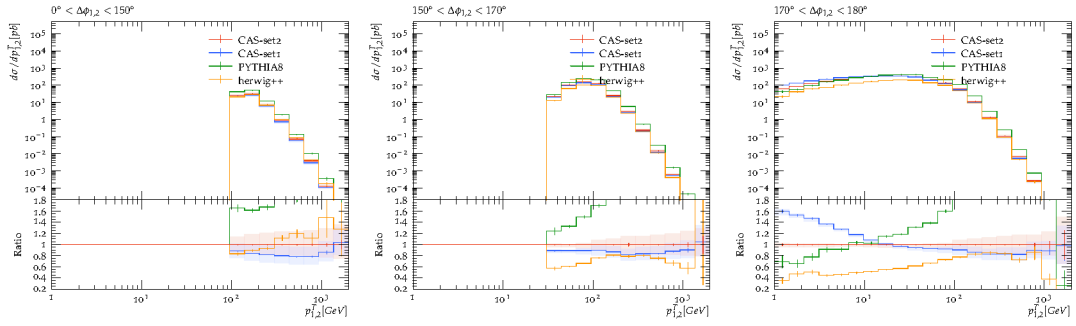


Figure 9.1: Dijet system p_{T} ($p^T_{1,2}$) in bins of $\Delta\phi_{1,2}$.

In Fig. 9.1 the $p^T_{1,2}$ (using the same event selection at particle level as presented in Sec. 4.3)

NLO+PS predictions from CAS-set2 (used as default in the predictions presented in this thesis) and CAS-set1 are shown. Both CASCADE3 predictions differ only in the TMD used: set1 and set2 were build using angular ordering but the argument of α_S is the scale μ_i^2 for set1 and $q_i^2 = (1 - z_i)^2 \mu_i^2$ for set2. The difference in the α_S scale definition can be seen in at low p_{T} for the back-to-back region when comparing CAS-set2 to CAS-set1.

These proposed measurements, among others, could help us to improve the understanding of QCD which is the major background of $W/Z(+jets)$ [149] and $t\bar{t}(+jets)$ [150] measurements in the full hadronic channel. The understanding of the QCD background also becomes important for new physics Higgs searches [151] and heavy resonances (X) in $X \to t\bar{t}$ [152] and $X \to VH$ [153, 154] for hadronic final states at the LHC.

The measurement of $p_{1,2}^T$ and the four leading jet p_{T} spectra in bins of $\Delta\phi_{1,2}$ not only have never been measured before but are of great interest to study QCD in more detail. For the dijet system $p_{1,2}^T$ small-x effects [155] can be studied in the small $0^\circ < \Delta\phi_{1,2} < 150^\circ$ region where a lot of extra jets are present. Also in the back-to-back region, where soft gluon resummation have a large contribution, $TMD-$factorization breaking effects [156–158] may arise. Therefore when comparing these measurements with predictions based on the factorization ansatz, we might find insights on this theoretical issue, which can be shown to be non-consistent with QCD being a gauge theory [156]. The upcoming Run-III with a potentially higher center-of-mass-energy of $\sqrt{s} = 14\,\mathrm{TeV}$ will bring new challenges and interesting precision measurements. With new challenges, the answers to many theoretical issues may arise. Not so far away, the high luminosity LHC [159] project, which would start around 2027, is expected to collect approximately 3 ab^{-1} (more than six times the luminosity collected by the LHC up to 2018). The study of this massive amount of data will further push our understanding of QCD and the SM in general, with potentially new findings and fascinating physics waiting for us.

Appendices

APPENDIX

A

INTRODUCTION TO MC INTEGRATION

In this appendix the MC techniques used in Monte Carlo Event Generators are discussed as additional material to Sec. 2.3. The discussion is an overview that goes from basic principles to some more advanced MC techniques used in MC simulations.

The basis of Monte Carlo integration methods relies on the "law of large numbers" [160,161]. This law states that for uniformly distributed random values x_i in the interval $[a, b]$ the sum of the probability density functions converges to the true estimate of the mean of the function $f(x)$:

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \to \frac{1}{b-a}\int_a^b f(x)dx \tag{A.1}$$

where $f(x)$ must be integrable inside $[a, b]$. For many random values, the MC estimates get closer or become a better estimate of the true integral, and for $N \to \infty$ it converges to the true integral. In addition, the Central Limit Theorem tells us how the convergence goes for a finite number of random variables N. It says that the sum of a large number of random variables follows a normal (Gauss) distribution no matter if the variables are generated uniformly or with any other probability density function (p.d.f) defined as:

$$\int_{-\infty}^{\infty} g(x)dx = 1 \tag{A.2}$$

where $g(x)$ gives the probability to find the values x in the interval $[x, x + dx]$. Then according to the Central Limit Theorem to obtain a Gauss distribution from a given number of random variables (x_i):

$$\frac{\sum_i x_i - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}} \to \mathcal{N}(0, 1) \tag{A.3}$$

Uncertainties as function of N	for 1 dim	for d dim
Monte Carlo	$N^{-1/2}$	$N^{-1/2}$
Trapezoid method	N^{-2}	$N^{-2/d}$
Simpson's method	N^{-4}	$N^{-4/d}$

Table A.1: MC integration efficiency compared to other integration methods.

here μ_i is the mean and σ_i is the variance squared of the random numbers. Then for example if one sums $n = 12$ numbers many times ($N \to \infty$) one obtains a normal distribution $\mathcal{N}(0, 1)$:

$$\mathcal{N}(0,1) \to \frac{R_n - n/2}{\sqrt{n/12}} = R_{12} - 6 \tag{A.4}$$

here $R_n = \sum_{i=1}^{n} x_i$

Then eq. A.1 can be obtained from the expectation value of the function $f(x)$ with the values distributed according to a uniform p.d.f $g(x)$ in the range $[a, b]$ (and zero in the rest), as follows:

$$E[f] = \int_{-\infty}^{\infty} f(x)g(x)dx = \frac{1}{b-1}\int_a^b f(x)dx \to \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{A.5}$$

The error of the MC integration or MC estimator is given by:

$$\sigma = \frac{\sqrt{V[f]}}{\sqrt{N}}, \text{ with variance } V[f] = E[(f(x) - E[f(x)])^2] = \int (f(x) - E[f(x)])^2 f(x)dx \tag{A.6}$$

where the variance can be defined for discrete x_i values as:

$$V[f] = \frac{1}{N-1}\sum_{i=1}^{N}[f(x_i) - \frac{1}{N}\sum_{i=1}^{N} f(x_i)]^2, \text{ where } E(f) = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{A.7}$$

From eq. A.6 one sees that by increasing the sample size (N), one gets more accurate results. Another solution would be to reduce the variance, but the first is simpler. Then if one compares the MC integration method with other numerical integration methods as in Tab. A.1, one can notice that for multidimensional integration with more than four dimensions ($d > 4$), the MC method is more efficient and gives a minor error. In practice, one finds pretty large multidimensional integrals when calculating observables in a given phase space, and the computational strength of the MC method makes it not only the most powerful one but the only choice to calculate complicated integrals.

Generating distributions

The x_i values in the MC integration are generated according to the $f(x)$ distribution. In general using a random number generator (details on random number generators can be found in Ref. [162]) in the interval $[0, 1]$ one can generate random values according to a distribution $f(x)$:

$$\int_a^b f(x)dx = R\int_a^x f(t)dt \tag{A.8}$$

here R is a random number. A practical example would be to generate values for $f(x) = 1/x$, and by using eq. A.8 one gets:

$$x_i = a\left(\frac{b}{a}\right)^{R_i} \tag{A.9}$$

But this method works when we know $f(x)$ is an analytic integrable function. If not, one can use the brute force hit and miss method: accepting the value if $f(x_i) < R_j \cdot f_{max}$ and rejecting for $f(x_i) > R_j \cdot f_{max}$. For this method we need to select first x_i in the range $[a, b]$ with $x_i = a + (b - a)R_i$ so in general one uses two random numbers R_i and R_j in the interval $[0, 1]$. This method can be improved by using a similar function to $f(x)$ but integrable $g(x)$ function, and then the selection condition would be $f(x_i) < R_j \cdot c \cdot g(x)$ with c being a constant such that $c \cdot g(x) > f(x)$ for all x in $[a, b]$ and generating x_i values according to $g(x)$.

Improved MC methods

We already saw that the integration error could be improved by reducing the variance or increasing the sample of generated values. These are some approaches to improve the accuracy of the MC integration which will be relevant in further discussions in this thesis:

Importance sampling In this method one can use a $g(x)$ p.d.f similar to $f(x)$ and approximate the integral:

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{g(x)} g(x)dx = E\left[\frac{f(x)}{g(x)}\right] \tag{A.10}$$

with $g(x)$ being normalized and integrable in $[a, b]$. The values x_i in $[a, b]$ can be then generated according to $g(x)$ by using eq. A.8 (or using brute force method as discussed above). Then the integral would be calculated as:

$$I = \frac{1}{N} \sum \frac{f(x_i)}{g(x_i)} \tag{A.11}$$

This method reduces the variance but can be dangerous if used when $g(x)$ becomes zero or approaches zero quickly [162].

Subtraction method This method (also called control variates in Ref. [162]) is useful and especially successful if $f(x)$ has a divergent part (this method is often used in NLO QCD calculations). By using $g(x)$ function which approximates the $f(x)$ function:

$$\int f(x)dx = \int g(x)dx + \int (f(x) - g(x))dx \tag{A.12}$$

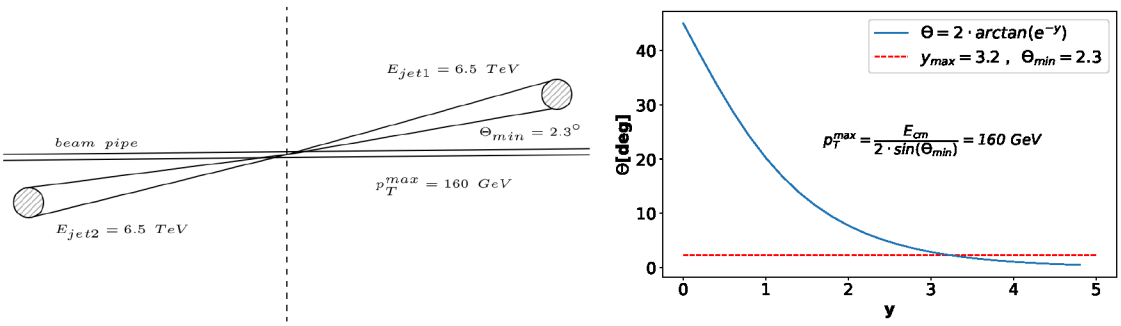
This way if one knows the integral of $g(x)$ the uncertainty comes from the integral of $(f(x) - g(x))$ which will have a smaller variance than $f(x)$ if the proper $g(x)$ function was chosen. This method is more stable than importance sampling explained before since zeros in $g(x)$ will not affect the calculation, given that $g(x)$ cannot induce singularities on $(f(x) - g(x))$.

APPENDIX

B

PRESELECTION STUDIES

In this appendix we study the effect of performing a preselection of jets with $|y| < 5.0$ in MC and in data, instead of $|y| < 3.2$ which is used in this thesis measurements (see Sec. 4.3). Having a look at Fig. B.1 on the left there is a representation of a two jet event (assuming energy-momentum conservation) in the kinematical limit of two jets with 6.5 TeV each the maximum $p_T^{max} = 160\,\text{GeV}$, in the right figure the p_T^{max} is calculated and the polar angle Θ is plotted as a function of the rapidity y. All this is tanking into account the high energy limit for which $y \approx \eta$ and $E \approx |\mathbf{p}|$.



(a) Representation of a two jets event in the pseudo rapidity plane.

(b) Polar angle Θ as function of the rapidity y.

Figure B.1: Two jet event with the maximum p_T allowed at $\sqrt{s} = 13$ TeV in the CMS detector in the pseudorapidity plane for the angle $\Theta_{min} = 2.3^{\circ}$ corresponding to a rapidity of 3.2.

This would mean that in general, at least at particle level for high leading jet $p_{T1} > 200\,\text{GeV}$ the cut on $|y| < 3.2$ should make no difference from $|y| < 5.0$. Then in Fig. B.2 and Fig. B.3 particle level of PYTHIA8 is plotted for both preselections, for the jet multiplicity and the four leading jets transverse momenta. It is found only a small difference of about 10% for the exclusive two jets at low $\Delta\phi_{1,2}$ and low p_{T1} for the jet multiplicity in Fig. B.2,.For the four

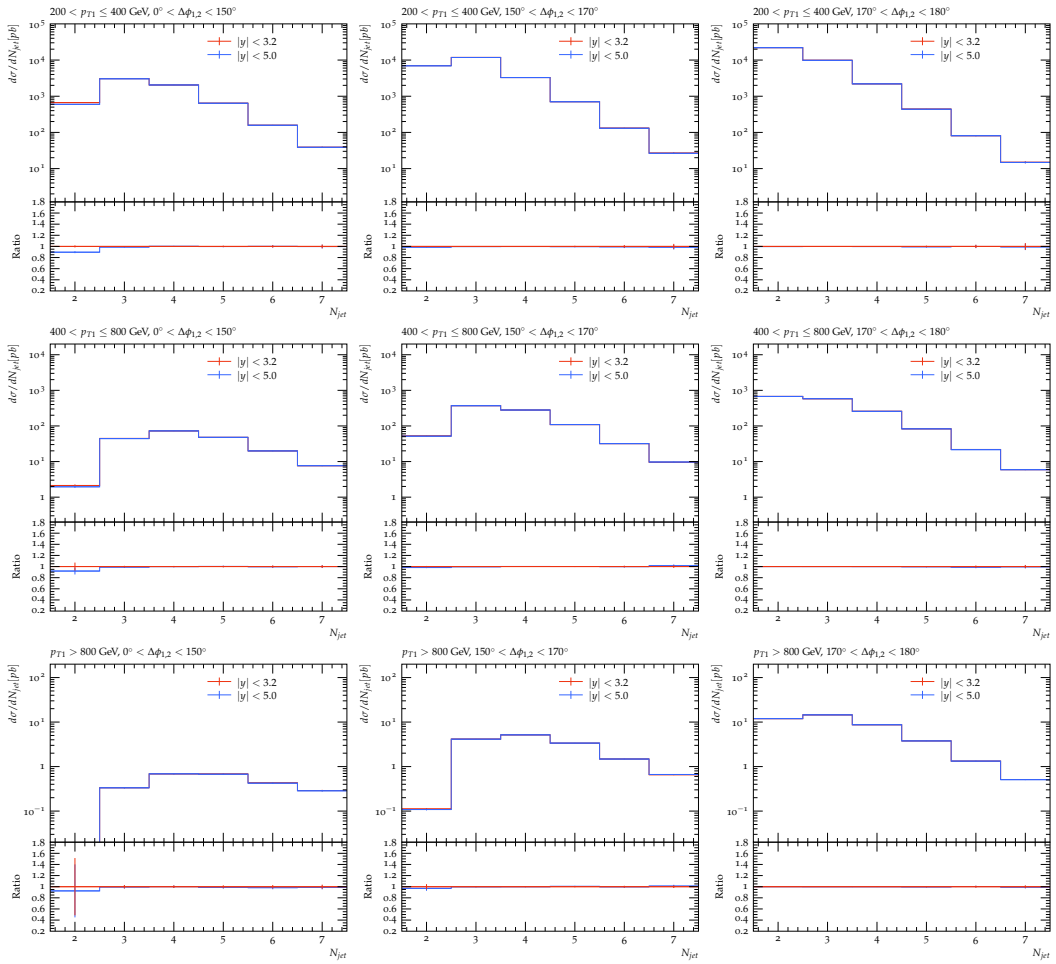


Figure B.2: Jet multiplicity at particle level for PYTHIA8 using $|y| < 5$ and $|y| < 3.2$ preselection.

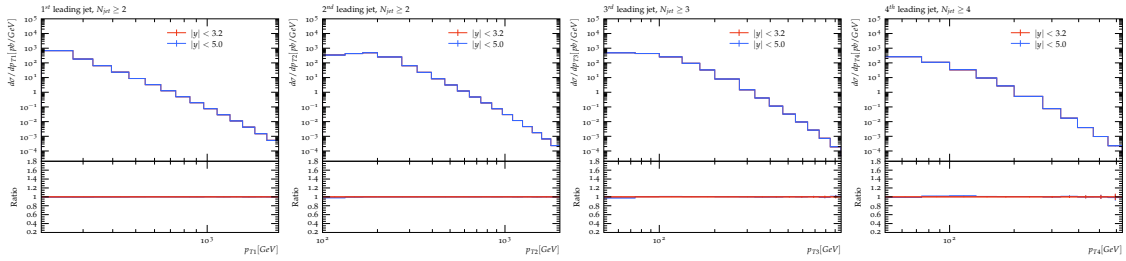


Figure B.3: Four leading jets p_{T} at particle level for PYTHIA8 using $|y| < 5$ and $|y| < 3.2$ preselection.

leading jets p_{T}, in Fig. B.3 the effect is almost undetectable, we only see tiny differences at low p_{T} for the p_{T3} and p_{T4}.

In addition, in the Fig.B.4 we can see that the effect is essentially the same in data and MC at detector level for the jet multiplicity observable. From this study, we conclude that using $|y| < 3.2$ does not change the physics, and one could draw the same physics conclusions as if one uses a larger acceptance in the preselection with $|y| < 5.0$.

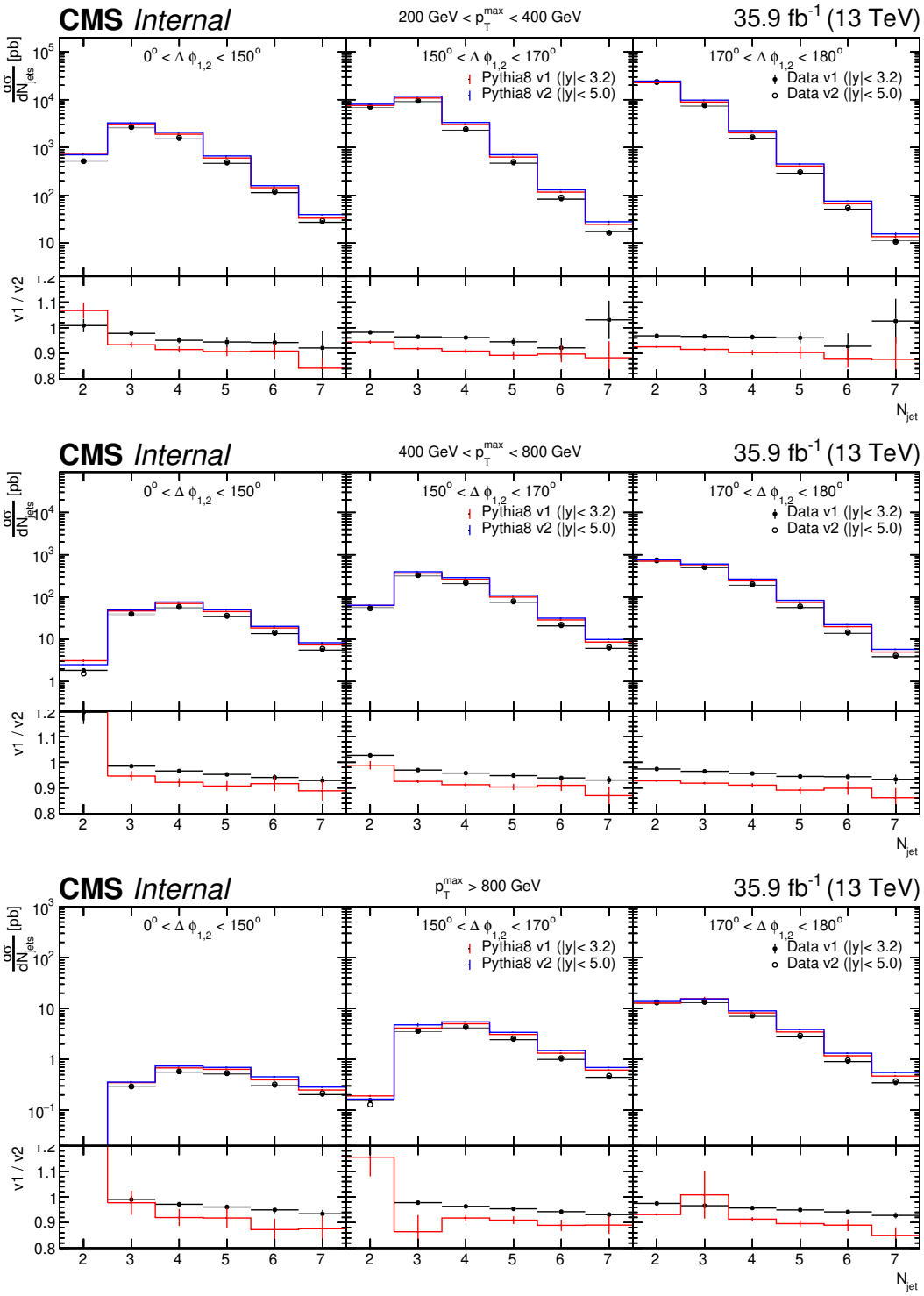


Figure B.4: jet multiplicity(N_{jets}) for dijet events at detector level in bins of leading jet p_T (p_T^{max}) and $\Delta\phi$, in the tracker acceptance $|y| < 2.5$. Here v1 refers to the default preselection in the analysis $|y| < 3.2$ and v2 refers to the preselection of $|y| < 5$.

APPENDIX

C

BACKGROUND STUDIES ON THE MEASURED OBSERVABLES

In Sec. 4.5 the background subtraction was presented. This appendix presents the effect of the MET fraction cut in each observable. In Fig. C.1 the jet multiplicity without MET fraction cut (left column) and with MET fraction cut (right column) is shown with all QCD and non-QCD contributions.

The QCD signal represents more than 98% in each p_{T1} bin, the non-QCD events represent less than 2%, as shown in the legend. In the figures comparing the fraction of non-QCD events for each jet multiplicity bin (red text in %), one can see that the MET fraction cut reduces the non-QCD contribution, especially in the 2-jets multiplicity bin (where a large MET fraction is expected) for the two last p_{T1} bins (400-800 and more than 800) where the reduction is about 7% and 12%. After the MET fraction cut, the non-QCD contribution in each jet multiplicity bin is in the same order as the statistical uncertainty in data (not shown in the plots).

For the jet p_{T} measurements, the effect of the MET fraction cut is not so visible (see Fig. C.2). It might be seen at high p_{T} for the leading jet. In this case, we can see in the ratio in Fig. C.2 b, for instance, that the non-QCD contributions become more important (larger than 1%) at high p_{T} for the two leading jets (about 1 TeV) and the extra jets (about 400 GeV). Still, the statistical uncertainty from data (blue line) starts to be comparable with the non-QCD contributions. We do not use the non-QCD contributions in the analysis in the following. After the MET fraction cut, the contamination from non-QCD events to the QCD signal is comparable with the data statistical uncertainty and much smaller than the QCD signal itself. It is also seen that the differences between the LO QCD samples (PYTHIA8, MADGRAPH, and HERWIG++) are more significant than the non-QCD contributions.

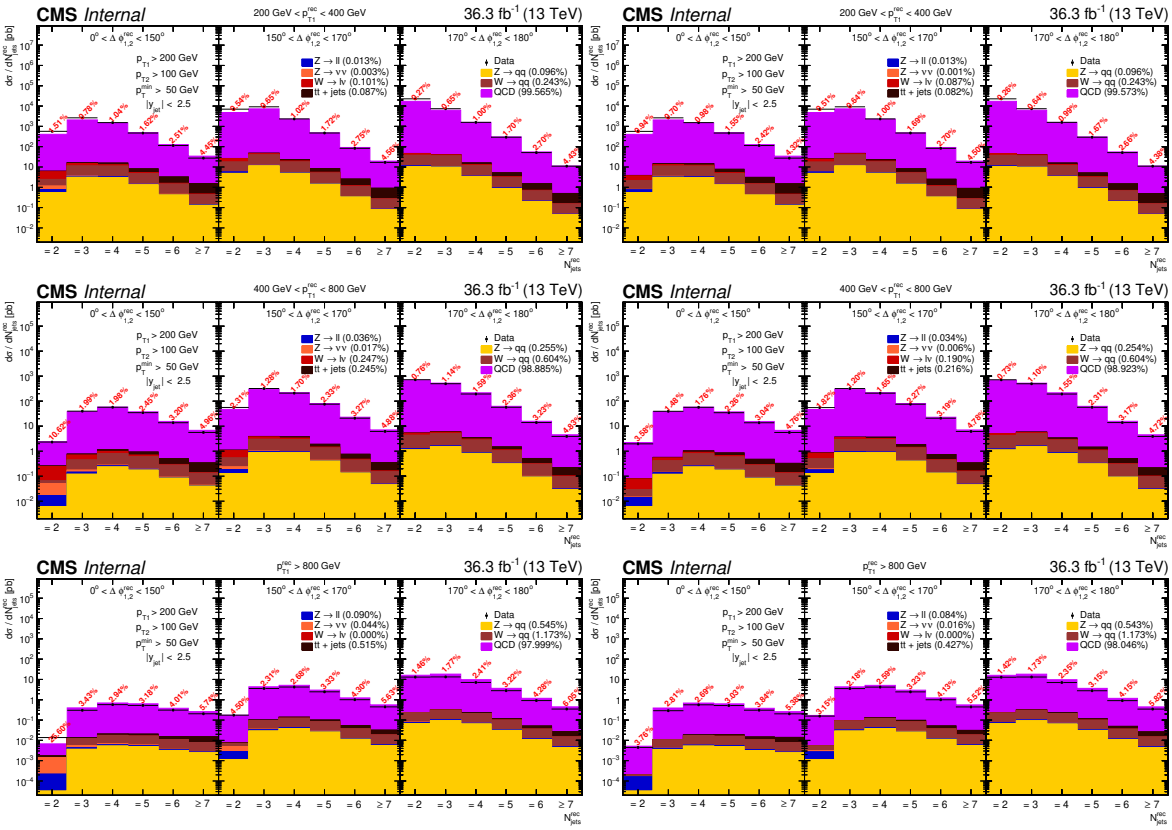
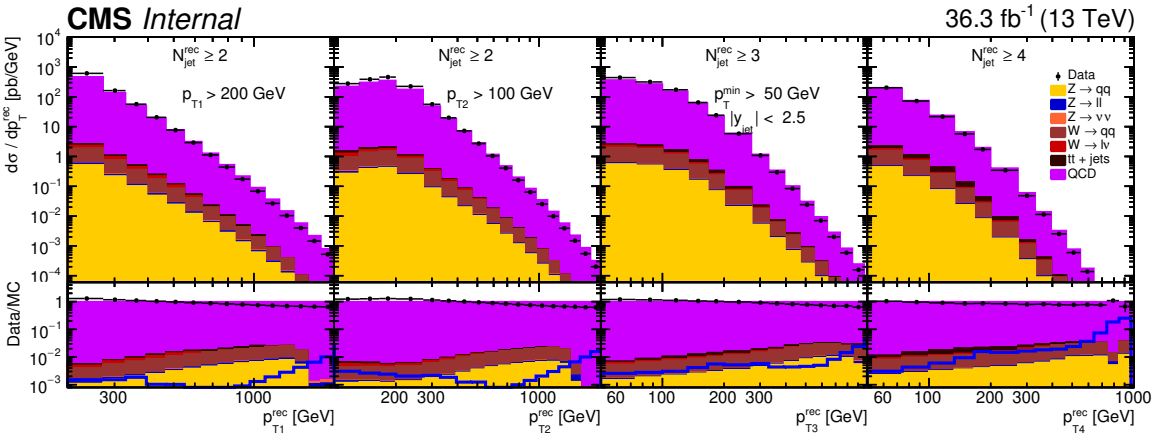
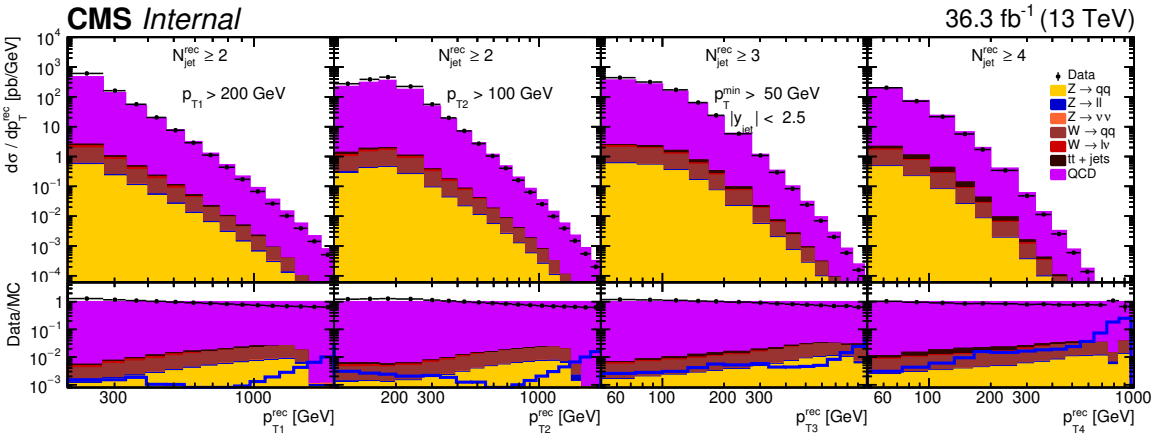


Figure C.1: Jet multiplicity distribution without MET fraction cut on the left, and with MET fraction cut of 0.1 on the right for all p_{T1} bins (at detector level). Contributions from non-QCD processes tt+jets, W/Z+jets(including $Z \to \nu\nu$ and $W \to l\nu$), and QCD are shown. Each MC contribution has the fraction of events from the total for each p_{T1} bin (in %) in the legend, above each jet multiplicity bin it is shown the fraction of non-QCD events($N_{non-QCD}/(N_{QCD}+N_{non-QCD})$ in %) with red text.



(a) Without MET fraction cut



(b) With MET fraction cut

Figure C.2: Transverse momentum of the first four leading jets with all MC contributions compared to data (at detector level). The blue line in the ratio represents the statistical uncertainty from data.

APPENDIX

D

PYTHIA AND MADGRAPH SAMPLES IN SLICES

In Sec. 5.1 the data and MC samples were presented. In this appendix, the cross section and luminosities of the PYTHIA8 and MADGRAPH samples $\hat{p_{\mathrm{T}}}$ and H_T slices are shown in Tab. D.1. In addition the different slices contribution with respect to the particle level p_{T} spectra are presented in Fig. D.1.

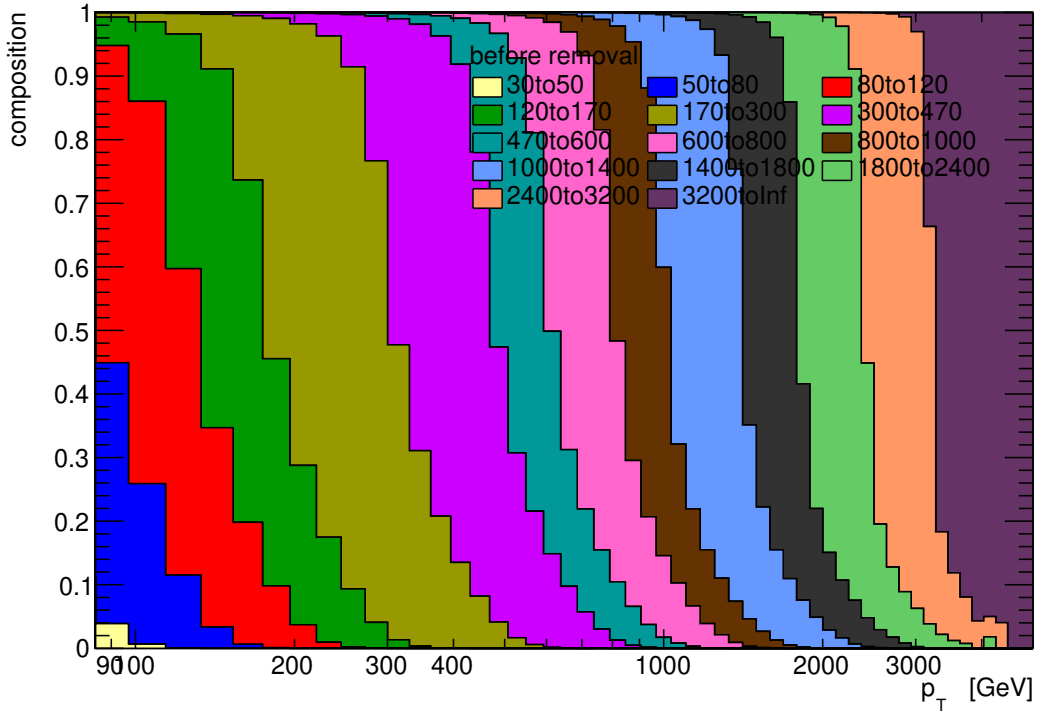
slice	# events	cross section [pb]	luminosity [pb^{-1}]
15→30	35613665	1821000000	0.01956
30→50	9980010	138800000	0.0719
50→80	9954330	19110000	0.5209
80→120	14161654	2735000	5.178
120→170	12616118	466200	27.06
170→300	14796734	117200	126.3
300→470	22470364	7763	2895
470→600	3959946	641	6178
600→800	13524707	185.7	72830
800→1000	19697052	32.02	615100
1000→1400	9846575	9.375	1050000
1400→1800	2873387	0.8384	3427000
1800→2400	1981998	0.1133	17490000
2400→3200	996090	0.006746	147700000
3200→ ∞	391689	0.0001623	2413000000

(a) PYTHIA8 in slices of $\hat{p}_T$

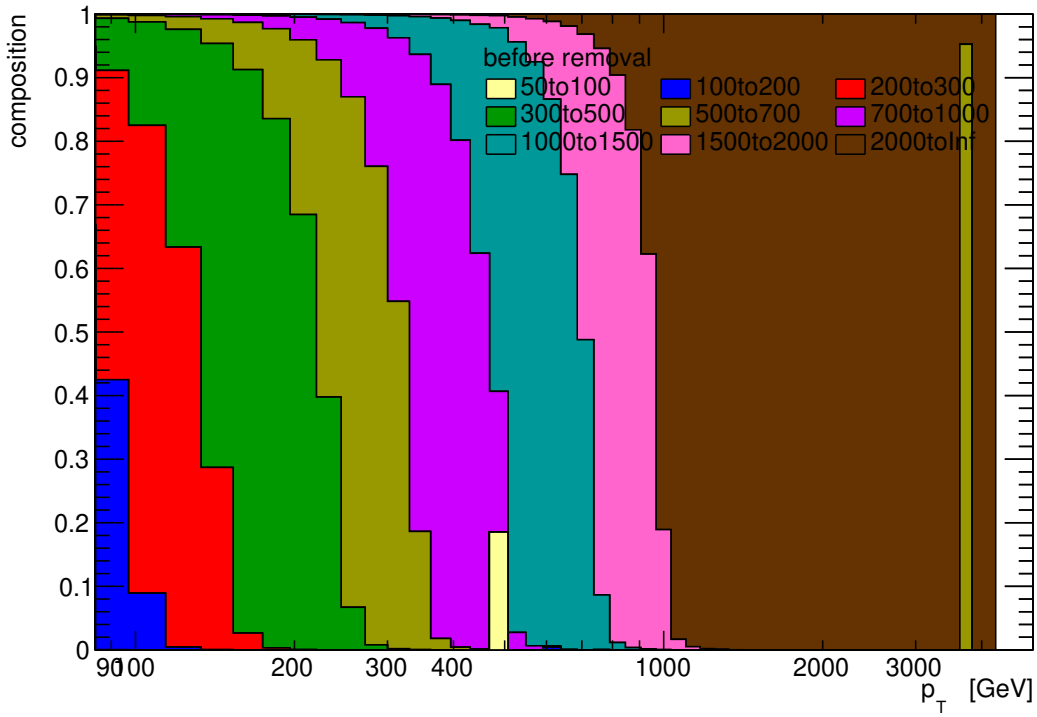
slice	# events	cross section [pb]	luminosity [pb^{-1}]
50→100	4180423	246400000	0.01697
100→200	71207048	27940000	2.549
200→300	56130206	1712000	32.79
300→500	54552806	347700	156.9
500→700	45770532	32150	1424
700→1000	15629207	6828	2289
1000→1500	15210893	1200	12680
1500→2000	11839311	120	98660
2000→ ∞	6019495	25.34	237500

(b) MADGRAPH in slices of H_T

Table D.1: Luminosity of PYTHIA8 and MADGRAPH samples in slices.



(a) PYTHIA8 in slices of $\hat{p}_T$



(b) MADGRAPH in slices of H_T

Figure D.1: Contributions from the different slices to the particle level p_T distributions directly after proper normalization of the slices by corresponding MC luminosities.

APPENDIX

E

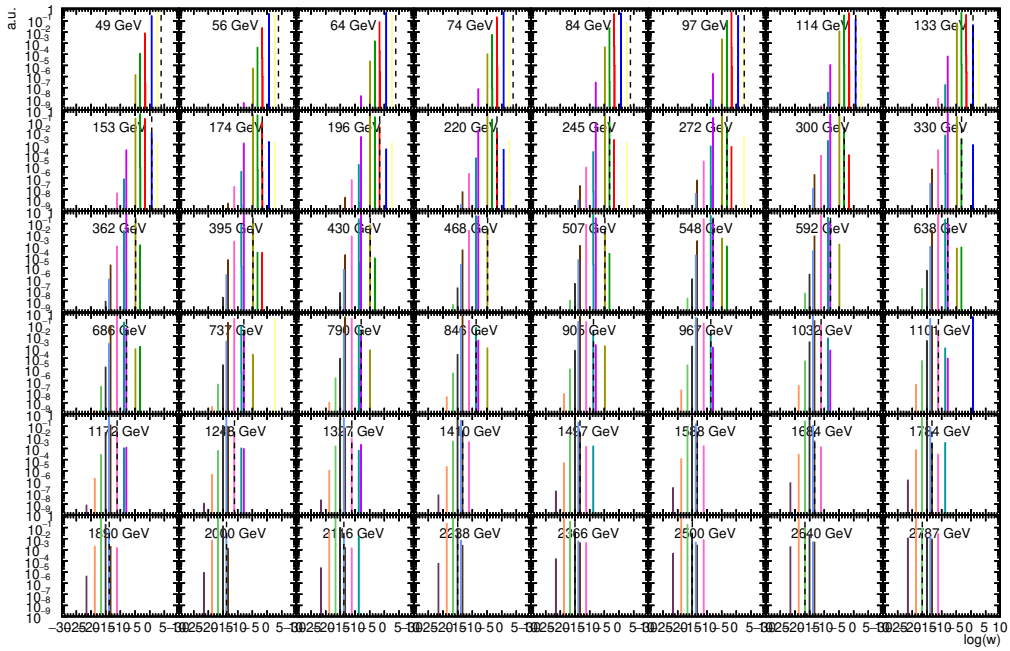
DETERMINATION OF THE MAXIMAL WEIGHT FOR EACH P_T^{REC} BIN

In this appendix, the determination of the maximal weight for each p_T^{rec} bin will be presented. As discussed in Sec. 5.2.3 two procedures are used to determine the overweighted events caused by the simulation of the pileup. Both procedures will be described in this appendix in more detail.

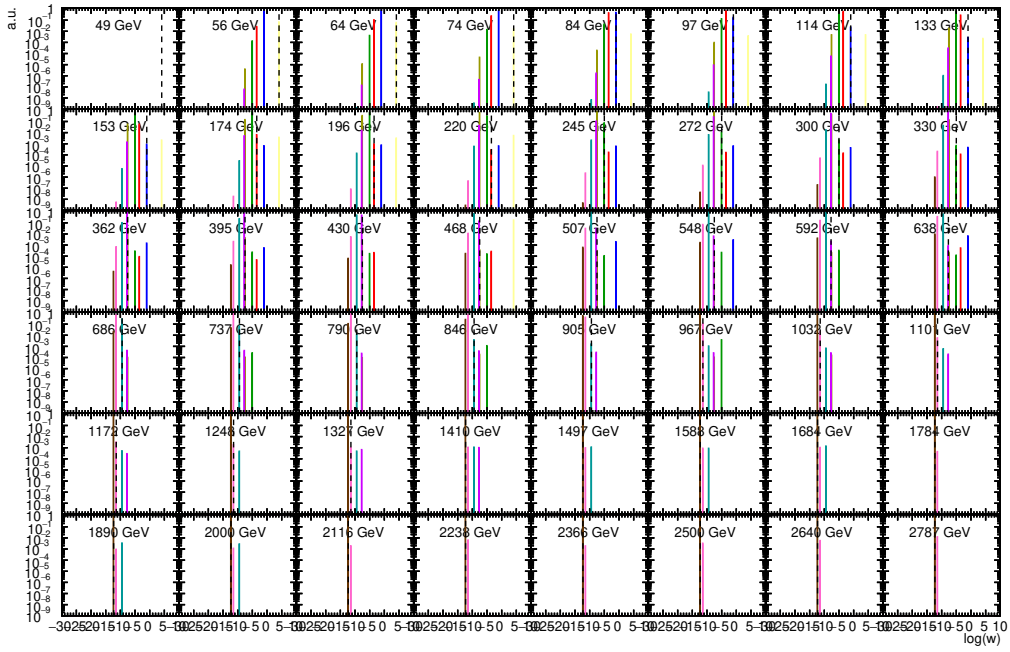
To determine the maximal weight for each p_T^{rec} bin, we consider only the leading jets and plot the histogram of the logarithm of the event weight for each p_T^{rec} bin. Then two different methods bin exist:

Samples in slices: We ensure that a low $\hat{p}_T$ slice cannot contribute over any higher $\hat{p}_T$ slice; in addition, we reject contribution to bins with less than a hundred entries the distributions are shown in Fig. E.1.

Flat samples: We find that a fit with a double Gaussian describes well the distribution, and cut off events above four sigmas to the right of the second Gaussianthe distributions are shown in Fig. E.2.

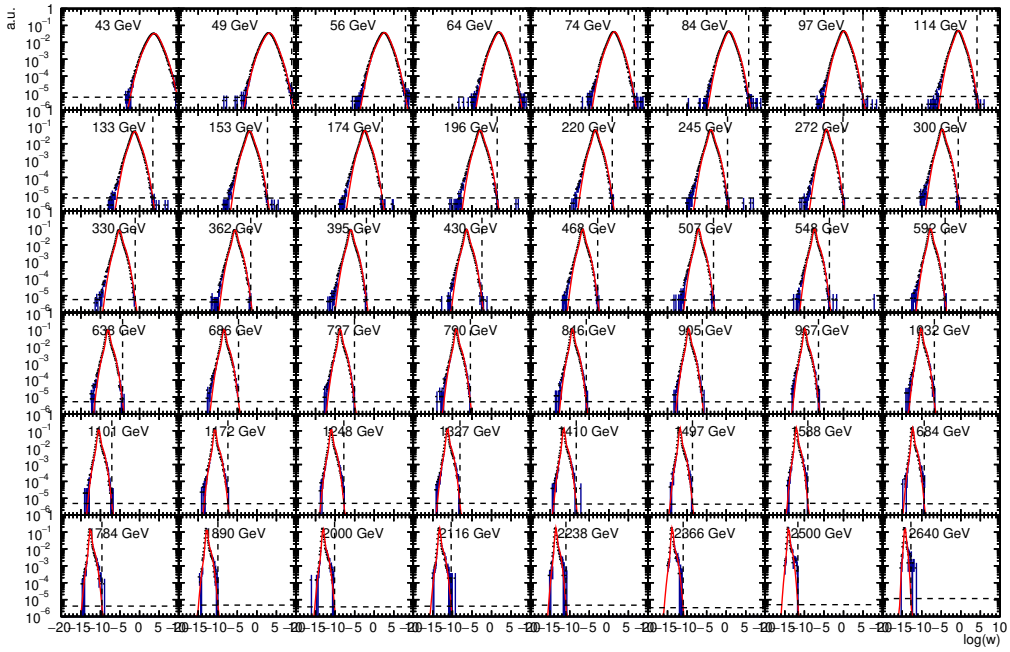


(a) PYTHIA8 in slices

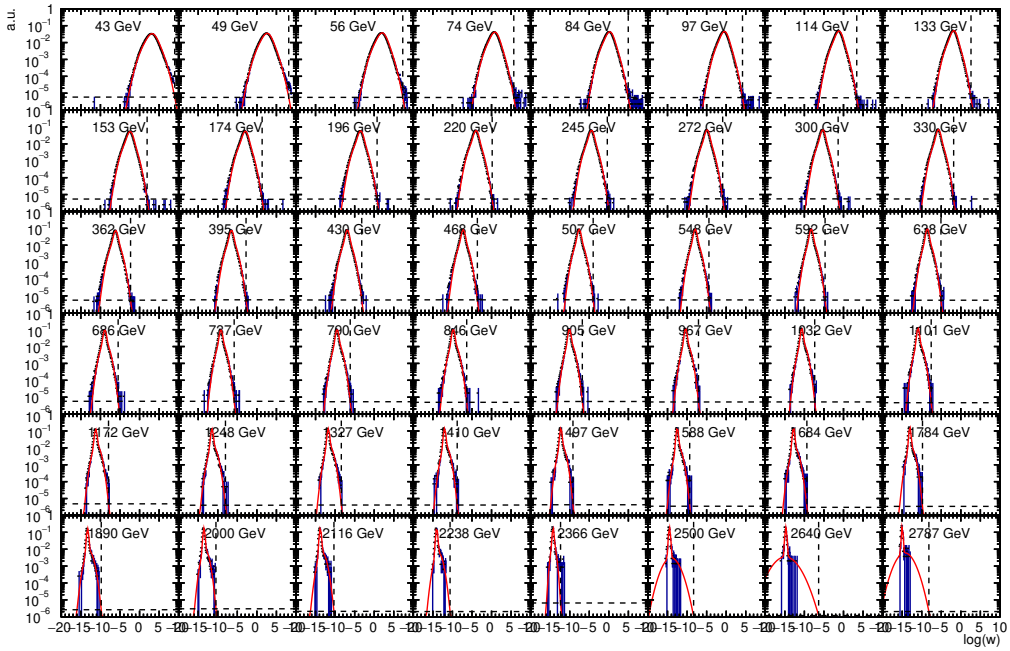


(b) MADGRAPH

Figure E.1: Logarithm of the event weight in the flat samples shown for each bin of reconstructed transverse momentum of the leading jet. Each color corresponds to a different slice (see explanation in Sec. 5.1). The number of entries corresponds to the effective number of entries (*i.e.* weighted proportionally to the cross section), normalized to unity in each cell. The dashed vertical line separates on the left (right) the slices that are used (rejected) to fill the reconstructed bin.

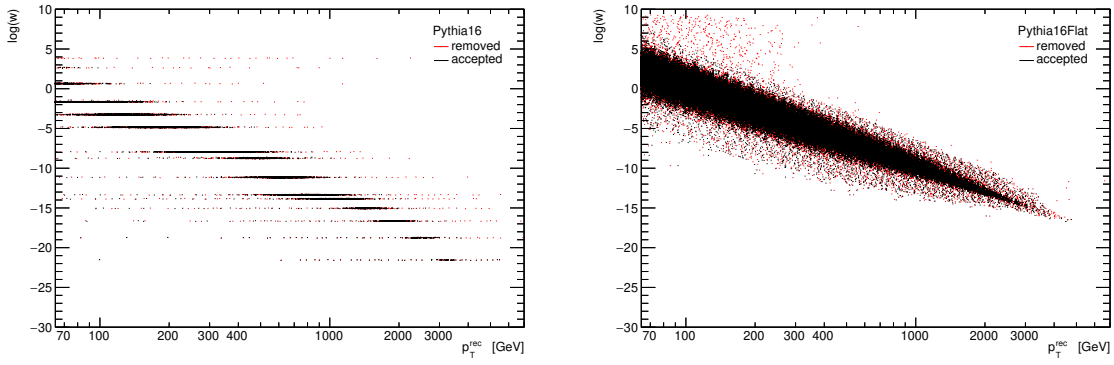


(a) PYTHIA8 flat



(b) HERWIG++

Figure E.2: Logarithm of the event weight in the flat samples shown for each bin of reconstructed transverse momentum of the leading jet. The number of entries is normalized to unity in each cell. The red line corresponds to a fit with a double Gaussian function (six parameters in total). The dashed vertical and horizontal lines correspond to four sigmas to the right of the second Gaussian, where the cut-off is considered.



(a) PYTHIA8 sample in slices.

(b) PYTHIA8 flat sample.

Figure E.3: Scatter plot of the reconstructed transverse momentum of the leading jet and logarithm of event weight (in the legends of the figures 16 stands for the 2016 year).

The effect of the procedure is also shown on a scatter plot in Fig. E.3 for the two PYTHIA8 samples. Here the sample is not divided into bins of the reconstructed p_{T} of the leading jet. If one rotates the y-axis in this plot and normalizes each p_{T} bin to unity, one gets a similar plot to what is shown before in Fig. E.1-E.2 for the slice and flat samples.

APPENDIX

F

PILEUP CLEANING EFFECTS

This appendix contains all the additional material to what was discussed in Sec. 5.2.3.

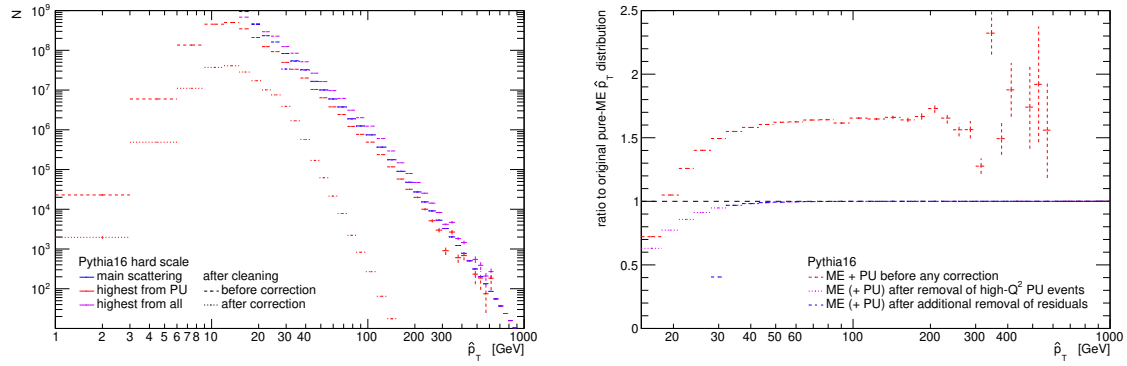
The pileup cleaning effects on the hard scale (p_T spectrum) at the different stages of the procedure is shown in Fig. F.1 (Figs. F.2-F.3), with PYTHIA8 in slices (with both PYTHIA8 and MADGRAPH samples):

- the red color stands for the highest scale of the PU events, while the two different styles show the contribution before and after the correction;
- the blue color stands for the scale of the "main" event (*a.k.a.* leading-vertex event), which nearly does not change with the effect of the correction.
- the violet color stands for the highest scale of all events, which, after correction, matches with the scale of the main event, as expected.

The impact on the detector-level p_T spectrum is huge, since the original normalization was not correct, as explained in Sec. 5.2.3.

F.1 PU cleaning effect on the measured observables in this thesis

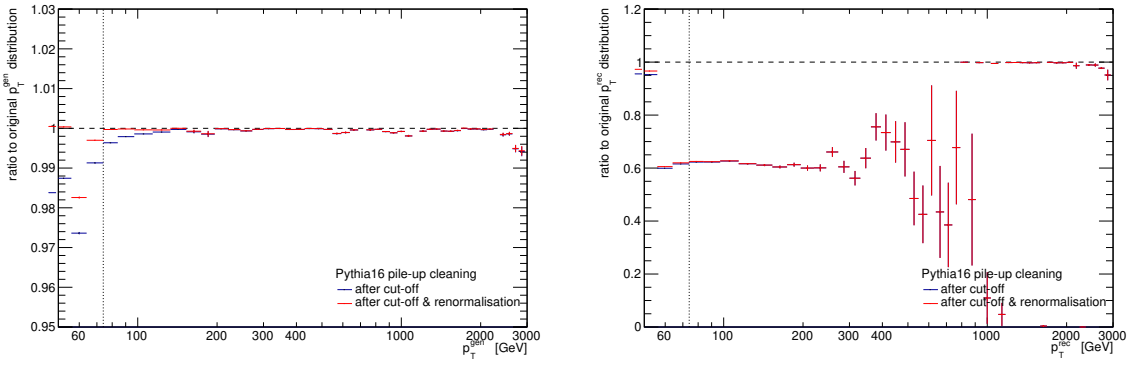
Here the PU cleaning effect is shown for the measured observables in this thesis. This plots are complementary to what was shown in Sec. 5.2.3. In Fig.F.6, F.7, F.8 and Fig.F.9 the predictions from PYTHIA8 for the jet multiplicity and the p_{T} spectra of the four leading jets are shown. In addition, the remaining p_{T1} bins for jet multiplicity from the MADGRAPH prediction are shown in Fig. F.4-F.5.



(a) Hard scales before and after applying a cut off based on the weight and on the transverse momentum.

(b) Ratio to original distribution of hard scale of inclusive jet production.

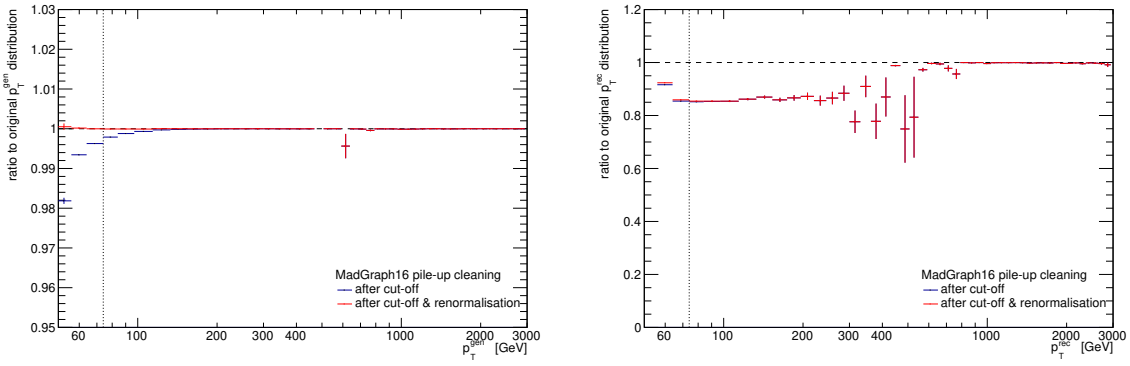
Figure F.1: Hard scales of the main scattering and of the PU scatterings present in a single bunch crossing, in PYTHIA8 sample.



(a) Generator level.

(b) Detector level.

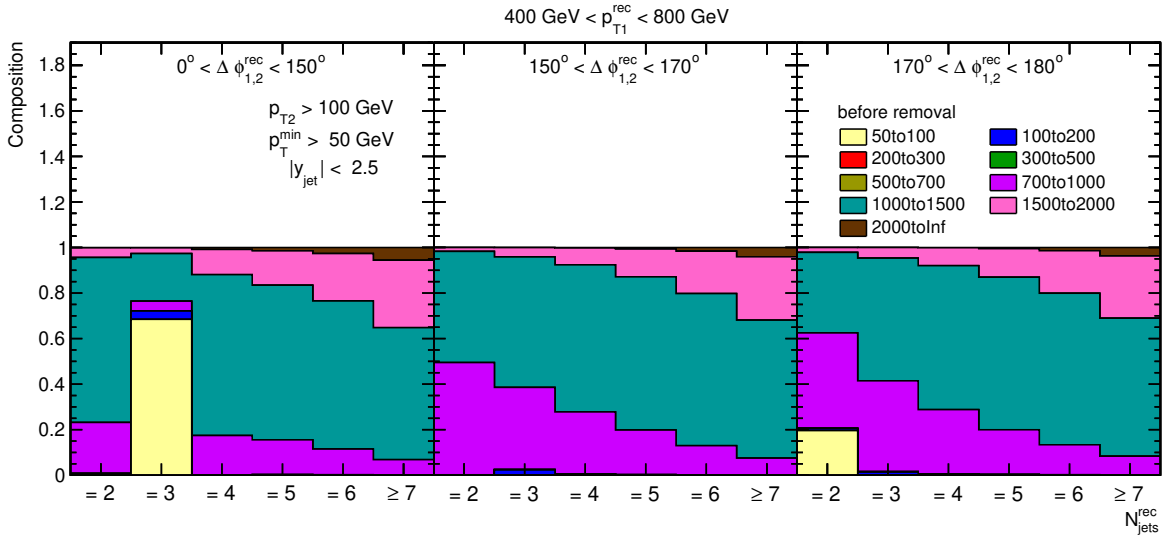
Figure F.2: Effect of PU cleaning on p_T distributions at both levels with PYTHIA8.



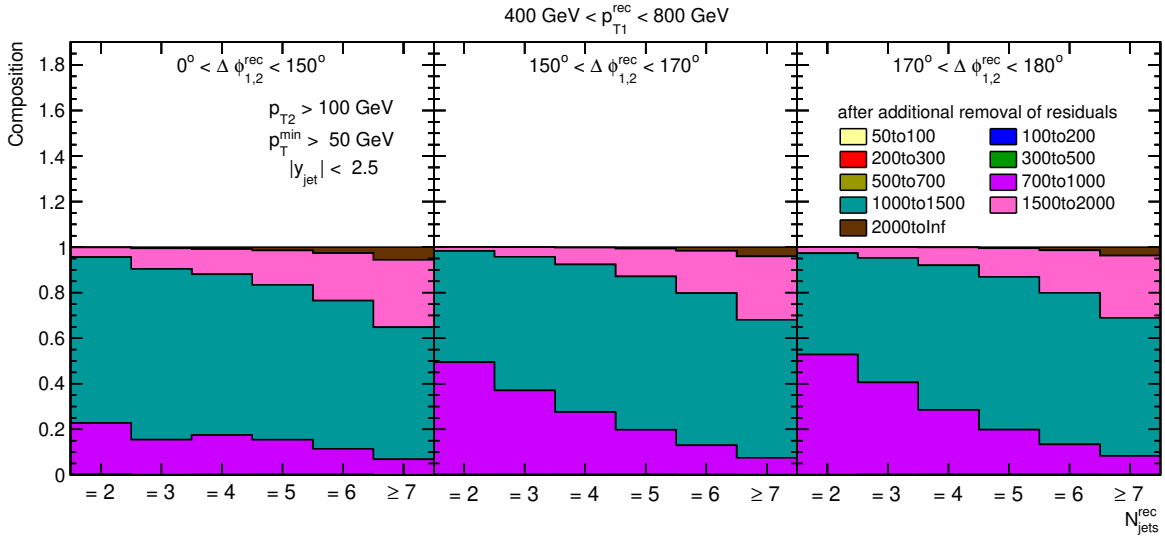
(a) Generator level.

(b) Detector level.

Figure F.3: Effect of PU cleaning on p_T distributions at both levels with MADGRAPH.

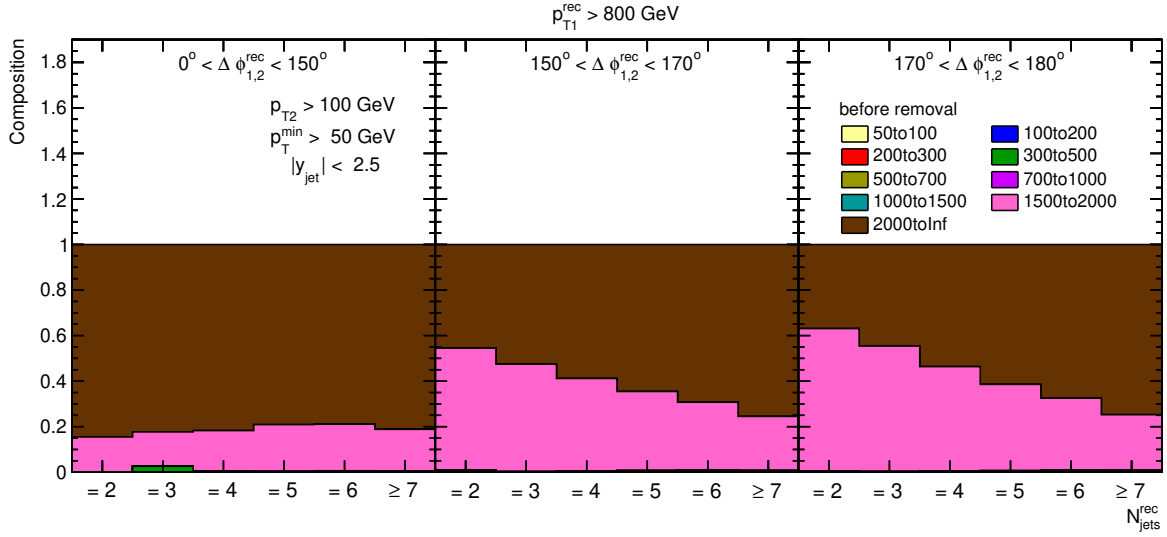


(a) Before cleaning

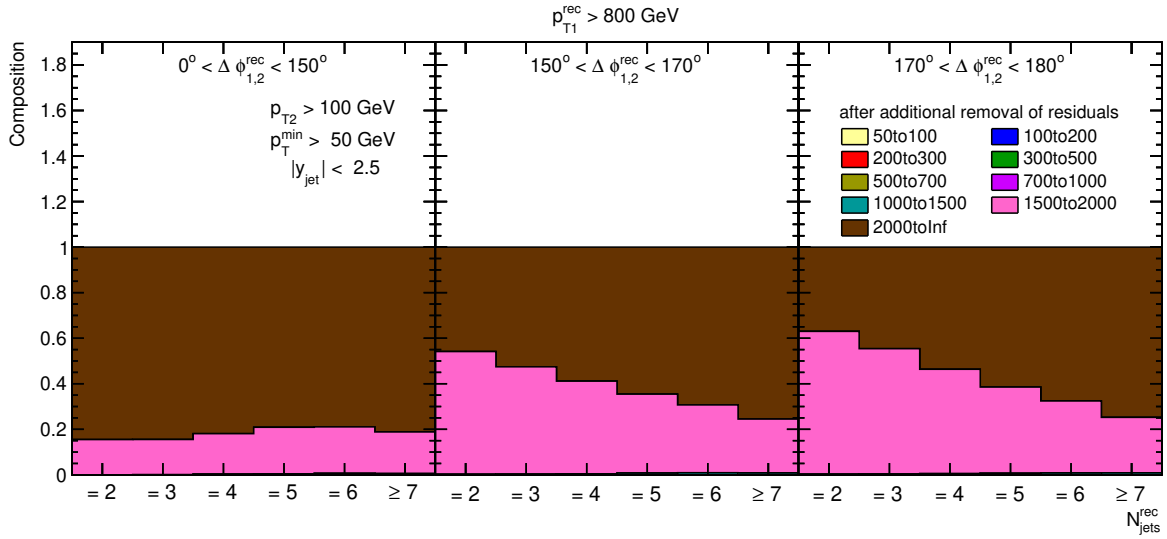


(b) After cleaning

Figure F.4: The second p_{T1} bin is shown.Contributions from the different $\hat{p}_T$ slices to the detector-level distributions of PYTHIA8 sample before and after correction for highly weighted PU events for the jet multiplicity distribution.

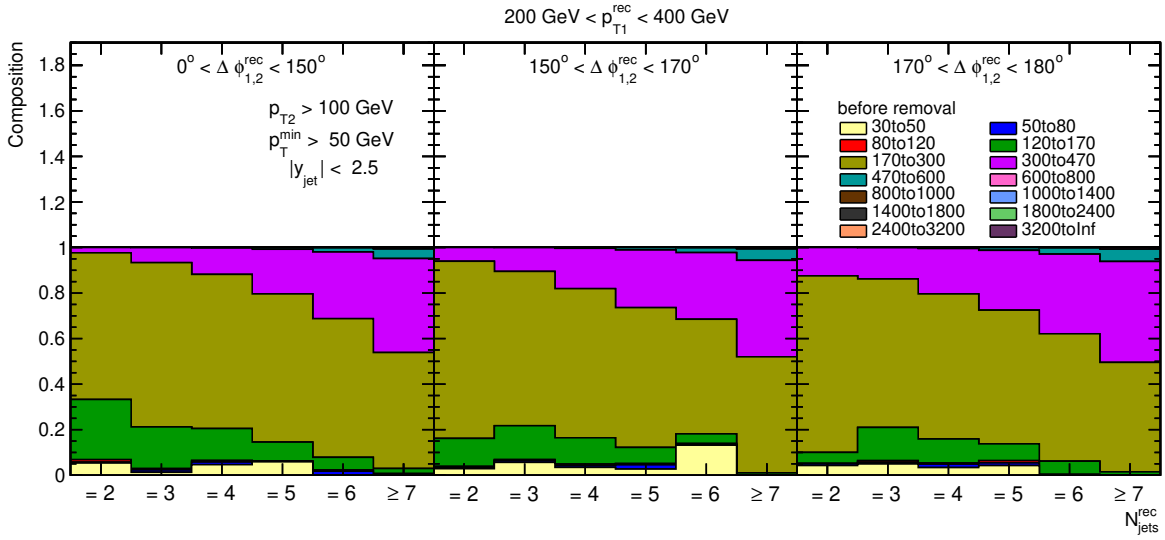


(a) Before cleaning

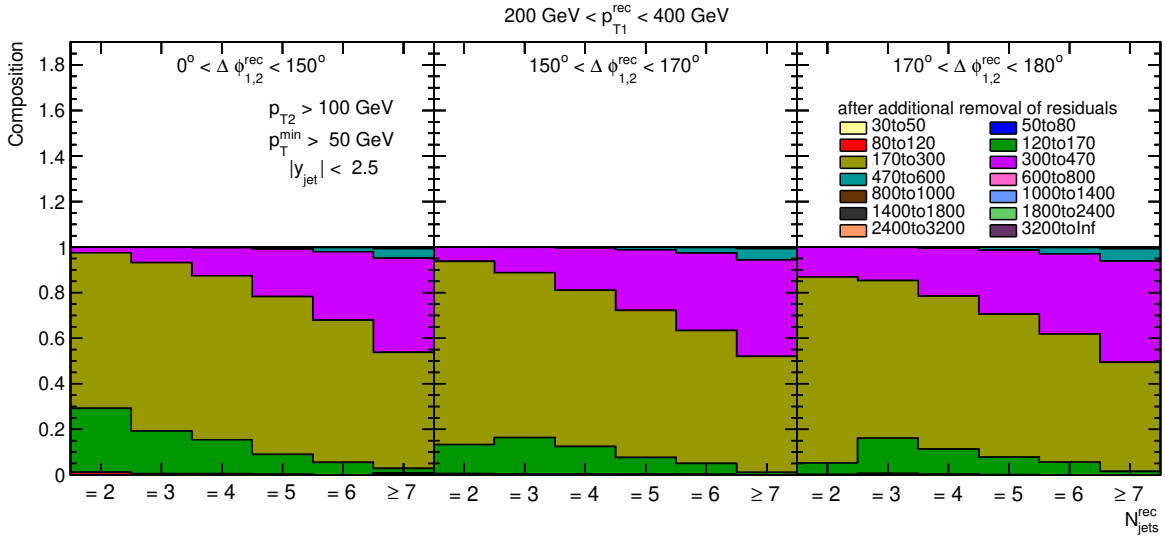


(b) After cleaning

Figure F.5: The third p_{T1} bin is shown.Contributions from the different $\hat{p}_T$ slices to the detector-level distributions of PYTHIA8 sample before and after correction for highly weighted PU events for the jet multiplicity distribution.

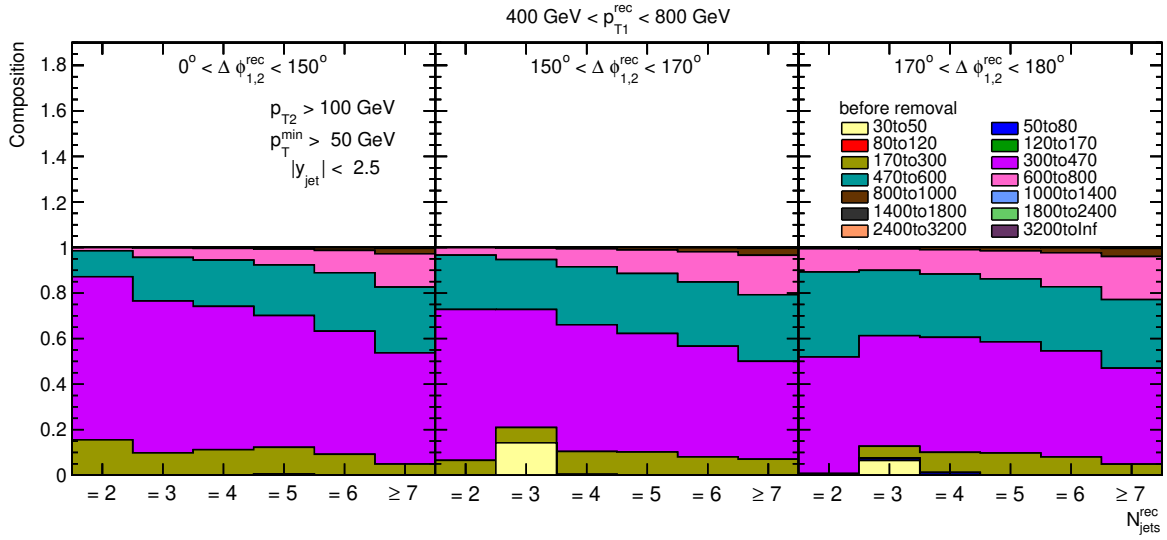


(a) Before cleaning

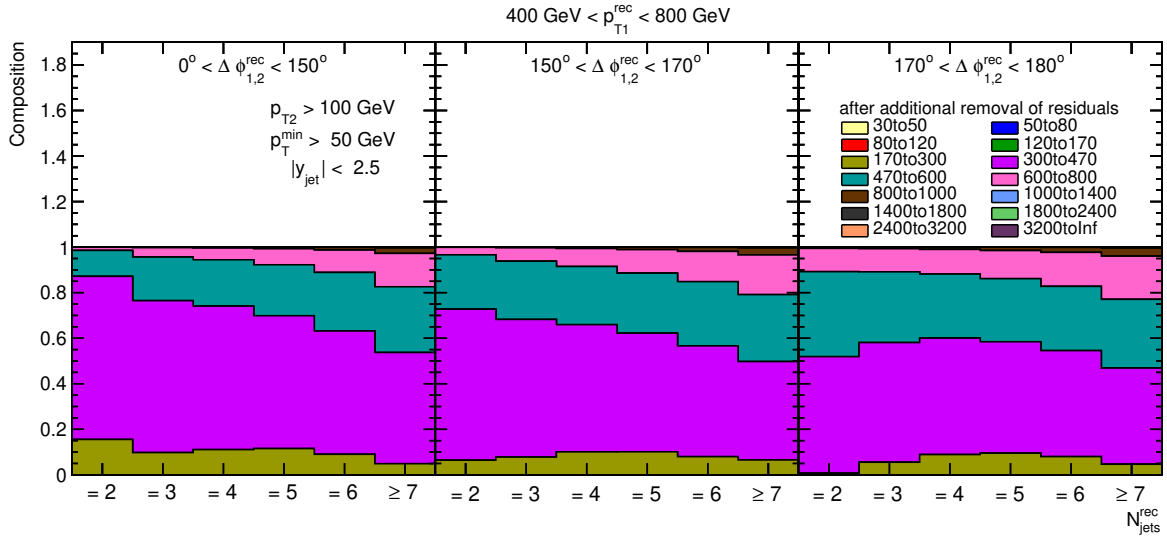


(b) After cleaning

Figure F.6: The first p_{T1} bin is shown.Contributions from the different $\hat{p}_T$ slices to the detector-level distributions of PYTHIA8 sample before and after correction for highly weighted PU events for the jet multiplicity distribution.

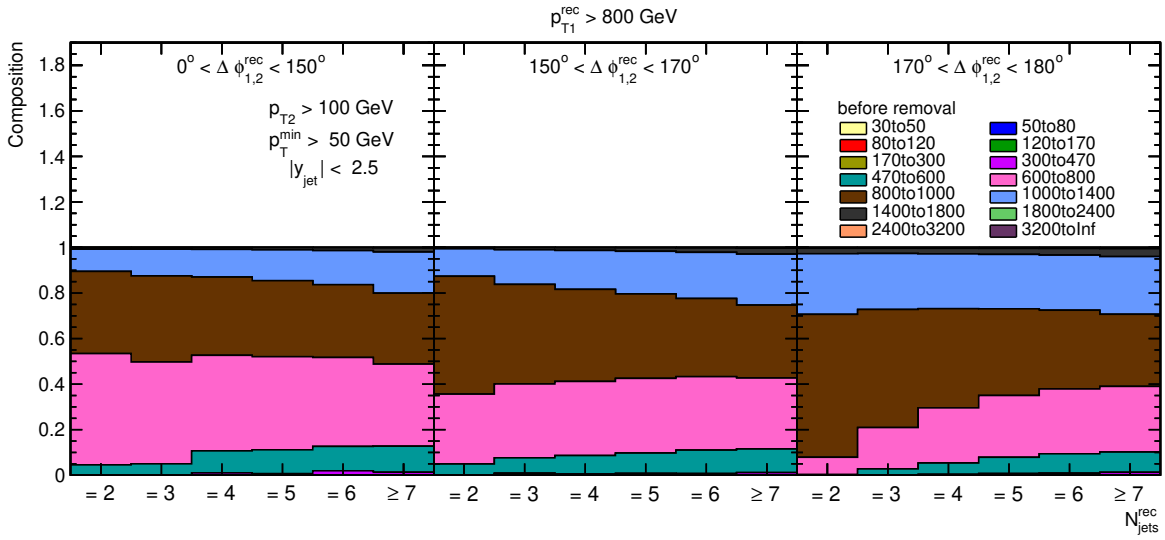


(a) Before cleaning

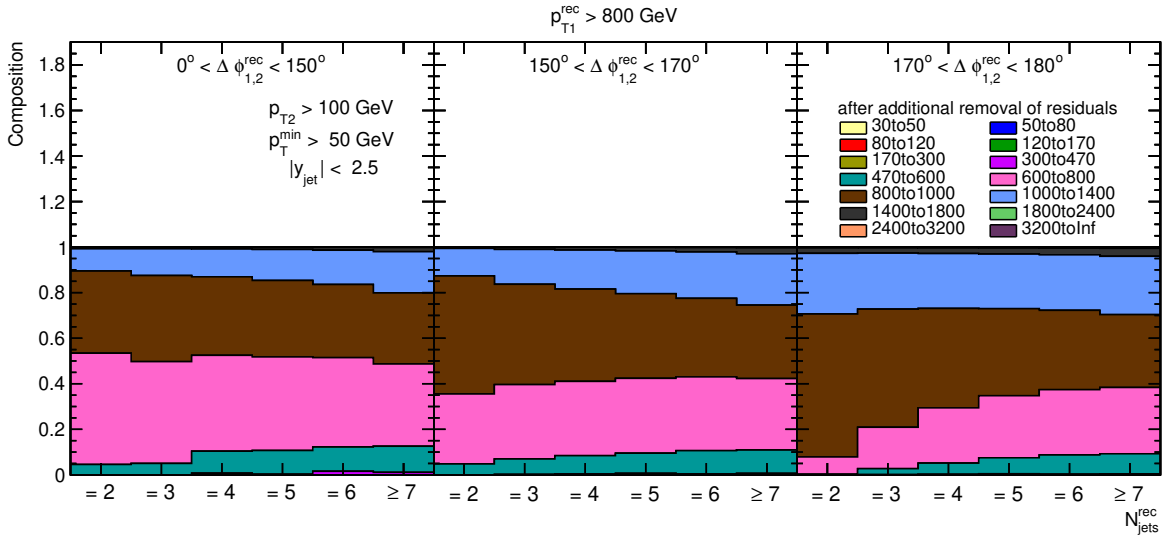


(b) After cleaning

Figure F.7: The second p_{T1} bin is shown.Contributions from the different $\hat{p}_T$ slices to the detector-level distributions of PYTHIA8 sample before and after correction for highly weighted PU events for the jet multiplicity distribution.

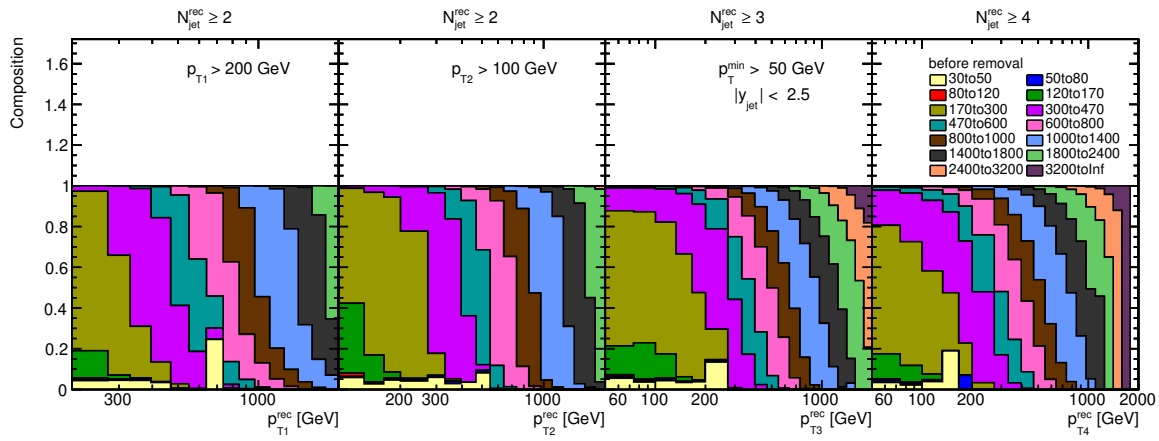


(a) Before cleaning

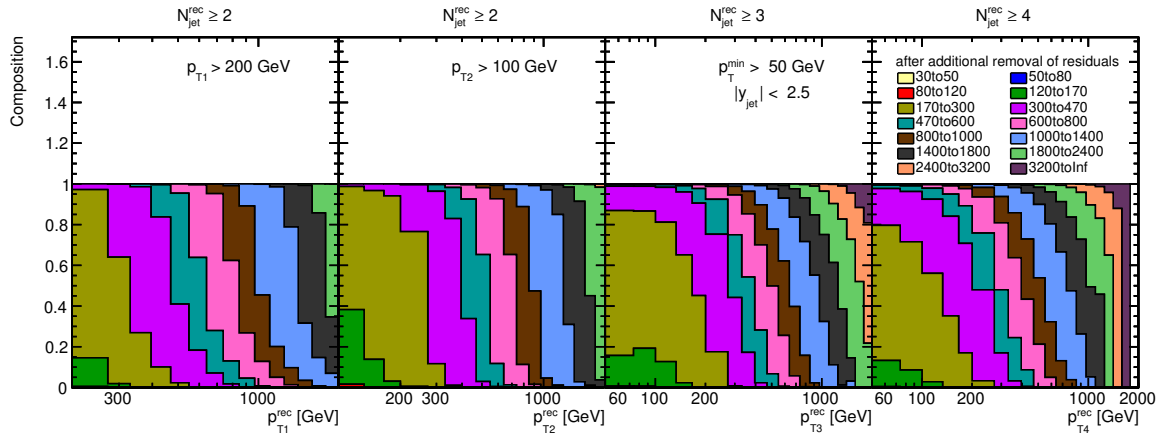


(b) After cleaning

Figure F.8: The third p_{T1} bin is shown. Contributions from the different $\hat{p}_T$ slices to the detector-level distributions of PYTHIA8 sample before and after correction for highly weighted PU events for the jet multiplicity distribution.



(a) Before cleaning



(b) After cleaning

Figure F.9: Contributions from the different $\hat{p}_T$ slices to the detector-level distributions of PYTHIA8 sample before and after correction for highly weighted PU events for the four leading jets p_{T} distributions.

APPENDIX

G

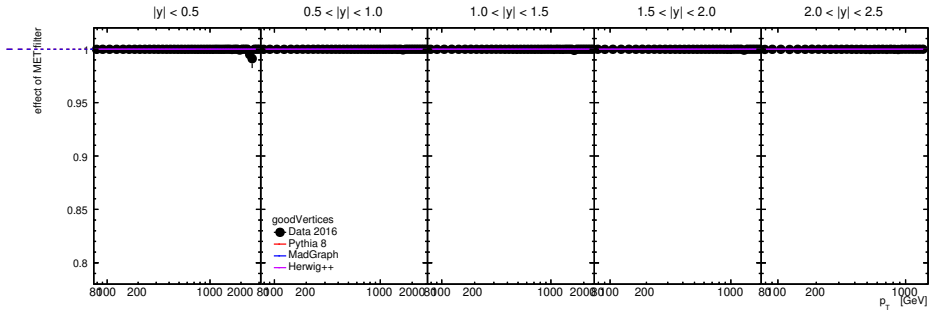
MET FILTERS STUDIES

In Sect. 5.5.1, the net effect of the application of all the MET filters was shown together. In this appendix, the detail effect of all the MET filters one by one will be presented in the following.

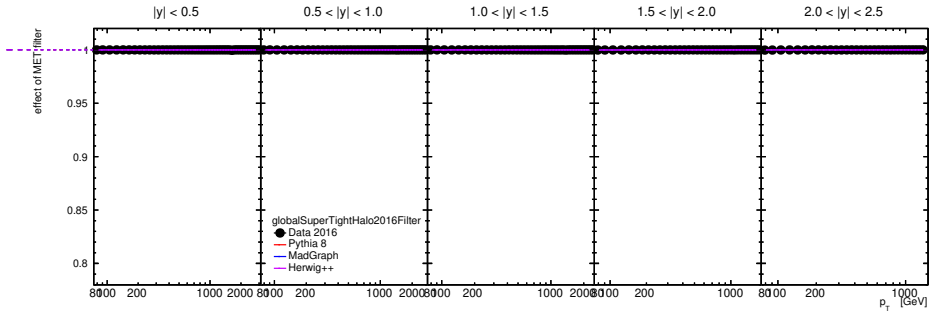
There is a total of eight MET filters as shown in Fig. G.1-G.2. Each filter is tested separately on top of the same data and simulation (*i.e.* they are not applied in series). It is found that only two of the recommended filters have a significant effect but only the first one makes a significant impact in the net effect:

- `HBHENoiseIsoFilter` (Fig. G.1c)
- `EcalDeadCellTriggerPrimitiveFilter` (Fig. G.2a)

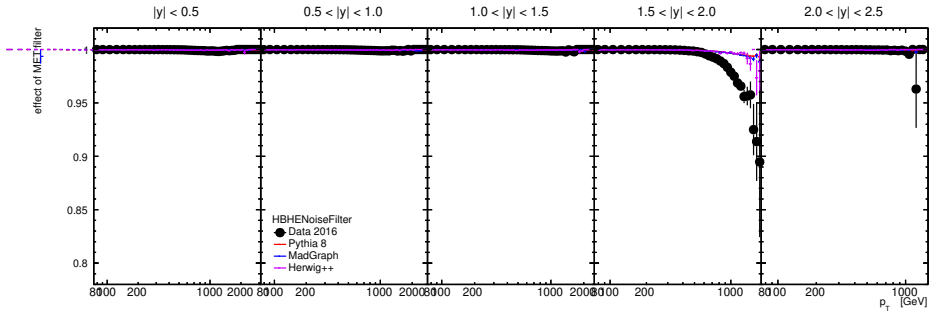
The effect of the no longer recommended filter `BadChargedCandidateFilter` is also shown in Fig. G.2c): while it was originally implemented in the context of the MIP mitigation, where jets matched with activity in the muon chambers were vetoed. However, very-high-p_T jets do leave signal in the muon chambers and should rather not be vetoed.



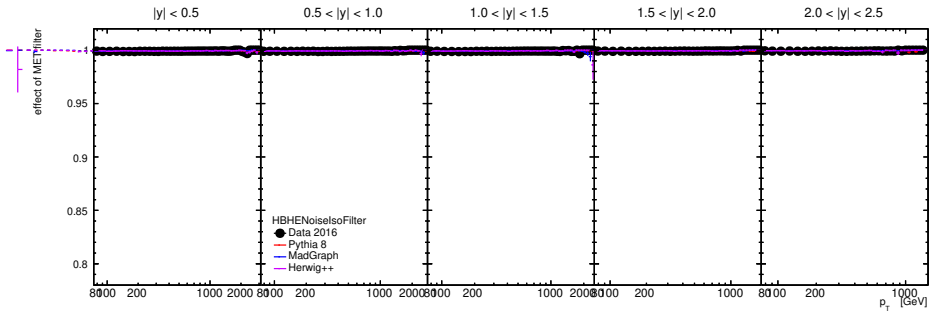
(a) `goodVertices`



(b) `globalSuperTightHalo2016Filter`

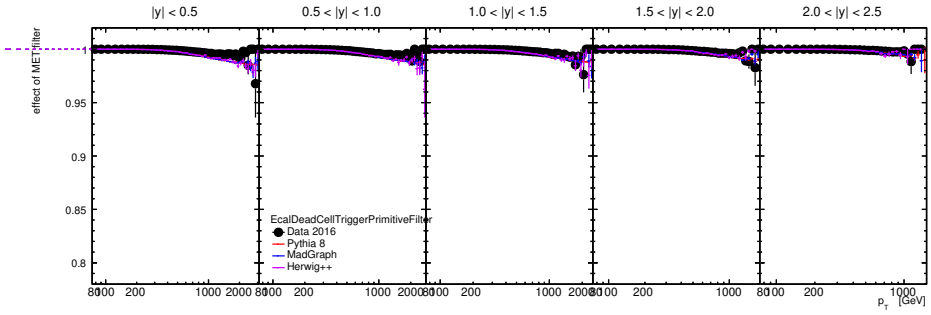


(c) `HBHENoiseFilter`

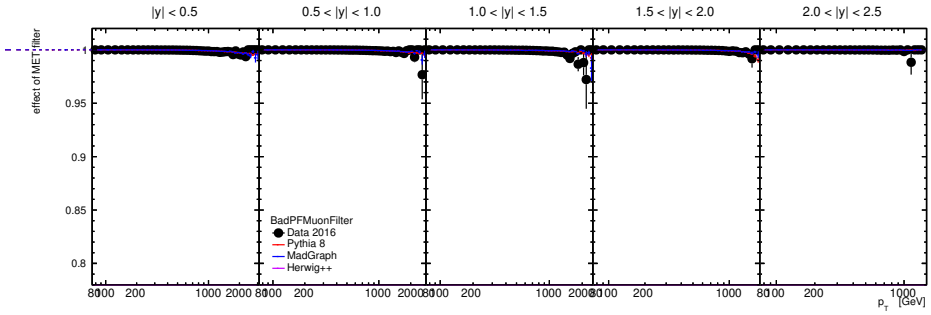


(d) `HBHENoiseIsoFilter`

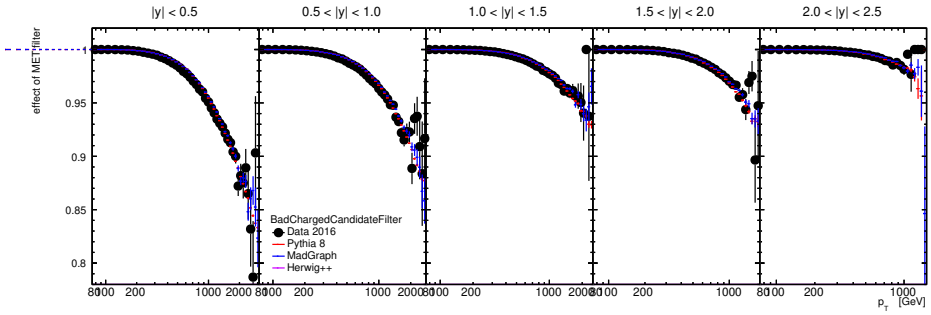
Figure G.1: Effect of MET filters applied separately in data and in simulation (part 1).



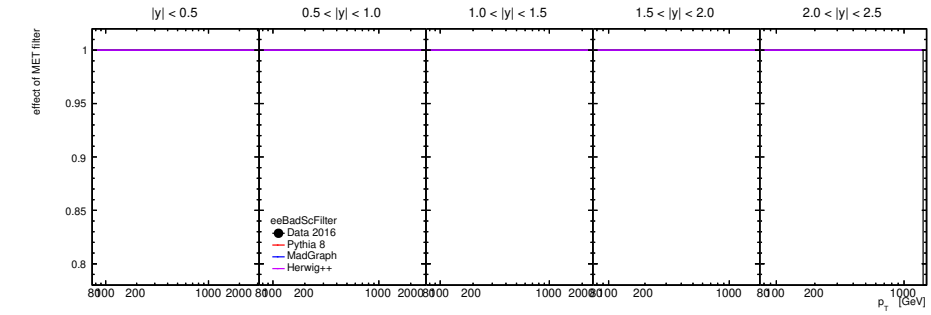
(a) `EcalDeadCellTriggerPrimitiveFilter`



(b) `BadPFMuonFilter`



(c) `BadChargedCandidateFilter`



(d) `eeBadScFilter`

Figure G.2: Effect of MET filters applied separately in data and in simulation (part 2).

APPENDIX

H

PROBABILITY MATRICES

The PM is no more than the normalization of the RM matrix to generator level axis. It give us the probability of having a generator level event to be reconstructed by the detector and measured by it. In this appendix the PM for all the used MC samples are shown. The HERWIG++, PYTHIA8 (PYTHIA8 flat sample) were used in addition to the MADGRAPH PM to study the effect of using different MC models in the unfolding. In the following all the PM for the jet multiplicity and the four leading jet p_{T} measurement will be presented.

H.1 Jet multiplicity measurement $(N_j, \Delta\phi_{1,2}, p_{T1})$

In Fig. H.1 the PM for MADGRAPH, PYTHIA8 (also flat sample is shown), and HERWIG++ samples are shown. Pythia16 refers to PYTHIA8 (full pt sliced sample), Pythia16Flat stands for the flat PYTHIA8 sample and MadGraph16 is MADGRAPH, and Herwig16 being HERWIG++, in all cases 16 stands for the dataset year (2016). Here the matrices **x** (genBinningID) and **y** (recBinningID) axis corresponds to TUnfold Binning which translate a multidimensional histogram into a vector of histograms and it gives a bin ID number from 1 to N, being N the sum of the size of each individual 3D histogram projection (in this particular case the projections are Nj (x-axis) , $\Delta\phi_{1,2}$ (y-axis), and p_{T1} (z-axis)). As we can see in the figure the condition number is always much smaller than 10, being safe to use real matrix inversion for unfolding the data.

H.2 Four leading jets p_T (p_{Ti}, n_i^{jet})

In Fig. H.2 the PM for MADGRAPH, PYTHIA8 (also flat sample), and HERWIG++ samples are shown for the four jet p_{T} spectra measurement. The condition number is always smaller than six.

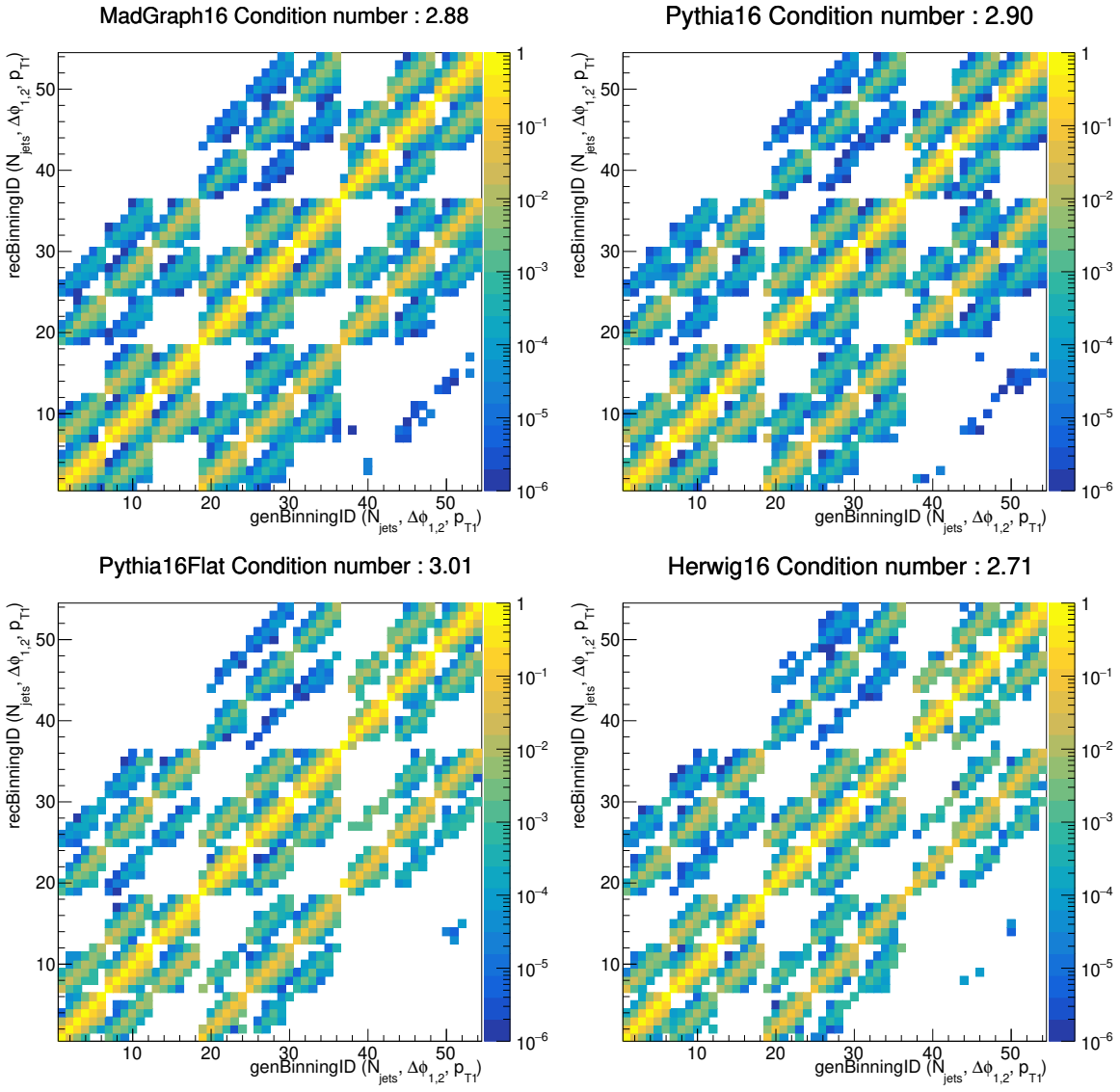


Figure H.1: Probability Matrices for 3D Jet Multiplicity distribution (for tight matching only) with its respective condition number.

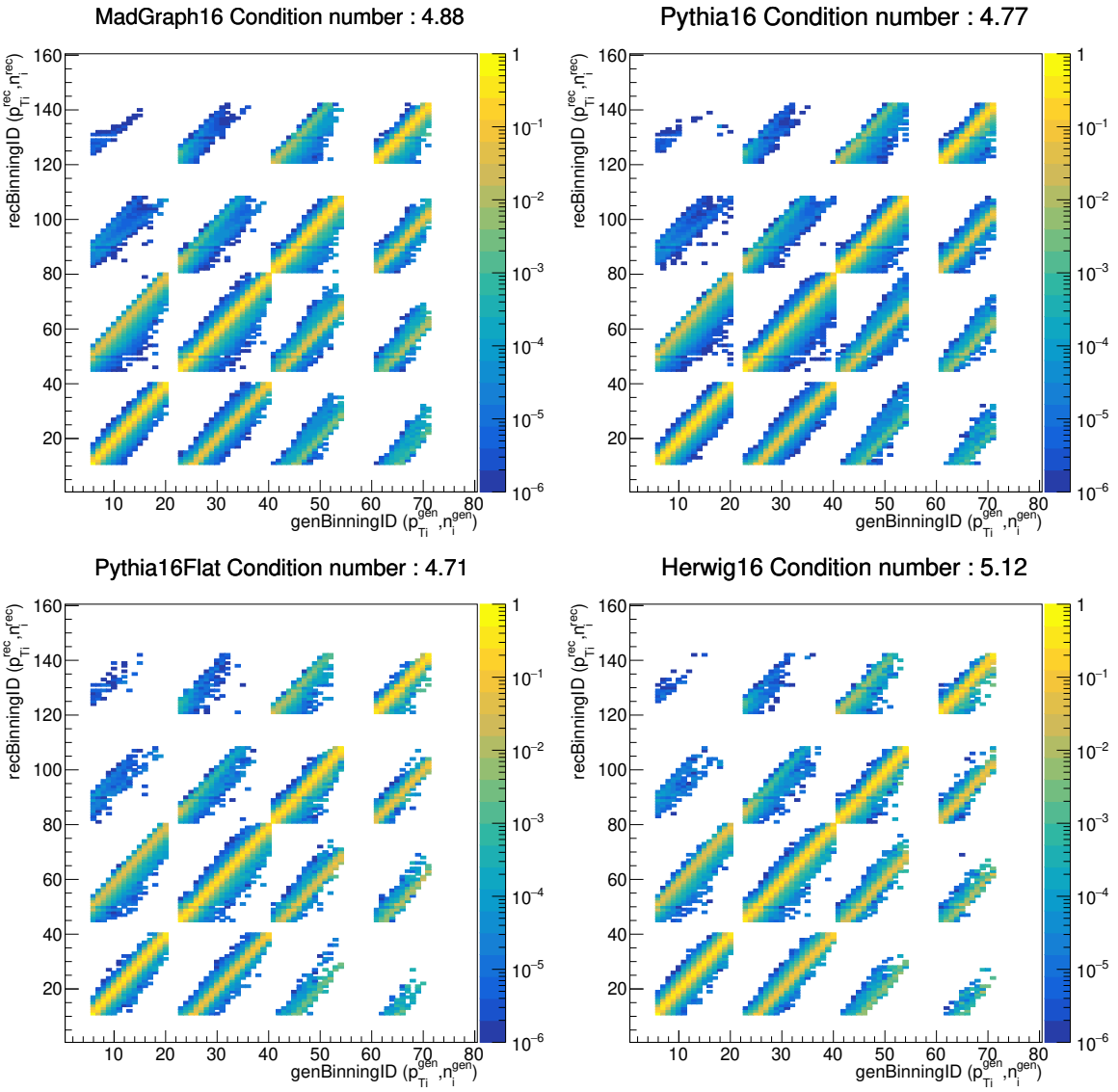


Figure H.2: Probability Matrices for 2D p_T of the first four leading jets, with its respective condition number. Here Pythia16 refers to PYTHIA8 (full pt sliced sample), Pythia16Flat stand for the flat PYTHIA8 sample and MadGraph16 is MADGRAPH, and Herwig16 is HERWIG++, in all cases 16 stands for the dataset year (2016).Using `TUnfoldBinning` the x and y-axis are defined $(p_{Ti}^{gen(rec)}, n_i^{rec(gen)})$ here $n_i^{rec(gen)}$ is the jet number with $i = 1, 2, 3, 4$.

APPENDIX

I

PREDICTIONS COMPARED TO DATA

Here the measured data is compared with the same LO and NLO predictions as in the results and discussion chapter (Ch.8) but without the normalization to the measured inclusive dijet cross section.

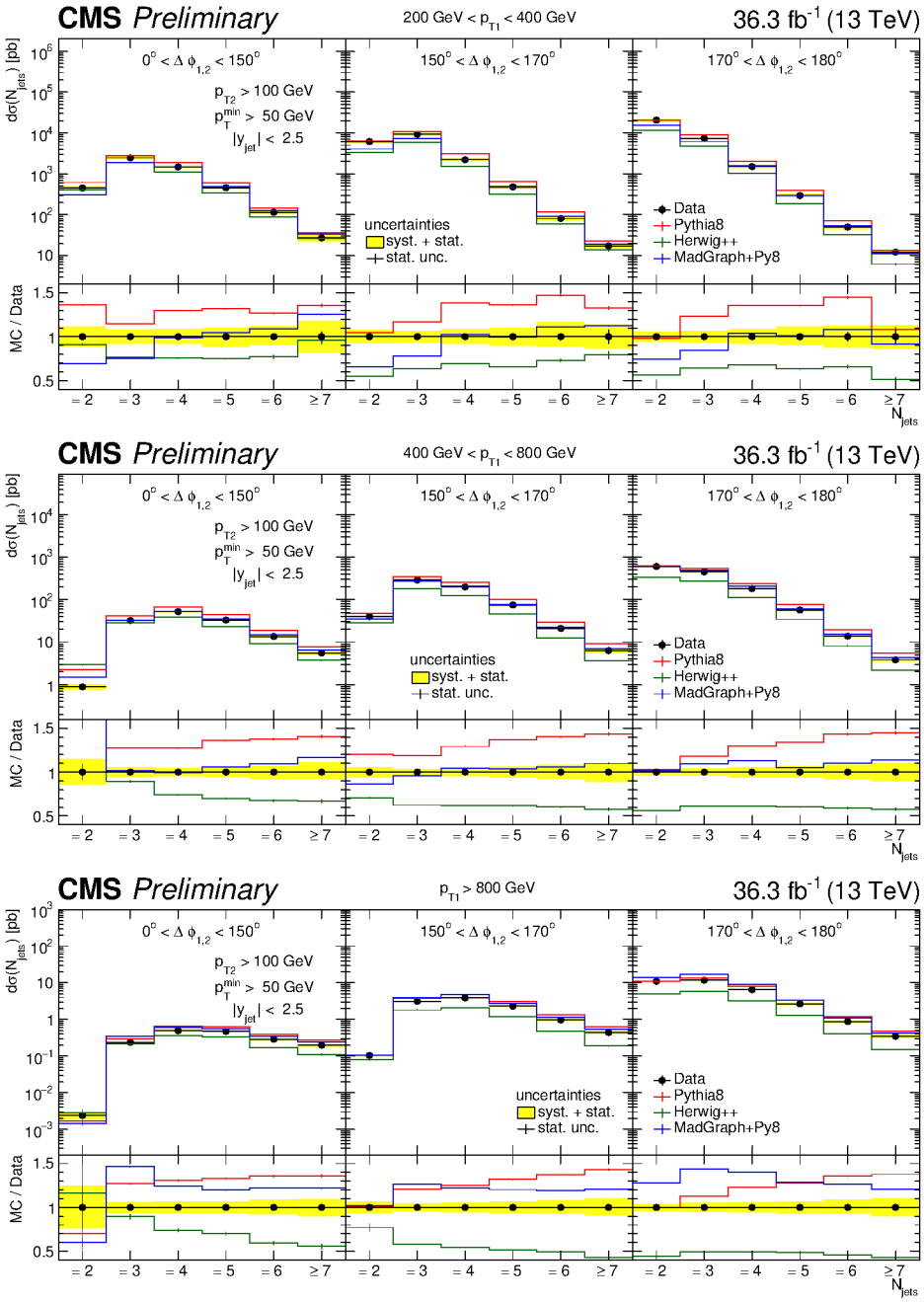


Figure I.1: Differential cross section as function of the exclusive jet multiplicity (inclusive for 7 jets) in bins of p_{T1} and $\Delta\phi_{1,2}$. The data is compared LO predictions. The yellow band in the data represents the total experimental uncertainty (with the statistical uncertainty as black vertical lines) and the predictions show the statistical uncertainty as vertical lines.

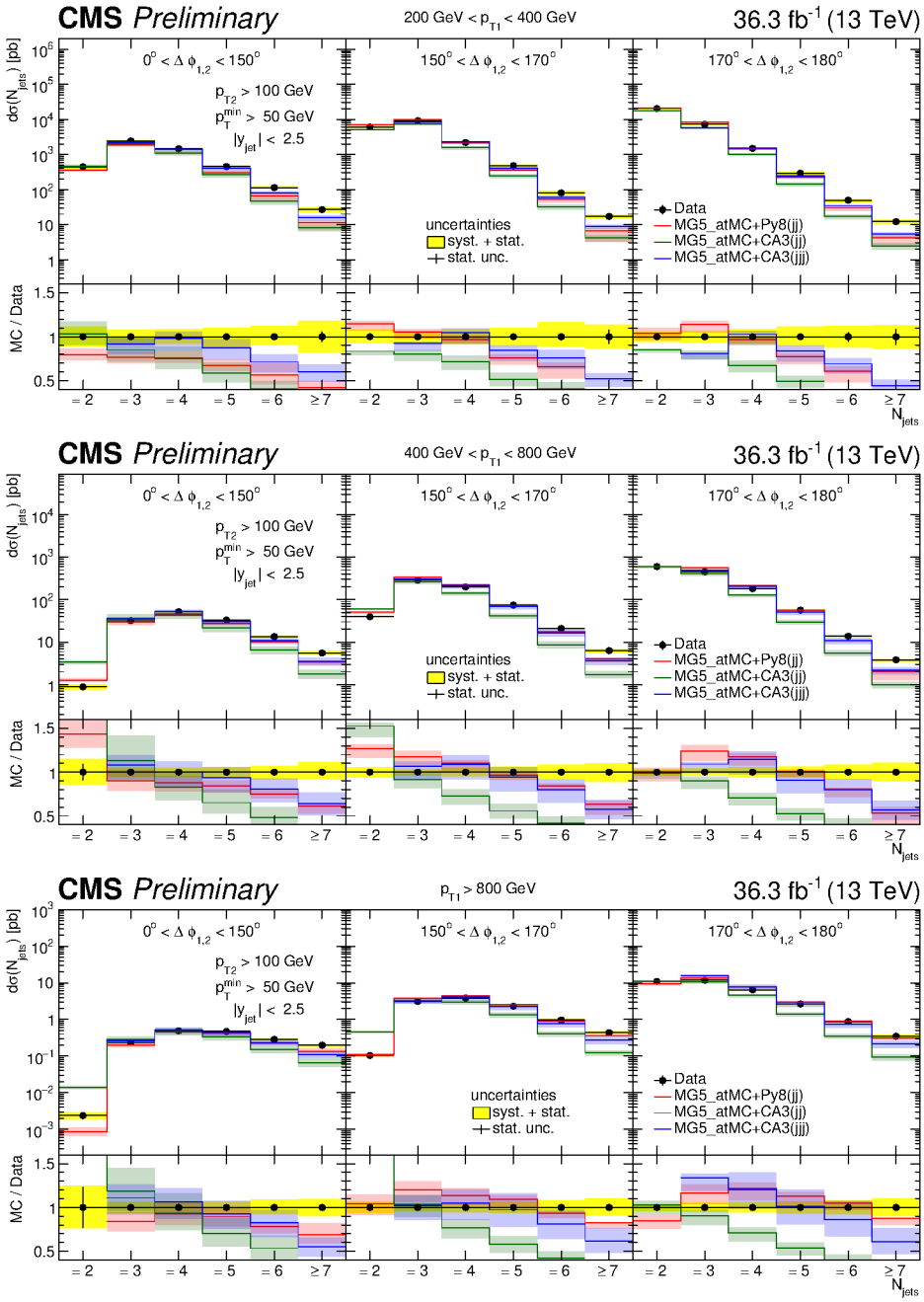
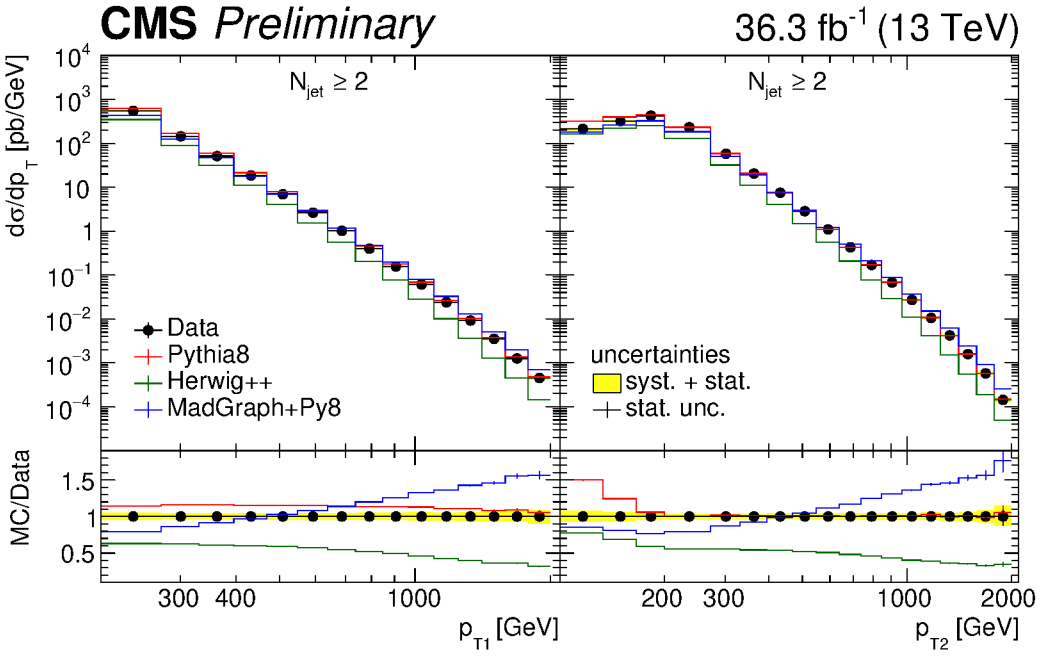
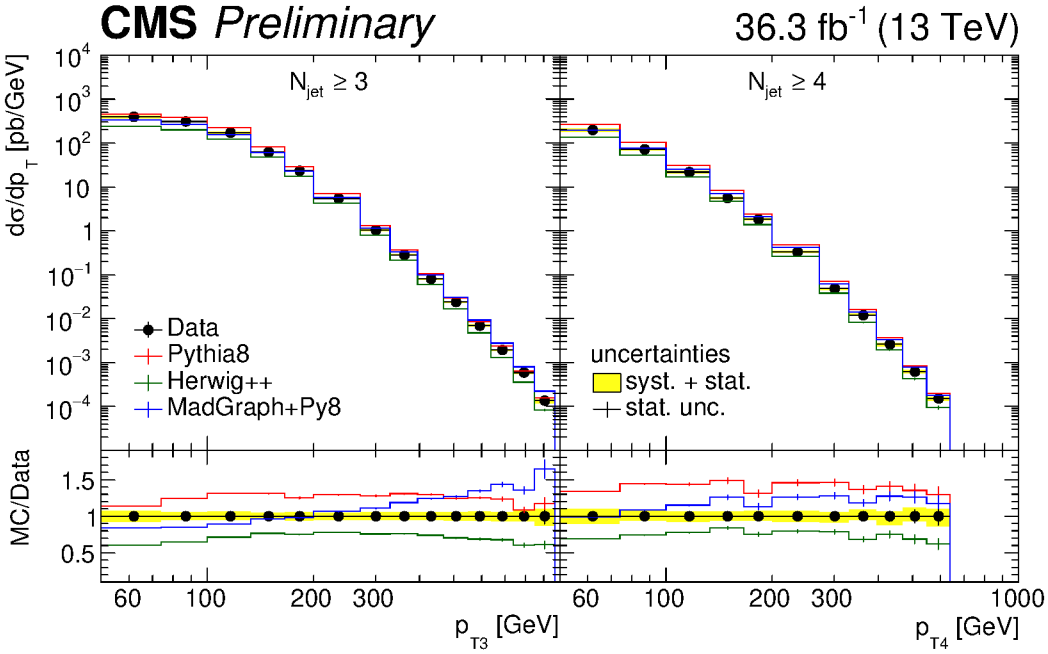


Figure I.2: Differential cross section as function of the exclusive jet multiplicity (inclusive for 7 jets) in bins of p_{T1} and $\Delta\phi_{1,2}$. The data is compared to NLO predictions from MG5_aMC+CA3 (jj) NLO, MG5_aMC+CA3 (jjj) NLO and MG5_aMC+Py8 (jj) NLO. The yellow band in the data represents the total experimental uncertainty (with the statistical uncertainty as black vertical lines) and the color bands in the theoretical predictions represent the scale uncertainty.

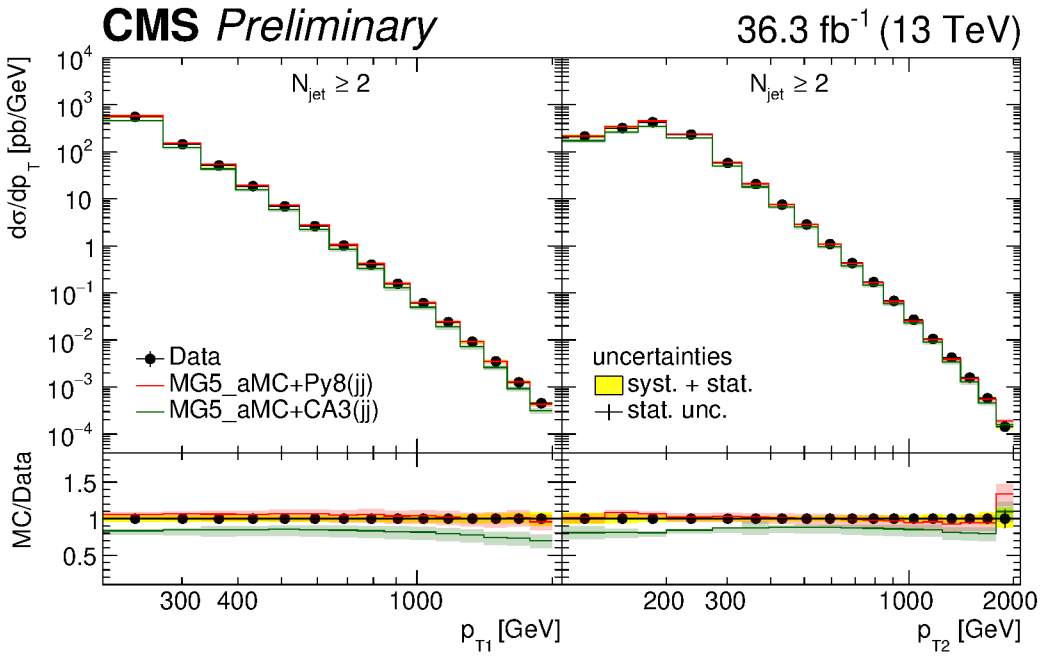


(a) Leading and 2nd leading jet p_{T} distributions.

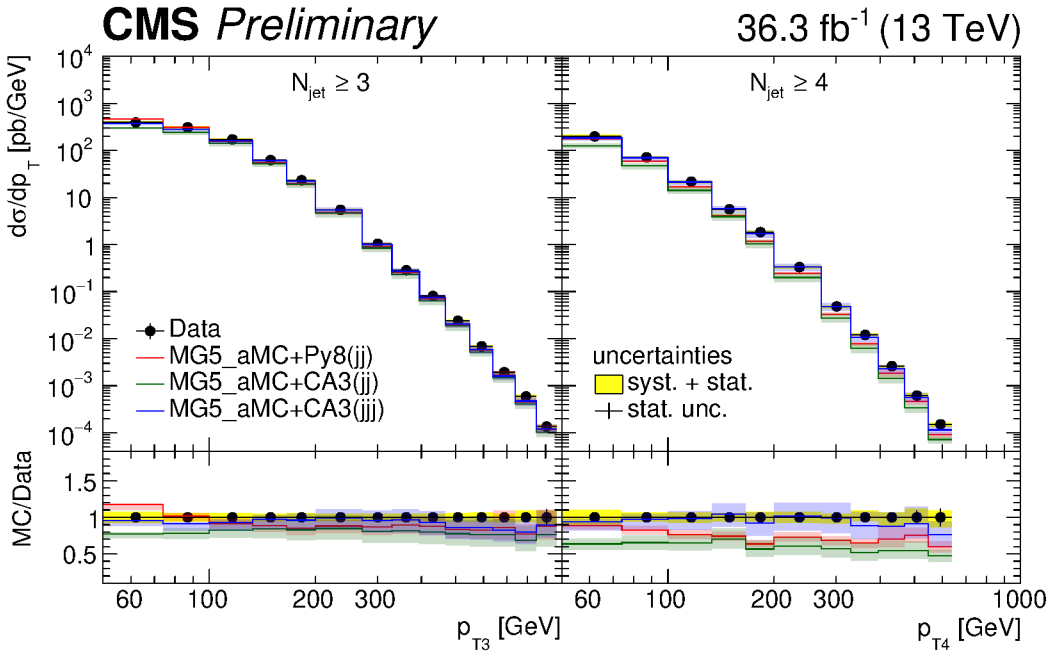


(b) 3rd and 4th leading jet p_{T} distributions.

Figure I.3: Transverse momenta of the first four leading jets compared to LO predictions.



(a) Leading and second leading jet p_{T} distributions.



(b) Third and fourth leading jet p_{T} distributions.

Figure I.4: Four leading jets p_{T} for data compared to NLO predictions from MG5_aMC+CA3 (jj) NLO, MG5_aMC+CA3 (jjj) NLO and MG5_aMC+Py8 (jj) NLO. The yellow band in the data represents the total experimental uncertainty and the color bands in the theoretical predictions represent the scale uncertainty.

APPENDIX

J

PUBLICATIONS

In this appendix, the publications which I have contributed and are directly related to the work in this thesis are presented. First is the CMS Physics Analysis Summary: "Cross section measurements of jet multiplicity and jet transverse momenta in multijet events at $\sqrt{s} = 13\,\text{TeV}$". This work is available at CERN Document Server in Ref. [116], and there the measurements from this thesis are presented. At the moment of writing this thesis, this CMS PAS is being reviewed to be published as a CMS paper.

Then two more publications (outside CMS) are presented: "Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method" as in [47] and "CASCADE3 A Monte Carlo event generator based on TMDs " as in [42]. These last two papers are directly related to the theoretical predictions presented in this thesis with CASCADE3 MC event generator.

CMS Physics Analysis Summary

Contact: cms-pag-conveners-smp@cern.ch 2021/07/25

Cross section measurements of jet multiplicity and jet transverse momenta in multijet events at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

Multijet events at large transverse momentum (p_T) are measured at $\sqrt{s} = 13$ TeV with data recorded with the CMS detector at the LHC corresponding to an integrated luminosity of 36.3 fb^{-1}. The multiplicity of jets with $p_T > 50$ GeV in addition to a high p_T dijet system is measured for different regions of the transverse momentum of the leading p_T jet and as a function of the azimuthal angle $\Delta\phi_{1,2}$ between the two leading jets in the dijet system. The differential cross section of the four jets leading in p_T is measured as a function of their transverse momentum. The measurements are compared to leading order matrix-element calculations supplemented with parton shower, hadronization and multiparton interactions. In addition next-to-leading order matrix-element calculations combined with conventional parton shower as well as with Parton Branching (PB) transverse momentum dependent (TMD) parton densities and PB-TMD initial state parton shower are compared to the measurements.

1 Introduction

The production of jets, which are reconstructed from a collimated stream of hadrons, coming from the fragmentation of high energetic partons are described by the theory of strong interactions, Quantum Chromodynamics (QCD). At leading order (LO) in the strong coupling α_S, two collinear partons from the protons scatter and produce two high p_T partons in the final state. The jets which originate from such a process are strongly correlated in the transverse plane, and the azimuthal angle between both, $\Delta\phi_{1,2}$, will be close to π. However, higher order corrections to the lowest order process will result in a decorrelation in the azimuthal plane and $\Delta\phi_{1,2}$ will deviate from π significantly. These higher order corrections can be either hard parton radiation, calculated at the matrix element level at next-to-leading order (NLO) or as softer multiple parton radiation treated by parton showers. In a recent approach, transverse momentum dependent (TMD) parton densities [1] obtained with the Parton Branching method (PB) [2, 3] are combined with next-to-leading order (NLO) matrix element calculations [4] supplemented with PB initial state parton showers [5], leading to predictions where the initial state parton shower is determined by the PB-TMD. It is of great interest to measure the multiplicity of such additional jets in different regions of $\Delta\phi_{1,2}$ as well as to measure the transverse momenta of the additional jets.

The azimuthal correlation in high p_T dijet events has been measured previously by the D0 Collaboration in $p\overline{p}$ collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV [6, 7], in pp collisions by the ATLAS Collaboration at $\sqrt{s} = 7$ TeV [8], and by the CMS Collaboration at $\sqrt{s} = 7, 8$, and 13 TeV [9–12].

In this note we describe new measurements in dijet events with transverse momenta of the leading jet $p_{T1} > 200$ GeV, and subleading jet $p_{T2} > 100$ GeV, in the range of pseudorapidity of $|y| < 2.5$. The multiplicity of jets with $p_T > 50$ GeV in addition to the two leading jets is measured in bins of p_{T1} and $\Delta\phi_{1,2}$. The jet multiplicity in bins of $\Delta\phi_{1,2}$ provides information on the number of jets needed to create the $\Delta\phi_{1,2}$ decorrelation. The cross section of the four leading jets is measured as a function of p_T of each jet, which will give additional information on the structure of the higher order corrections.

This note is organized as follows: in chapter 2 a brief summary of the CMS detector and the relevant components is given, in chapter 3 the theoretical models for comparison at detector level as well as with the final results are given. Chapter 4 gives an overview over the analysis with the event selection in section 4.1, data correction in section 4.2 and discussion of uncertainties in section 4.3. The final results and comparison with theoretical predictions is given in chapter 5. The final chapter gives a summary and conclusions.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections.

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4 μs [13]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event

rate to around 1 kHz before data storage [14].

During the 2016 data-taking period, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region $|\eta| > 2.0$, also known as prefiring, caused some trigger inefficiencies [13]. For events containing a jet with p_T larger than 100 GeV, in the region $2.5 < |\eta| < 3.0$ the efficiency loss is 10–20%, depending on p_T, η, and time. Correction factors were computed from data and applied to the acceptance evaluated by simulation.

The particle-flow algorithm (PF) [15] aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

Jets are reconstructed from PF objects, clustered using the anti-k_T algorithm [16, 17] with a distance parameter of $R = 0.4$. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_T spectrum and detector acceptance. Additional proton-proton interactions within the same or nearby bunch crossings can contribute additional tracks and calorimetric energy depositions, increasing the apparent jet momentum. To mitigate this effect, tracks identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions. Jet energy corrections are derived from simulation studies so that the average measured energy of jets becomes identical to that of particle level jets. In situ measurements of the momentum balance in dijet, photon+jet, Z+jet, and multijet events are used to determine any residual differences between the jet energy scale in data and in simulation, and appropriate corrections are made [18]. Additional selection criteria are applied to each jet to remove jets potentially dominated by instrumental effects or reconstruction failures. The jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [18].

The missing transverse momentum vector $\vec{p}_T^{\,\text{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as p_T^{miss} [19]. The $\vec{p}_T^{\,\text{miss}}$ is modified to account for corrections to the energy scale of the reconstructed jets in the event. Anomalous high-p_T^{miss} events can be due to a variety of reconstruction failures, detector malfunctions or non collisions backgrounds. Such events are rejected by event filters that are designed to identify more than 85–90% of the spurious high-p_T^{miss} events with a mistagging rate less than 0.1% [19].

The candidate vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex. The physics objects are the jets, clustered using the jet finding algorithm [16, 17] with the tracks assigned to candidate vertices as inputs, and the associated missing transverse momentum.

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [20].

3 Theoretical predictions

Predictions from Monte Carlo event generators at leading-order (LO) are used for comparisons with the measured distributions.

- PYTHIA 8 [21] simulates LO $2 \to 2$ hard processes. The parton shower is generated in a phase space ordered in transverse momentum and longitudinal momentum of the emitted partons and the colored strings are hadronized with Lund string fragmentation. The CUETP8M1 [22] tune (with the parton distribution function (PDF) set NNPDF2.3LO [23]) sets the parameters for multiparton interactions.
- HERWIG++ [24] simulates LO $2 \to 2$. The emitted partons in the parton shower obey angular ordering conditions, and the cluster fragmentation is used to transform colored partons into observable hadrons. The CUETHppS1 [22] tune (with the PDF set CTEQ6L1 [25]) is applied.
- MADGRAPH5_aMC@NLO [4] version 2.3.3 event generator (labeled as MADGRAPH+PY8) is used in the LO mode of up to four noncollinear high-p_{T} partons, supplemented with parton showering and multiparton interactions by PYTHIA 8 with CUETP8M1 tune and merged according to the k_{T}-MLM matching procedure [26] with a matching scale of 10 GeV.

At next-to-leading order (NLO) in the strong coupling, different theoretical predictions are used. The factorization and renormalization scales $\mu_f = \mu_r$ are set to half the sum over the scalar transverse momenta of all produced particles/partons $1/2 \sum_i H_{T\,i}$ (as is the default in MADGRAPH). The uncertainty bands of the NLO predictions are determined from the variation of the factorization and renormalization scales by a factor of two, using the 7-point variation (avoiding the largest scale differences).

- MG5_aMC [4] interfaced with PYTHIA 8 (labelled MG5_aMC+PY8 (jj) NLO), with MEs computed at NLO for the process $pp \to jj$. The NNPDF 3.0 NLO PDF [27] is used and $\alpha_s(m_{\mathrm{Z}})$ is set to 0.118. The CUETP8M1 tune is used for PYTHIA 8. This calculation is used to study the effect of multi-parton interactions.
- MG5_aMC [4] interfaced with CASCADE3 [5] (labelled MG5_aMC+CA3 (jj) NLO), with MEs computed at NLO for the process $pp \to jj$. The HERWIG6 subtraction terms in MCatNLO are used, as they are closest to the needs for applying PB TMDs (Transverse Momentum Dependent) parton densities, as described in Ref. [5]. The NLO PB TMD set 2 [1] with $\alpha_s(m_{\mathrm{Z}}) = 0.118$ is used. The inclusion of the transverse momentum k_{T} and initial state PB-TMD parton shower is performed with CASCADE3 [5], final state radiation (since not constrained by TMDs) as well as hadronization is performed with PYTHIA6 [28]. The initial state parton shower follows the PB TMD distribution, and has no free parameters left. Multiparton interactions are not simulated in this approach.
- MG5_aMC [4] interfaced with CASCADE3 (labelled MG5_aMC+CA3 (jjj) NLO), with MEs computed at NLO for the process $pp \to jjj$. The same PB TMD distribution and parton shower, as for MG5_aMC+CA3 (jj) NLO is applied.

In Tab. 1 the used Monte Carlo event generators are summarized. All calculations are normalized to the observed dijet cross section.

Events generated by PYTHIA 8, MADGRAPH+PY8 and HERWIG++ are passed through a full detector simulation based on GEANT4 [29]. The simulated events are reconstructed with standard CMS programs.

Table 1: Description of the simulated samples used in the analysis.

generator	PDF	matrix element	tune
PYTHIA 8 [21]	NNPDF 2.3 (LO) [23]	LO $2 \to 2$	CUETP8M1 [22]
MADGRAPH+PY8 [4]	NNPDF 2.3 (LO) [23]	LO $2 \to 2,3,4$	CUETP8M1 [22]
HERWIG++ [24]	CTEQ6L1 (LO) [25]	LO $2 \to 2$	CUETHppS1 [22]
MG5_AMC+PY8 (jj)	NNPDF 3.0 (NLO) [27]	NLO $2 \to 2$	CUETP8M1 [22]
MG5_AMC+CA3 (jj)	PB set 2 (NLO) [1]	NLO $2 \to 2$	–
MG5_AMC+CA3 (jjj)	PB set 2 (NLO) [1]	NLO $2 \to 3$	–

4 Data analysis

The data samples recorded in 2016 corresponding to an integrated luminosity of 36.3 fb^{-1}, collected with single-jet high-level triggers (HLT), were used in this measurement. For each single-jet HLT trigger at least one jet with p_{T} higher than the p_T^{HLT} trigger threshold is required. All triggers, except the last one with the highest p_T^{HLT} threshold, were prescaled. We consider events only if the leading jet, reconstructed with PF algorithm, can be matched with a HLT jet. In Table 2 we show the integrated luminosity $\mathcal{L}$ for each trigger. The trigger efficiency ($> 99.5\%$) is estimated using triggers with lower $p_{\text{T}}^{\text{HLT}}$ thresholds, and for the trigger with lowest p_T^{HLT} threshold a tag-and-probe method is used to determine the p_{T}^{PF} threshold. The trigger inefficiency due to prefiring (see Sec. 2) is included in the simulated event samples.

The jets are corrected using the standard Jet Energy Scale (JES) correction procedure in CMS [18], and an additional smoothing procedure is applied to the JES correction in order to avoid a non-smooth behavior in the cross sections calculated in this analysis. The simulated samples are corrected to take into account the Jet Energy Resolution (JER) by smearing the p_{T} of the jets according to the resolution extracted from data. Jets reconstructed in regions of the detector corresponding to hot zones in the calorimeters are excluded from the measurement.

In the simulated samples events in which the primary vertex (PV) may be taken from pileup are excluded. In addition the pileup profile from simulation is reweighted to reproduce the one in data.

Table 2: The integrated luminosity for each trigger sample considered in this analysis with the p_{T} thresholds for HLT(PF) reconstruction.

$p_{\text{T}}^{\text{HLT}}$ (GeV)	40	60	80	140	200	260	320	400	450
p_{T}^{PF} (GeV)	74	97	133	196	272	362	430	548	592
$\mathcal{L}(\text{pb}^{-1})$	0.267	0.726	2.76	24.2	103	594	1770	5190	36300

4.1 Event selection

Each event is required to have at least one off-line reconstructed vertex [30] with at least one PV. The position of the proton-proton collision along the beamline needs to be $|z(PV)| < 24\,\text{cm}$, where $z(PV) = 0$ indicates the center of CMS detector. The radius of the PV in the x-y plane is required to be $\rho < 2\,\text{cm}$, and the vertex fit should be constrained for more than four degrees of freedom (ndof > 4).

All events which have jets clustered using the anti-k_{T} algorithm [16, 17] with a distance parameter of $R = 0.4$ and reconstructed with rapidity $|y| < 3.2$ and transverse momentum $p_{\text{T}} > 20\,\text{GeV}$ are considered. From these events the ones with at least two jets with $p_{T1} > 200\,\text{GeV}$, $p_{T2} > 100\,\text{GeV}$ and $|y^{1,2}| < 2.5$ are selected (events with one of the leading two jets with $|y| > 2.5$ are vetoed). Additional jets must have $p_{\text{T}} > 50\,\text{GeV}$ and $|y| < 2.5$. In addition, the

jets must satisfy a quality criteria based on the jet constituents, in order to efficiently reject fake jets [31].

In order to remove contributions from non-QCD processes ($tt + jets$ and $W/Z + jets$), especially contributions from $Z \to \nu\nu$ and $W \to l\nu$, events with a ratio of missing transverse energy over sum of transverse energies, $E_T^{miss} / \sum E_T > 0.1$, are rejected.

Within the event selection presented before, we measure the exclusive jet multiplicity (N_{jets}) up to 6 jets (inclusive for $N_{jets} \geq 7$), in three bins of p_{T1} ($200 < p_{T1} < 400\,\text{GeV}$, $400 < p_{T1} < 800\,\text{GeV}$ and $p_{T1} > 800\,\text{GeV}$) and for three different $\Delta\phi_{1,2}$ regions ($0^\circ < \Delta\phi_{1,2} < 150^\circ$, $150^\circ < \Delta\phi_{1,2} < 170^\circ$, and $170^\circ < \Delta\phi_{1,2} < 180^\circ$). Also in addition the p_T spectra of the four leading jets are measured.

4.2 Correction for detector effects

The measured cross sections are corrected for smearing and inefficiencies of the detector for comparison with particle level predictions (unfolding). The response matrix (RM) maps the true distribution onto the measured one. The RM is constructed using the simulated MC samples: PYTHIA 8, MADGRAPH+PY8, and HERWIG++. The MADGRAPH+PY8 sample, with the smaller statistical uncertainty, is used as default for constructing the RM, and the HERWIG++ and PYTHIA 8 samples are used to evaluate the effect of using different models.

The RM is constructed by matching the detector level and particle (hadron) level distribution. If events (or jets) cannot be matched they contribute to the background or inefficiency distributions. For the jet multiplicity, the dijet system (leading and subleading jets) is matched if the jets coincide within $\Delta R < 0.2$ (half of the jet radius of $R = 0.4$). For the p_T distributions, jets are matched with $\Delta R < 0.2$, and from the matched candidates the one highest in p_T is selected (only events with at least two jets are considered in the matching).

In Fig. 1 the Probability Matrix (PM) is shown for the p_T of the first four jets and Fig. 2 shows the PM for the multiplicity distributions. This PM ($\mathbf{A}$) comes from the normalization of the RM to the particle level axis (row-by-row normalization), and describes the probability that a particle level jet (or event) generated in a bin is reconstructed (migrates) in (to) another bin at detector-level. Mathematically the method for unfolding the particle-level distribution $\mathbf{x}$ from the detector-level distribution $\mathbf{y}$, using the background $\mathbf{b}$, can be represented as follows:

$$\chi^2 = \min_{\mathbf{x}} \left[(\mathbf{y} - \mathbf{b} - \mathbf{A}\mathbf{x})^\intercal \, {\mathbf{V}_{\mathbf{yy}}}^{-1} \, (\mathbf{y} - \mathbf{b} - \mathbf{A}\mathbf{x}) \right], \tag{1}$$

where $\mathbf{V_{yy}}$ represents the covariance matrix of $\mathbf{y}$. It is important to mention that this method works for $\mathbf{n} \geq \mathbf{m}$ where $\mathbf{n}(\mathbf{m})$ is the dimension (or number of bins) of $\mathbf{y}(\mathbf{x})$. Twice the number of bins at detector level compared to particle level are used in the unfolding. Condition numbers of the matrices (value of the ratio between the biggest and smallest auto-value of the matrix) smaller than 10 are required for real (pseudo) inversion. The condition number is ~ 5 for the p_T distributions (Fig. 1), a condition number of 3 for the jet multiplicity distribution is obtained (Fig. 2). Therefore, real (pseudo) inversion for unfolding the jet multiplicity (p_T of the four leading jets) can be applied, respectively, and no additional regularization is needed. The TUnfold (version 17.9) package [32], which treats also background and inefficiencies, is used.

4.3 Uncertainties

The statistical uncertainties from the measured spectra and response matrices are propagated to final results by the unfolding procedure. In Fig. 3 the statistical correlations are shown for

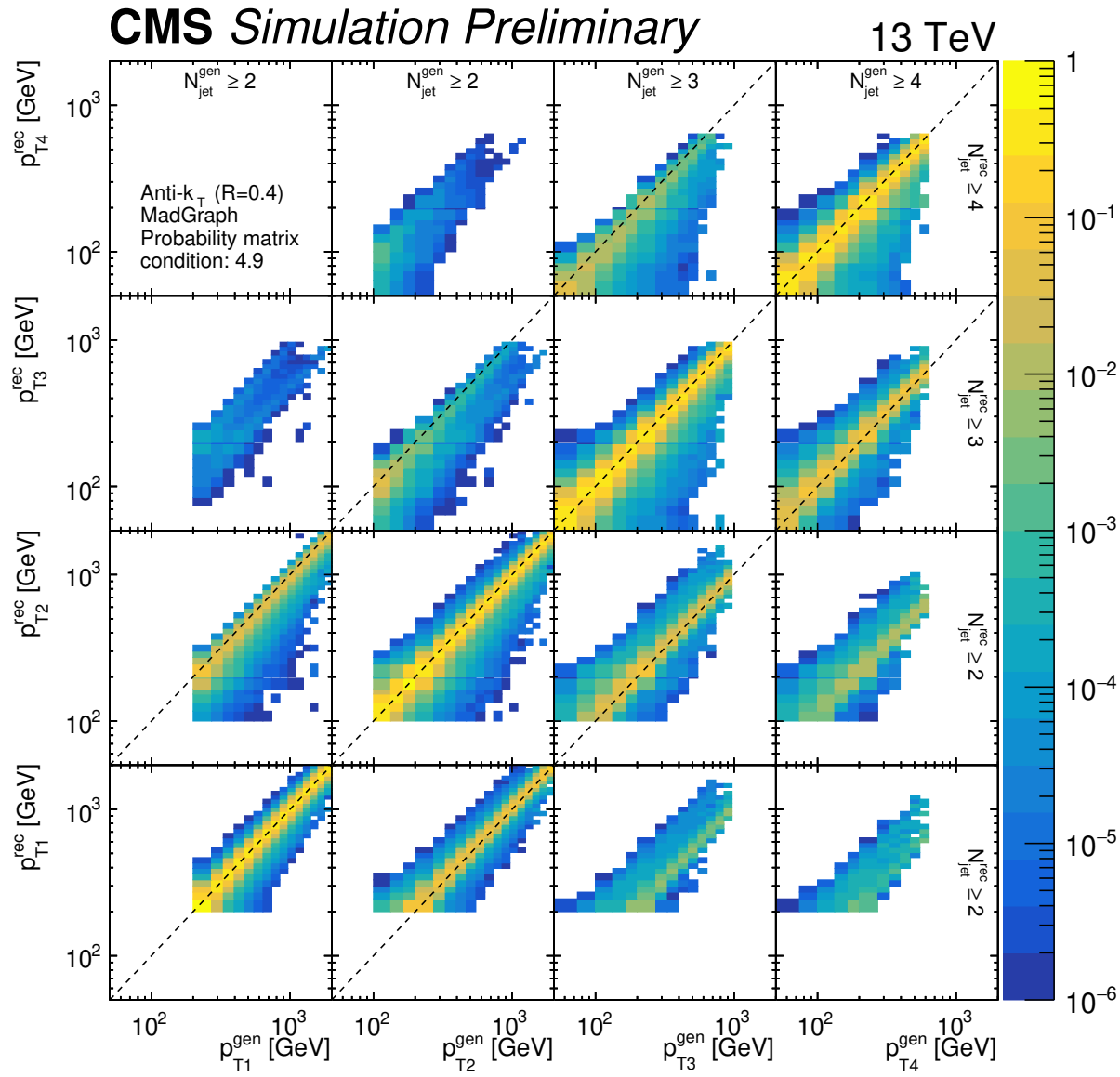


Figure 1: Probability matrix (condition number: 4.9) for the p_T of the four leading jets constructed with the MADGRAPH+PY8 sample. The global 4×4 sectors corresponds to each one of the first four jets p_T distributions, the x-axis corresponds to the hadron(gen) level and y-axis corresponds to the detector(rec) level as labeled in each axis. The z-axis covers a range from 10^{-6} to 1 indicating the probability of migrations from the hadron level bin to the correspondent detector level bin.

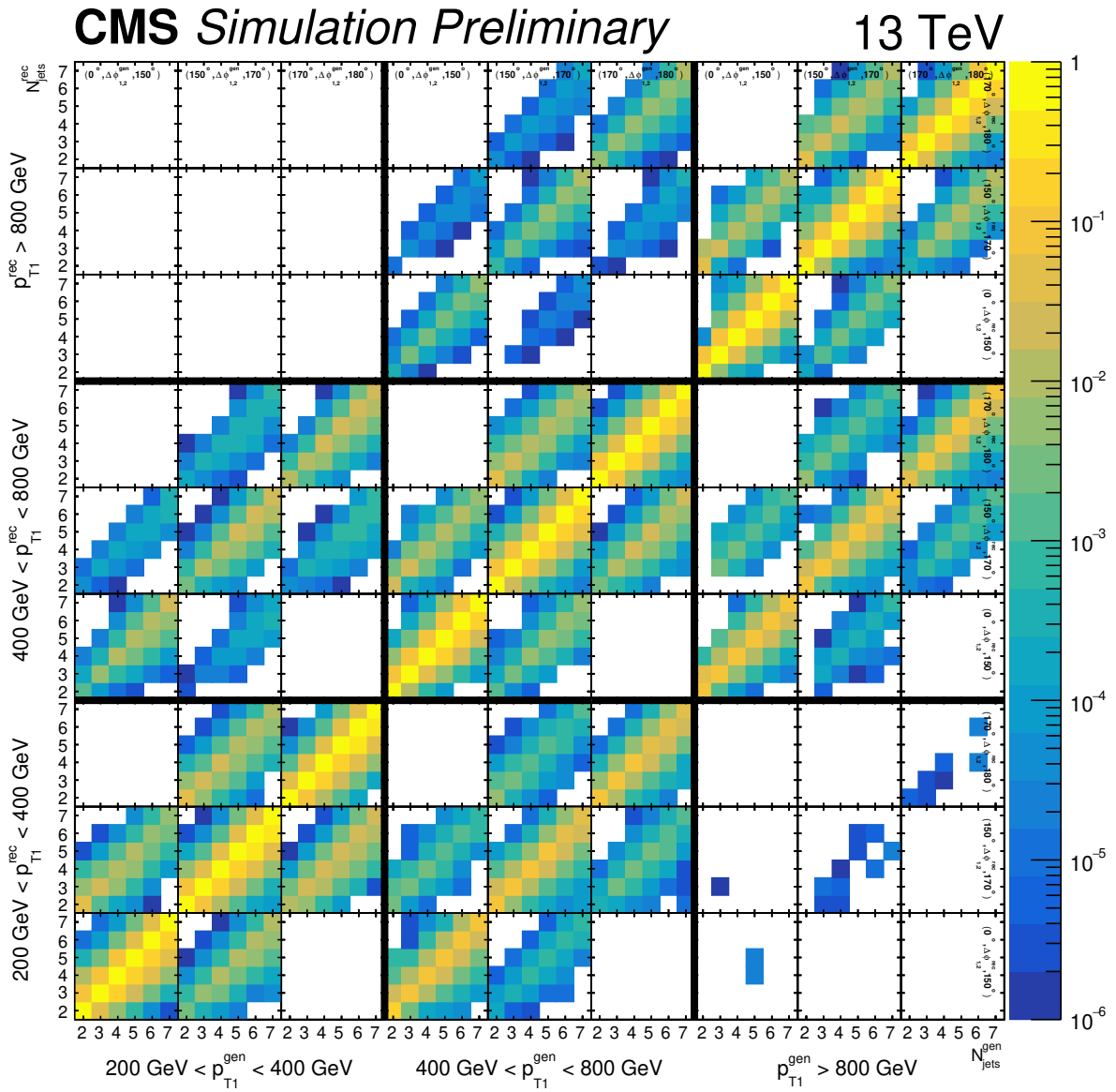


Figure 2: Probability matrix (condition number: 3.0) for the jet multiplicity distribution constructed with the MADGRAPH+PY8 sample. The global 3×3 sectors (delimited by the thick black lines) corresponds to the p_{T1} bins, indicated by the labels in the x(down) and y(left) axis; and inside this ones there are smaller 3×3 structures corresponding to the $\Delta\phi_{1,2}$ bins, indicated in the uppermost row and rightmost column, the $x(y)$-axis of these $\Delta\phi_{1,2}$ cells corresponds to the jet multiplicity at hadron(detector) level. The z-axis covers a range from 10^{-6} to 1 indicating the probability of migrations from the hadron level bin to the correspondent detector level bin.

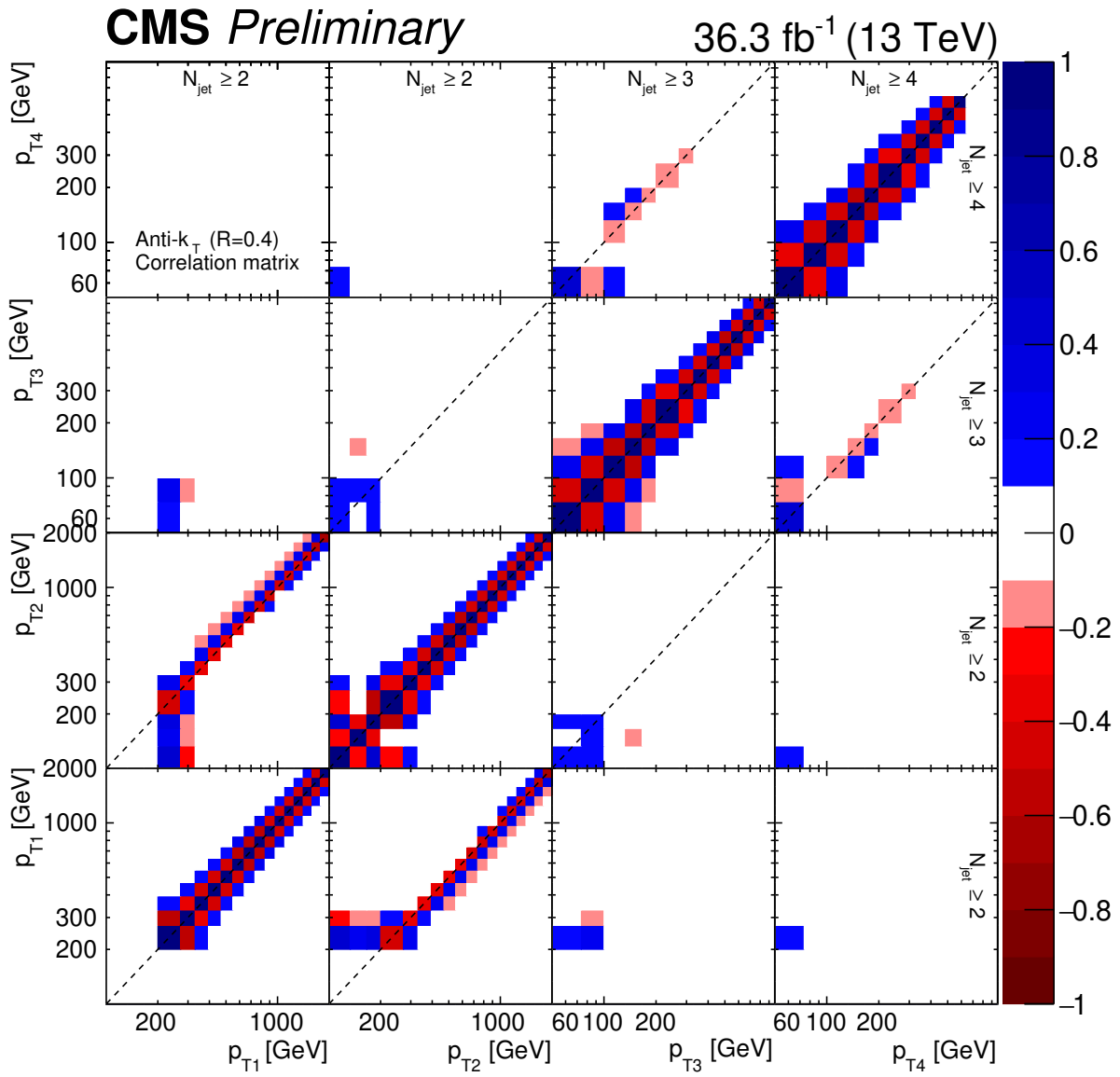


Figure 3: Correlation matrix for the hadron level p_T of the four leading jets. It contains contributions form the data recorded in 2016 and from the limited statistics from the MADGRAPH+PY8 sample. Here each one of the 4×4 sectors corresponds to one of the p_T spectra measured, indicated by the x and y-axis labels. The z-axis covers a range from -1 to 1 indicating the correlations in blue shades and anti-correlations in red shades, the values between -0.1 and 0.1 are represented in white.

the measurement of the p_T spectra of the four leading jets. In Fig. 4 we show correlation matrix for the jet multiplicity.

The systematic uncertainties originate from the following sources:

- Jet energy scale (JES)
  The JES uncertainty is estimated by using the variations of the JES corrections applied to data (at detector level) and repeating the whole unfolding procedure for each variation.
- Jet Energy Resolution (JER)
  The JER uncertainty is estimated by varying the resolution in the simulation sample,

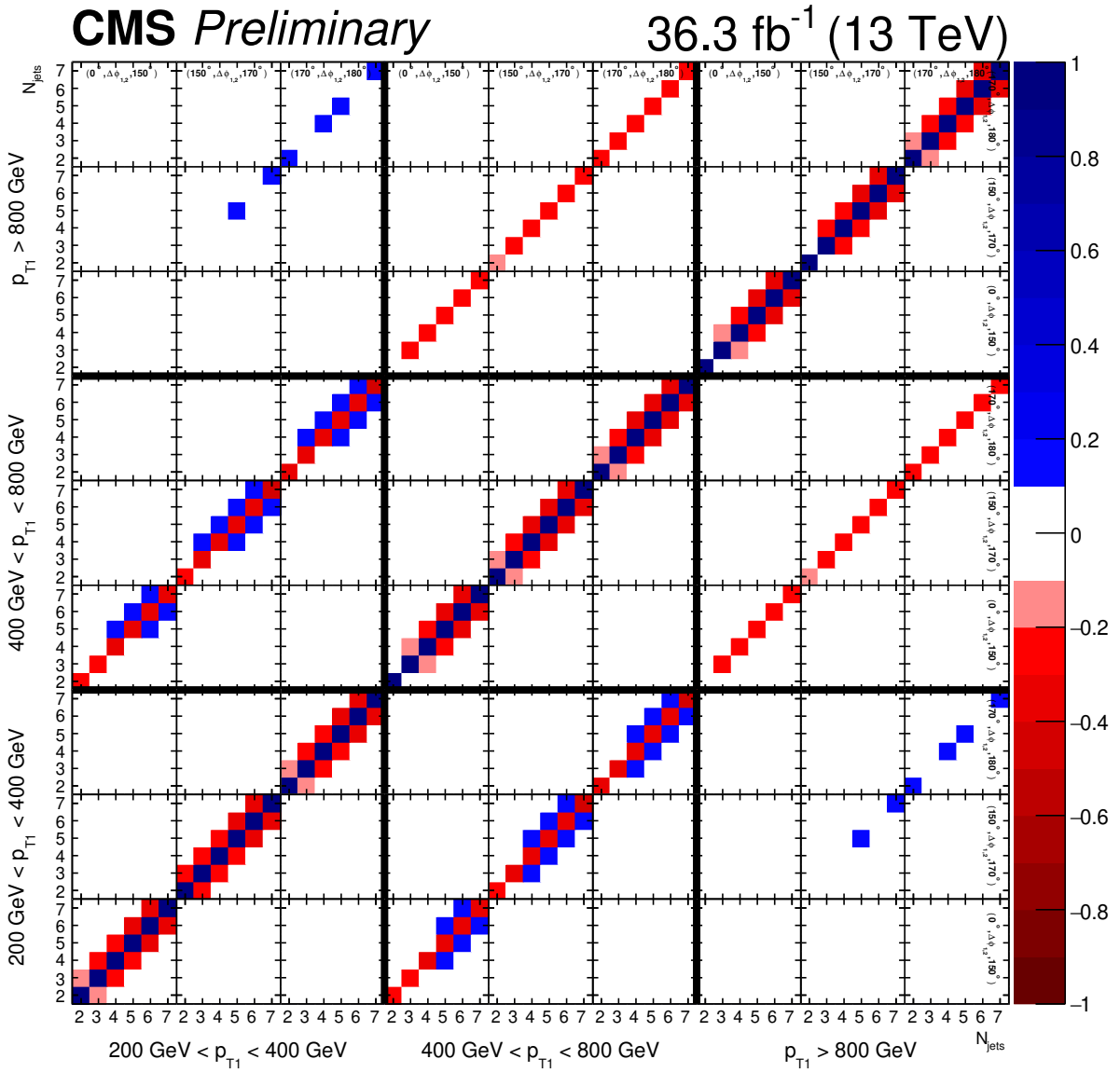


Figure 4: Correlation matrix at hadron level for the jet multiplicity distribution. It contains contributions form the data recordered in 2016 and from the limited statistics from the MADGRAPH+PY8 sample. The global 3×3 sectors (delimited by the thick black lines) corresponds to the p_{T1} bins, indicated by the labels in the x(down) and y(left) axis; and inside this ones there are smaller 3×3 structures corresponding to the $\Delta\phi_{1,2}$ bins, indicated in the uppermost row and rightmost column, the x and y-axis of these $\Delta\phi_{1,2}$ cells corresponds to the jet multiplicity. The z-axis covers a range from -1 to 1 indicating the correlations in blue shades and anti-correlations in red shades, the values between -0.1 and 0.1 are represented in white.

and repeating the unfolding for each variation.

- Luminosity (Lumi)
  The uncertainty of the integrated luminosity is 1.2% [33] and is applied as global scaling factor to the cross section.
- Pileup (PU)
  The pileup uncertainty is estimated by varying the minimum bias cross section by 5% in the simulation, affecting the measurement by less than 1%.
- Prefiring uncertainty
  The prefiring uncertainty is estimated by varying the simulated samples correction for this detector effect and repeating the unfolding for each variation, resulting in an uncertainty of $1-3\%$.
- Model uncertainty
  The model uncertainty is estimated by varying the hard scale distribution in the MC sample, and additionally varying background and inefficiencies separately by 15%. The final uncertainty is the quadratic sum of each of the uncertainties. The model uncertainty is validated with the HERWIG and PYTHIA samples, and it is about $1-7\%$.

The total systematic uncertainty is obtained by adding all systematic uncertainties in quadrature assuming independent sources.

In Fig. 5 the relative uncertainties for the jet multiplicity in bins of p_{T1} and $\Delta\phi_{1,2}$ are shown. The dominating uncertainty is `JES`. The total statistical uncertainty (`all uncorr.unc.`) is mainly composed by the data statistics (`stat.unc.`). The total experimental uncertainty (`Total`) is of the order of 10% to 15%.

In Fig. 6 the relative uncertainties as a function of p_{T1} for the four leading jets are shown. The dominant uncertainty is `JES`. The measurement is limited by systematics, and the total experimental uncertainty is of the order of 5% to 10%.

5 Results

Particles are considered stable if their mean decay length is larger than 1 cm. The phase space at stable particle level is defined by jets clustered using the anti-k_{T} algorithm [16, 17] with a distance parameter of $R = 0.4$ within $|y| < 3.2$ and requiring two highest p_{T} jets with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ and $|y| < 2.5$. For the additional jets $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ and $|y| < 2.5$ is required.

5.1 Jet multiplicity distribution

The multiplicity of jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ in $|y| < 2.5$ in addition to the dijet system is measured for different regions of the transverse momentum of the leading jet, p_{T1} and for different regions of the azimuthal angle $\Delta\phi_{1,2}$ between the two leading jets. The leading jet (subleading jet) must have a transverse momentum $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ ($p_{\mathrm{T}} > 100\,\mathrm{GeV}$), respectively. In Fig. 7 the multiplicity of the additional jets is shown for different regions of $\Delta\phi_{1,2}$ and for different regions of the transverse momentum of the leading jet, p_{T1}.

As a measure for the jet multiplicity we compare the production rate for 3 jets with the one for 7 jets. In the region of low p_{T1} ($200 < p_{T1} < 400\,\mathrm{GeV}$) a large number of additional jets is observed at low $\Delta\phi_{1,2}$ ($0^\circ < \Delta\phi_{1,2} < 150^\circ$), the production rate for 3 and 7 jets changes by two orders of magnitude. In the large $\Delta\phi_{1,2}$ region ($170^\circ < \Delta\phi_{1,2} < 180^\circ$), where the leading jets are nearly to back-to-back, the production rate for 3 and 7 jets changes by three orders of

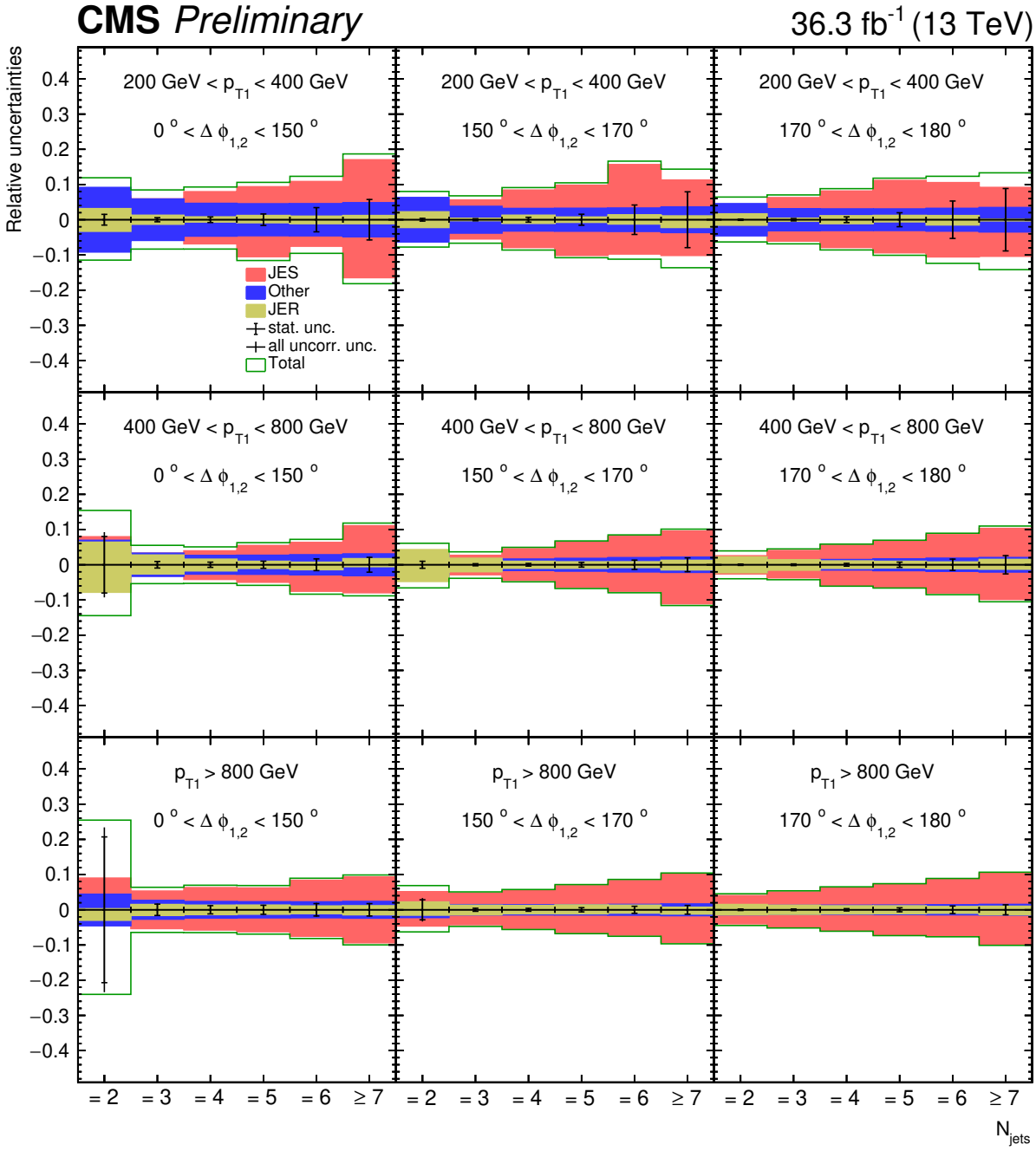


Figure 5: Relative uncertainties for the jet multiplicity distribution in bins of p_{T1} and $\Delta\phi_{1,2}$. Here `Other` includes luminosity, pileup, prefiring and unfolding model uncertainties added in quadrature.

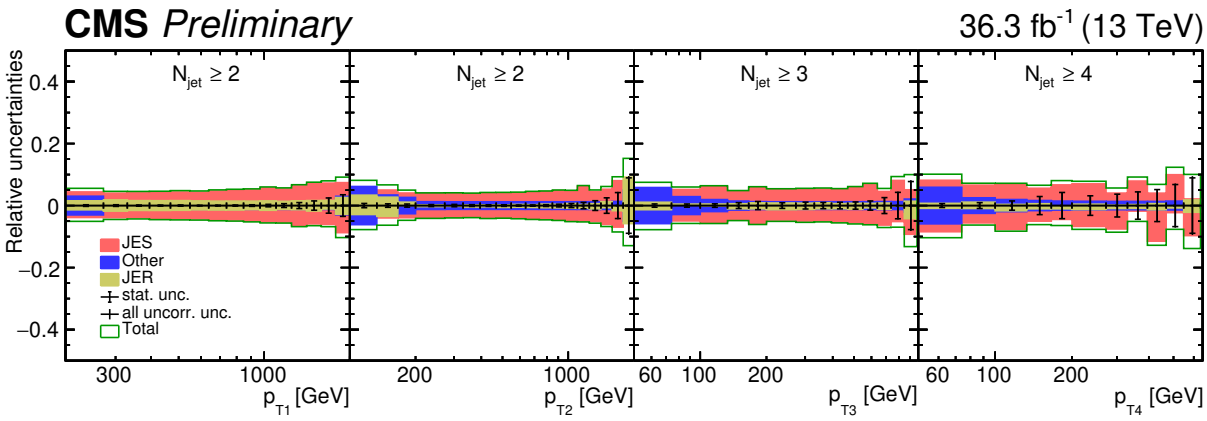


Figure 6: Relative uncertaities for the p_T distributions of the four leading jets. Here `Other` includes luminosity, pileup, prefiring and unfolding model uncertainties added in quadrature.

magnitude. It is interesting to note, that even in this back-to-back region a large number of additional jets is observed.

In the region of large p_{T1} ($p_{T1} > 800\,\text{GeV}$) we observe that the rate of additional jets at low $\Delta\phi_{1,2}$ is essentially constant, the rate between 3 and 7 jets changes only weakly, indicating that many jets participate in the compensation of the $\Delta\phi_{1,2}$ decorrelation. In the large $\Delta\phi_{1,2}$ region ($170^\circ < \Delta\phi_{1,2} < 180^\circ$) the rate between 3 and 7 jets changes by less than 2 orders of magnitude, in very contrast to the low p_{T1} region.

The measurement of the multiplicity of additional jet shows that over the whole range in $\Delta\phi_{1,2}$ more than two or three additional jets at large p_{T1} are needed to describe the data.

In Fig. 7 predictions from the LO $2 \to 2$ generators PYTHIA 8 and HERWIG++ including parton showering and multiparton interactions are shown. The shape of the predictions coming from PYTHIA 8 is different to what is observed in the measurement, especially in the region of large $\Delta\phi_{1,2}$. The shape of the prediction from HERWIG++ agrees rather well with the measurement, especially in the large $\Delta\phi_{1,2}$ region. The difference between PYTHIA 8 and HERWIG++ in jet multiplicity is coming from the different treatment of the parton shower. In addition the prediction from MADGRAPH+PY8 with up to four noncollinear high-p_T partons, supplemented with parton showering and multiparton interactions is shown. The prediction from MADGRAPH+PY8 agrees rather well in shape with the measurement.

All predicitons (LO and NLO) are normalized to the measured dijet cross section.

The calculations with NLO matrix elements matched with parton shower compared to the measurements are shown in Fig. 8. The uncertainty bands of the predictions comes from the variation of the factorization and renormalization scales by a factor of two (using 7-point variations avoiding the largest scale differences). The normalization of MG5_AMC+PY8 (jj) NLO calculation is in reasonable agreement with the measured cross section even for three jets. For higher jet multiplicities the prediction falls below the measurement. The prediction of MG5_AMC+CA3 (jj) NLO is about 20% smaller than the measurement for 2 and three jets, which is coming from the PBset2 fitted only to HERA data, corrected here by the normalization factor. Apart from the normalization, the MG5_AMC+CA3 (jj) NLO predicts a smaller cross section for more than three jets than observed in the measurement. The MG5_AMC+CA3 (jjj) NLO calculation (using the same normalization factor as for MG5_AMC+CA3 (jj)) gives a rather good description of the three and four jet cross section, while the higher jet multiplicities are still underestimated. In general, it is especially in the low p_{T1} region, where the measure-

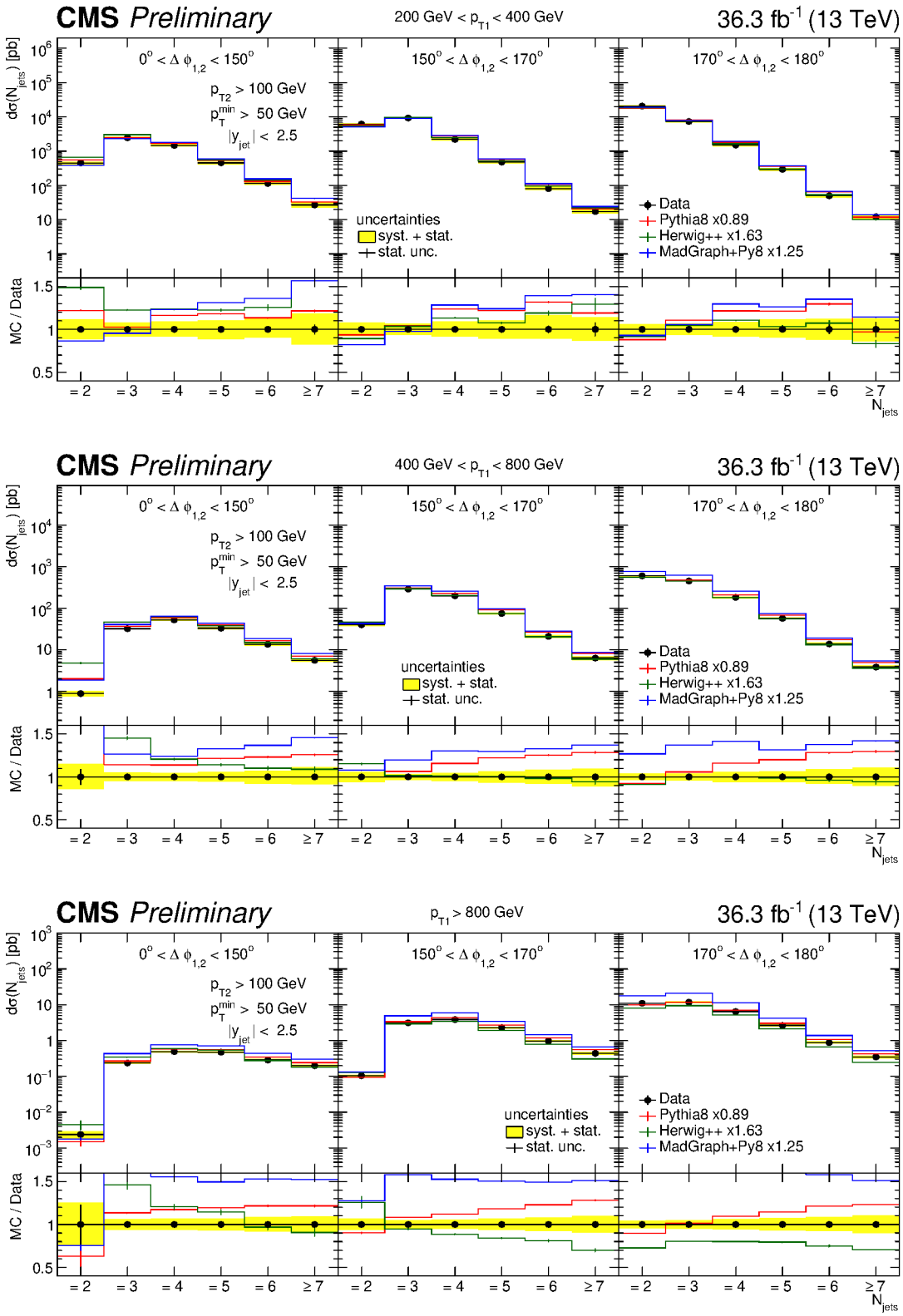


Figure 7: Differential cross section as function of the exclusive jet multiplicity (inclusive for 7 jets) in bins of p_{T1} and $\Delta\phi_{1,2}$. The data are compared to LO predictions normalized to the measured inclusive dijet cross section using the scaling factors shown in the legend.

ment is larger than the predictions.

5.2 Transverse momenta of the four leading jets

The measured differential jet cross section as a function of the jet transverse momentum, p_T, for the four leading p_T jets in dijet events is shown in Fig. 9. The p_T of the jets reaches the TeV scale. We observe that the shape of p_T spectrum for the third and fourth leading jets is similar to the one of the two leading jets, however, the cross section is different. The turn on form 100 GeV to 200 GeV for the second leading jet is a consequence of the higher minimum p_T cut (200 GeV) applied for the leading jet of the event.

In Fig. 10 the measured differential cross section as a function of the p_T for the four leading jets is shown and compared to LO predictions (using the same normalization factors as in Fig. 7). In contrast to the multiplicity distribution, none of the predictions is able to describe the measurement in shape. The prediction obtained with PYTHIA 8 describes reasonably the shape of the two leading jets, except the region $p_T < 200$ GeV for the subleading jet. The shape the third and fourth jet distributions is not well described, and PYTHIA 8 overestimates the rate up to 50 %. Similarly, the predictions from HERWIG are not in agreement in shape and rate with the measurements, the differences are up to 50% even for the leading and subleading jets at large p_T. The prediction from MADGRAPH+PY8 gives a significantly different shape of the p_T spectrum for the first 3 jets.

The predictions obtained with NLO matrix elements are shown in Fig. 11 using the same normalization factors as in Fig. 8. The prediction of MG5_AMC+PY8 (jj) describes the normalization and the shape of the first three jets rather well, while the prediction for the fourth jet (which comes from the parton shower) falls below the measurement. MG5_AMC+CA3 (jj) predicts the shapes of the distribution, even the one of the fourth jet, reasonably well. The prediction of MG5_AMC+CA3 (jjj) describes the third and fourth jets rather well described within uncertainties (predictions for the first and second jet are meaningless for MG5_AMC+CA3 (jjj) and therefore not shown). It is the first time, that calculations using PB-TMDs together with the calculation of the matrix elements in the MC@NLO frame are compared to jet measurements over a wide range in transverse momentum and jet multiplicities.

6 Conclusions

A study of multi-jet events has been performed in pp collisions at a center-of-mass energy of 13 TeV with data collected with the CMS detector corresponding to an integrated luminosity of 36.3 fb^{-1}. The measurements are performed by selecting a dijet system containing a jet with $p_T > 200$ GeV and a subleading jet with $p_T > 100$ GeV within $|y| < 2.5$.

For the first time, the jet multiplicity in bins of the leading jet p_T and the azimuthal angle between the two leading jets $\Delta\phi_{1,2}$ is measured. The differential cross section of the four leading p_T jets is measured up to the TeV scale. The jet multiplicity distributions show that even in the back-to-back region of the the dijet system, up to seven jets are measurable. The measurement of the differential cross section as a function of the jet p_T for the first four leading p_T jets is an important benchmark for Standard Model multijet cross section calculations, and especially for the simulations including parton showers for higher jet multiplicity.

The measured multiplicity distribution of jets in addition to the dijet system with $p_T > 50$ GeV and $|y| < 2.5$ is reasonably described by the LO multijet MADGRAPH+PY8 simulation, nonetheless in the back-to-back region HERWIG++ provides a better shape description. The measured differential cross section as a function of the transverse momentum of the four leading p_T jets

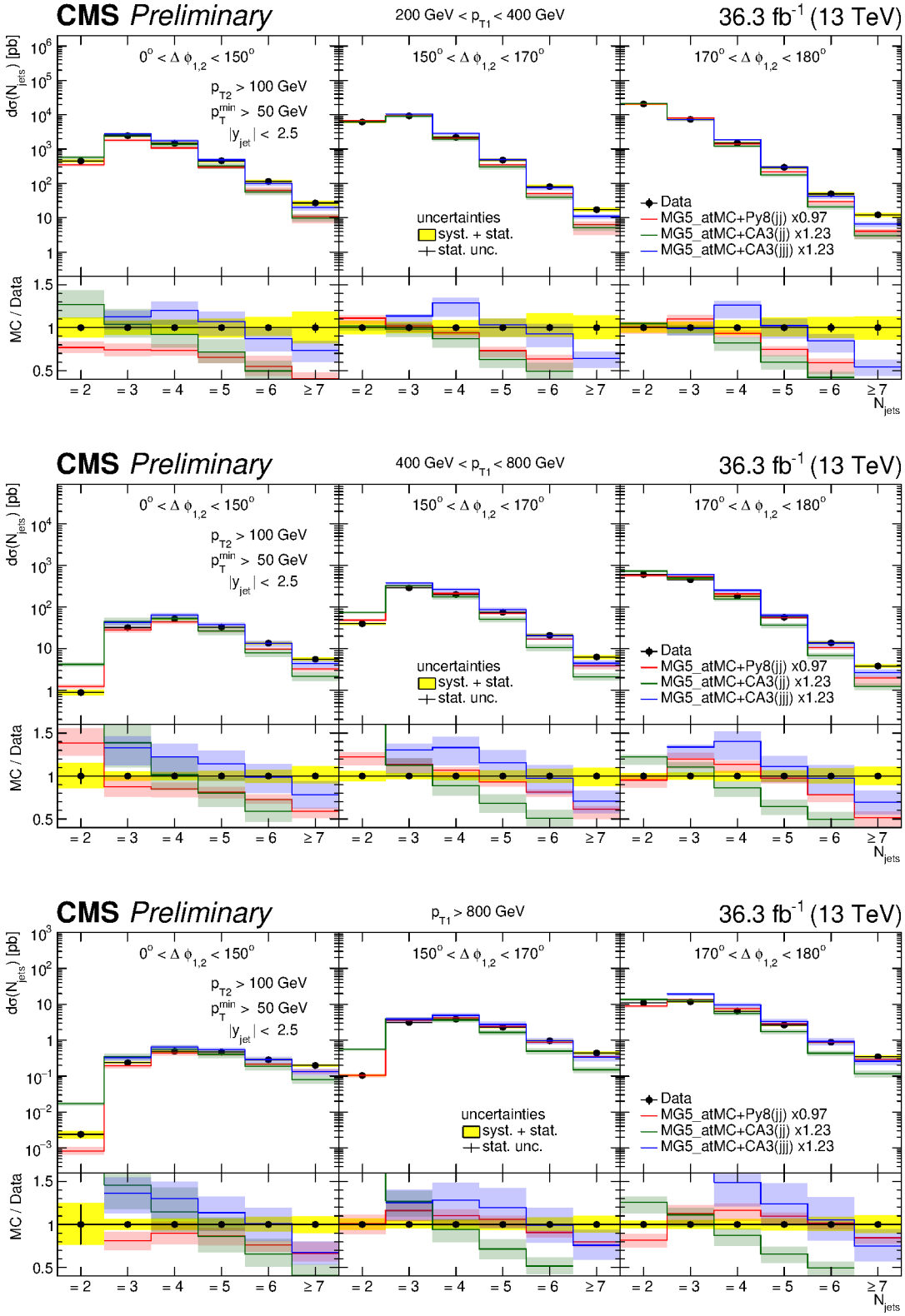


Figure 8: Differential cross section as function of the exclusive jet multiplicity (inclusive for 7 jets) in bins of p_{T1} and $\Delta\phi_{1,2}$. The data are compared to NLO dijet predictions MG5_aMC+Py8 (jj) and MG5_aMC+CA3 (jj) as well as the NLO three-jet prediction of MG5_aMC+CA3 (jjj). The predictions are normalized to the measured inclusive dijet cross section using the scaling factors shown in the legend.

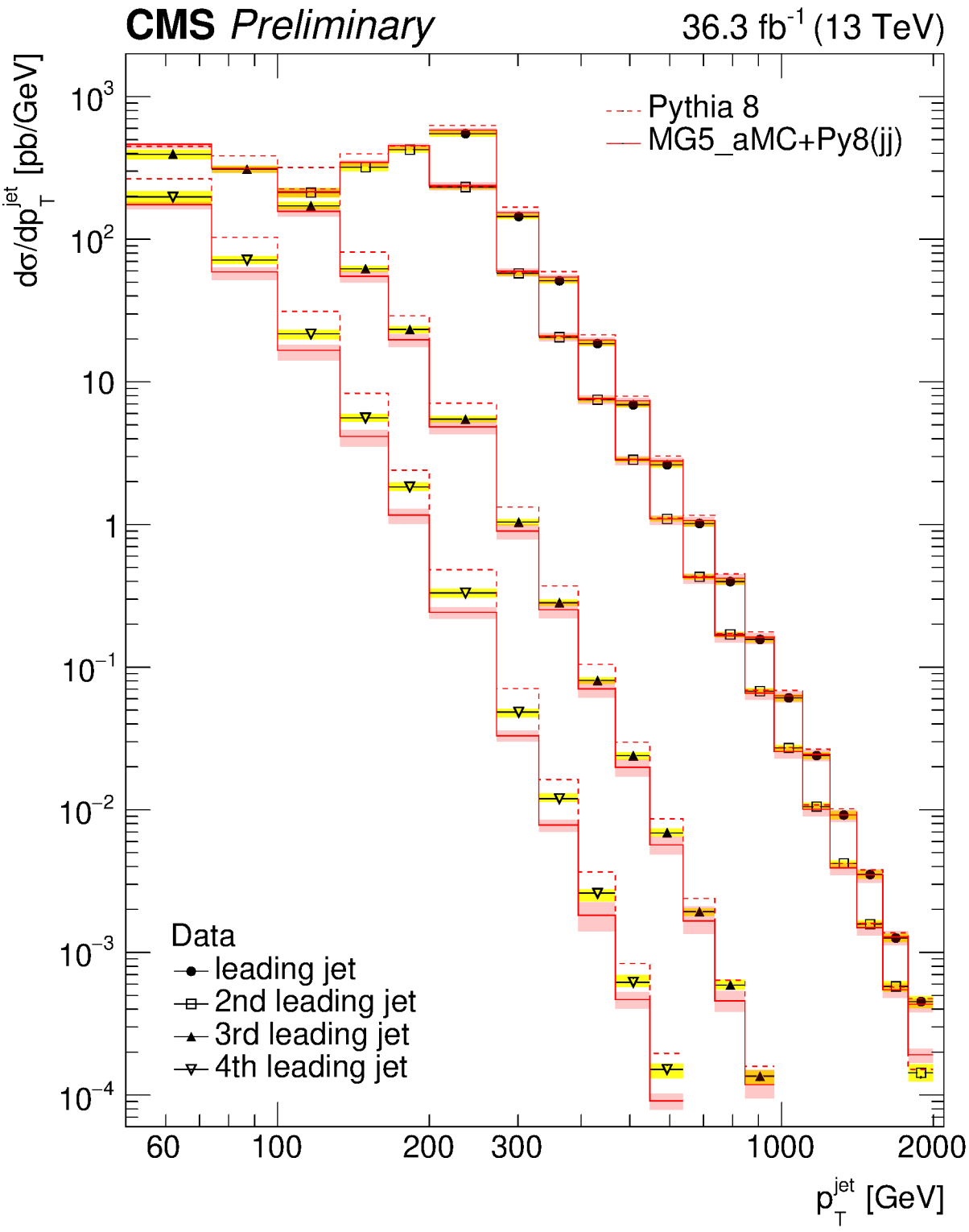
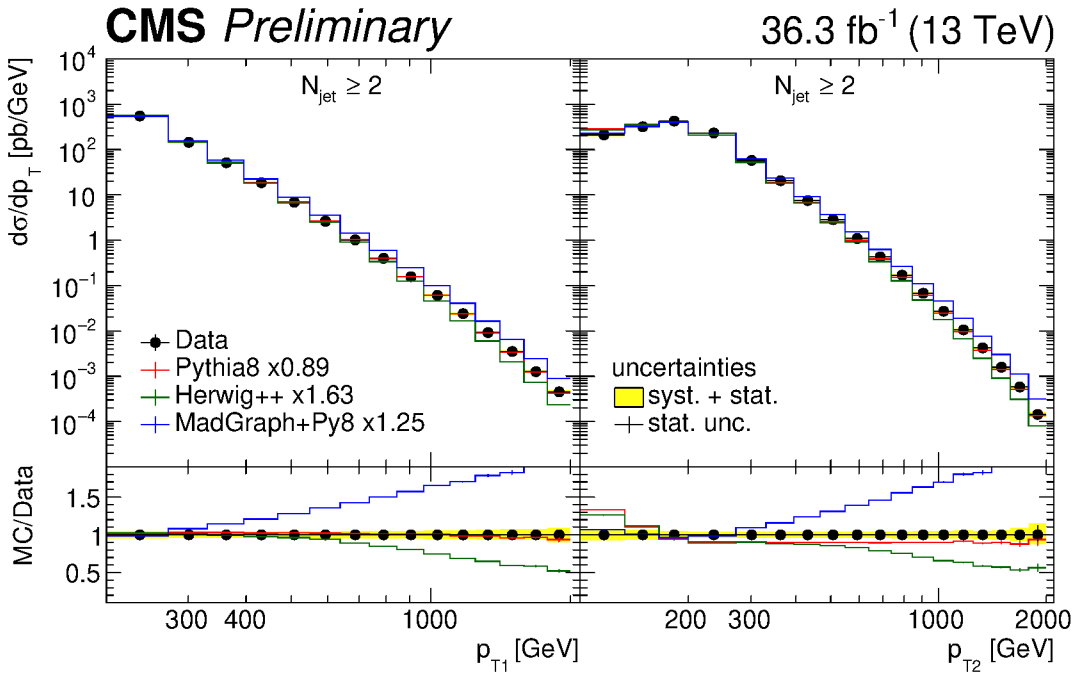
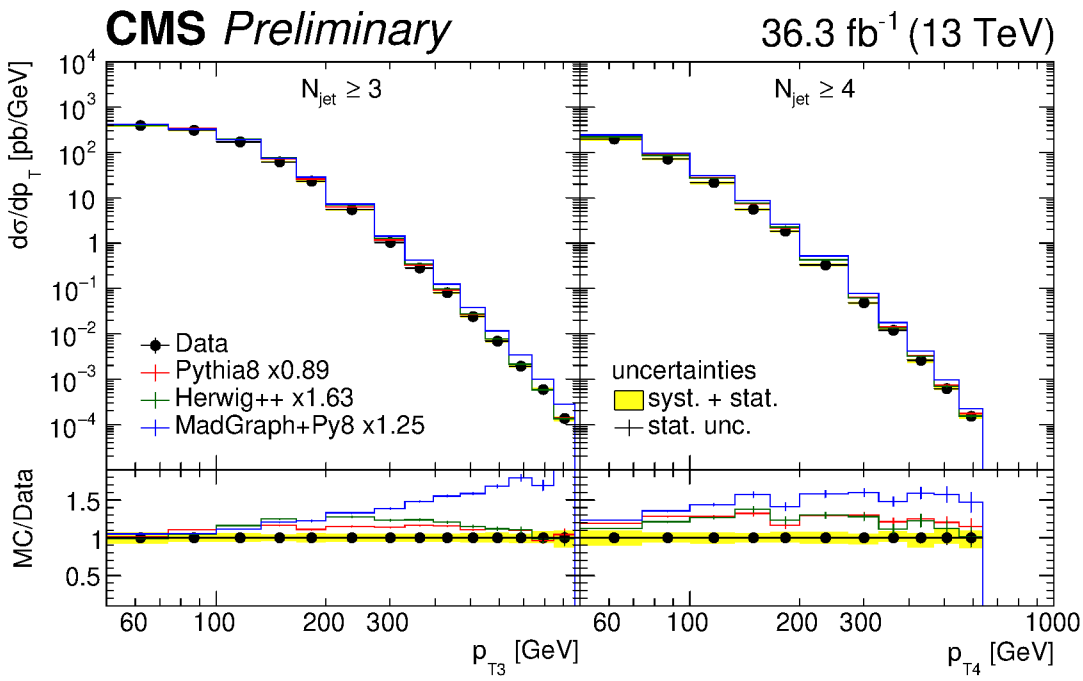


Figure 9: Transverse momenta of the measured four leading jets, here the yellow band represents the total experimental uncertainty. Data is compared to LO (PYTHIA 8) and NLO (MG5_AMC+PY8) predictions. The red band in the NLO prediction represents the scale uncertainty.

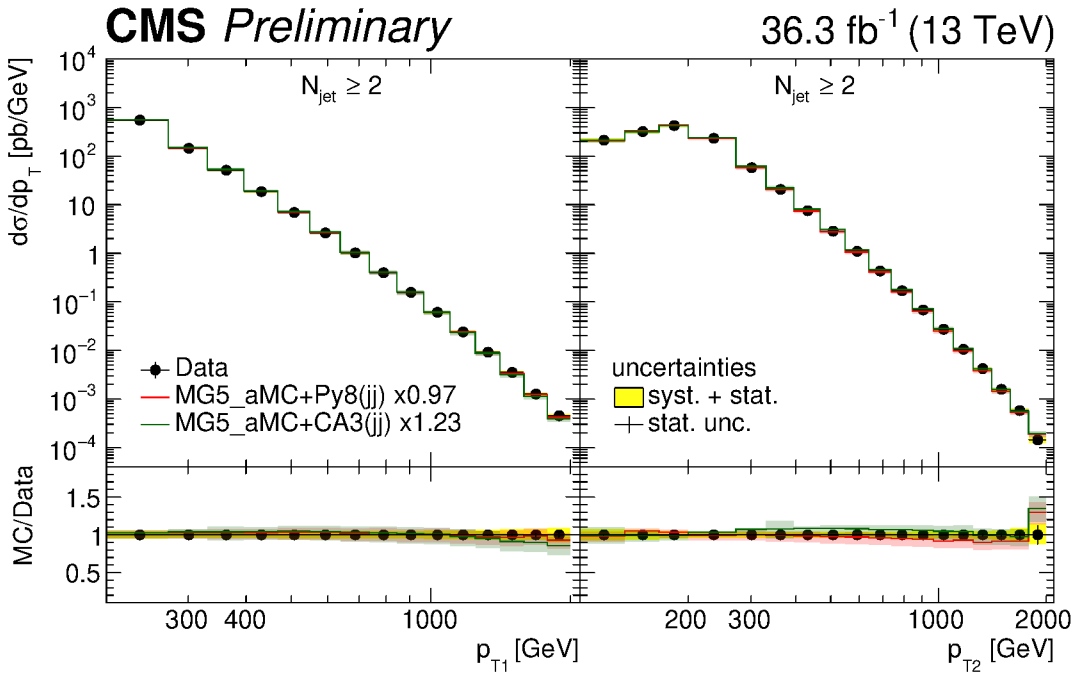


(a) Leading and second leading jet p_{T} distributions.

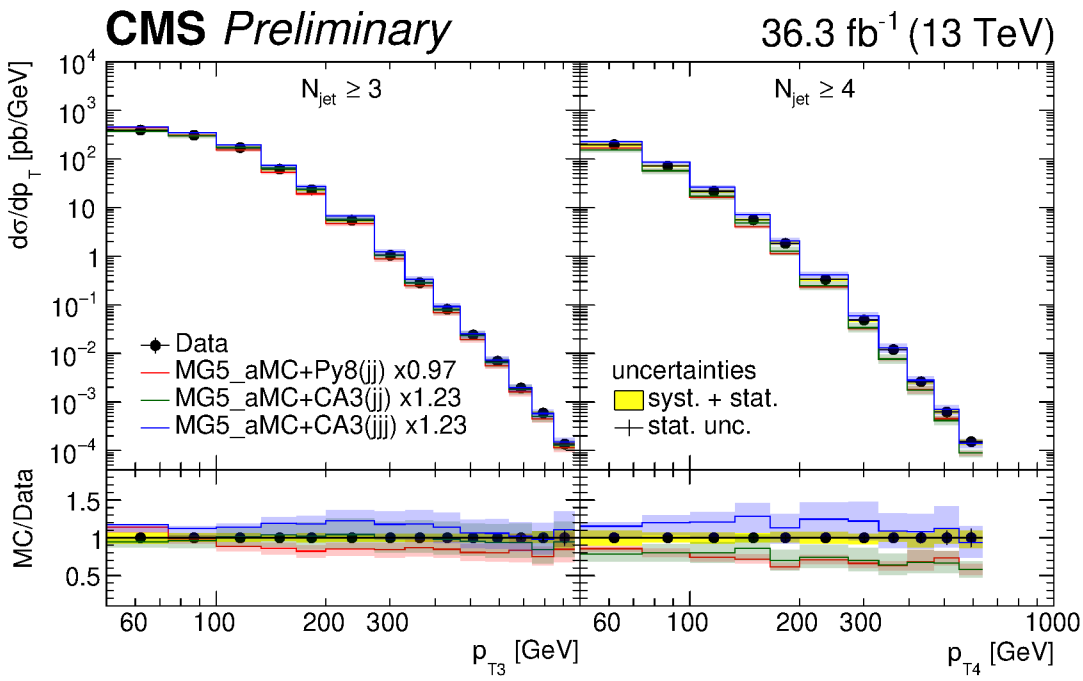


(b) Third and fourth leading jet p_{T} distributions.

Figure 10: Transverse momenta of the four leading jets compared to LO predictions normalized to the inclusive dijet data cross section using the scaling factors shown in the legend.



(a) Leading and second leading jet p_{T} distributions.



(b) Third and fourth leading jet p_{T} distributions.

Figure 11: Transverse momenta of the four leading jets compared to NLO predictions normalized to the inclusive dijet data cross section by using the scaling factors shown in the legend.

is not well described by any of the predictions, especially cross sections for the third and fourth jets are not described in normalization and shape.

The predictions using dijet NLO matrix elements, MG5_aMC+Py8 (jj) and MG5_aMC+CA3 (jj) describe the lower multiplicity regions rather well, as well as the transverse momenta of the leading jets. The cross section of the third and fourth jet is described in shape only by MG5_aMC+CA3 (jj). The three jet NLO calculation MG5_aMC+CA3 (jjj) describes very well the cross section of the third and fourth jet.

While the description of the lower jet multiplicity cross section obtained with NLO dijet calculations supplemented with conventional parton shower or with PB-TMDs and TMD parton shower is rather good, the higher jet multiplicities are not described with either parton shower approach. The measurements presented here allow for very stringent tests of theory predictions in the perturbative high p_{T} and high jet multiplicity regions.

References

[1] A. Bermudez Martinez et al., "Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method", *Phys. Rev. D* **99** (2019) 074008, doi:10.1103/PhysRevD.99.074008, arXiv:1804.11152.

[2] F. Hautmann et al., "Soft-gluon resolution scale in QCD evolution equations", *Phys. Lett. B* **772** (2017) 446, doi:10.1016/j.physletb.2017.07.005, arXiv:1704.01757.

[3] F. Hautmann et al., "Collinear and TMD quark and gluon densities from parton branching solution of QCD evolution equations", *JHEP* **01** (2018) 070, doi:10.1007/JHEP01(2018)070, arXiv:1708.03279.

[4] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[5] S. Baranov et al., "CASCADE3 A Monte Carlo event generator based on TMDs", *Eur. Phys. J. C* **81** (2021) 425, doi:10.1140/epjc/s10052-021-09203-8, arXiv:2101.10221.

[6] D0 Collaboration, "Measurement of dijet azimuthal decorrelations at central rapidities in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **94** (2005) 221801, doi:10.1103/PhysRevLett.94.221801, arXiv:hep-ex/0409040.

[7] D0 Collaboration, "Measurement of the combined rapidity and p_{T} dependence of dijet azimuthal decorrelations in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Lett. B* **721** (2013) 212, doi:10.1016/j.physletb.2013.03.029, arXiv:1212.1842.

[8] ATLAS Collaboration, "Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 172002, doi:10.1103/PhysRevLett.106.172002, arXiv:1102.2696.

[9] CMS Collaboration, "Dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 122003, doi:10.1103/PhysRevLett.106.122003, arXiv:1101.5029.

[10] CMS Collaboration, "Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **76** (2016) 536, doi:10.1140/epjc/s10052-016-4346-8, arXiv:1602.04384.

[11] CMS Collaboration, "Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **78** (2018) 566, doi:10.1140/epjc/s10052-018-6033-4, arXiv:1712.05471.

[12] CMS Collaboration, "Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **79** (2019) 773, doi:10.1140/epjc/s10052-019-7276-4, arXiv:1902.04374.

[13] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[14] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[15] CMS Collaboration, "Particle-flow reconstruction and global event description with the cms detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[16] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[17] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[18] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[19] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector", *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[20] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[21] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[22] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[23] NNPDF Collaboration, "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO", *Nucl. Phys. B* **855** (2012) 153, doi:10.1016/j.nuclphysb.2011.09.024, arXiv:1107.2652.

[24] M. Bahr et al., "Herwig++ physics and manual", *Eur. Phys. J. C* **58** (2008) 639, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[25] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.

[26] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[27] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[28] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[29] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[30] CMS Collaboration, "Tracking and primary vertex results in first 7 TeV collisions", CMS Physics Analysis Summary CMS-PAS-TRK-10-005, CERN, Geneva, 2010.

[31] CMS Collaboration, "Calorimeter jet quality criteria for the first CMS collision data", CMS Physics Analysis Summary CMS-PAS-JME-09-008, CERN, Geneva, 2010.

[32] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *JINST* **7** (2012) T10003, doi:10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.

[33] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", 2021. arXiv:2104.01927. Submitted to *Eur. Phys. J. C*.

Eur. Phys. J. C (2021) 81:425
https://doi.org/10.1140/epjc/s10052-021-09203-8

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

CASCADE3 A Monte Carlo event generator based on TMDs

S. Baranov[1], A. Bermudez Martinez[2], L. I. Estevez Banos[2], F. Guzman[3], F. Hautmann[4,5], H. Jung[2,a], A. Lelek[4], J. Lidrych[2], A. Lipatov[6], M. Malyshev[6], M. Mendizabal[2], S. Taheri Monfared[2], A. M. van Kampen[4], Q. Wang[2,7], H. Yang[2,7]

[1] Lebedev Physics Institute, Moscow, Russia
[2] DESY, Hamburg, Germany
[3] InSTEC, Universidad de La Habana, Havana, Cuba
[4] Elementary Particle Physics, University of Antwerp, Antwerp, Belgium
[5] RAL and University of Oxford, Oxford, UK
[6] SINP, Moscow State University, Moscow, Russia
[7] School of Physics, Peking University, Beijing, China

Received: 1 February 2021 / Accepted: 2 May 2021 / Published online: 16 May 2021

Abstract The CASCADE3 Monte Carlo event generator based on Transverse Momentum Dependent (TMD) parton densities is described. Hard processes which are generated in collinear factorization with LO multileg or NLO parton level generators are extended by adding transverse momenta to the initial partons according to TMD densities and applying dedicated TMD parton showers and hadronization. Processes with off-shell kinematics within k_t-factorization, either internally implemented or from external packages via LHE files, can be processed for parton showering and hadronization. The initial state parton shower is tied to the TMD parton distribution, with all parameters fixed by the TMD distribution.

1 Introduction

The simulation of processes for high energy hadron colliders has been improved significantly in the past years by automation of next-to-leading order (NLO) calculations and matching of the hard processes to parton shower Monte Carlo event generators which also include a simulation of hadronization. Among those automated tools are the MADGRAPH5_aMC@NLO [1] generator based on the MC@NLO [2–5] method or the POWHEG [6,7] generator for the calculation of the hard process. The results from these packages are then combined with either the HERWIG [8] or PYTHIA [9] packages for parton showering and hadronization. Different jet multiplicities can be combined at the matrix element level and then merged with special procedures, like the MLM [10] or CKKW [11] merging for LO processes, the FxFx [12] or MiNLO method [13] for merging at NLO, among others. While the approaches of matching and merging matrix element calculations and parton showers are very successful, two ingredients important for high energy collisions are not (fully) treated: the matrix elements are calculated with collinear dynamics and the inclusion of initial state parton showers results in a net transverse momentum of the hard process; the special treatment of high energy effects (small x) is not included.

The CASCADE Monte Carlo event generator, developed originally for small x processes based on high-energy factorization [14] and the CCFM [15–18] evolution equation, has been extended to cover the full kinematic range (not only small x) by applying the Parton Branching (PB) method and the corresponding PB Transverse Momentum Dependent (TMD) parton densities [19,20]. The initial state evolution is fully described and determined by the TMD density, as it was in the case of the CCFM gluon density, but now available for all flavor species, including quarks, gluons and photons at small and large x and any scale μ. For a general overview of TMD parton densities, see Ref. [21].

With the advances in determination of PB TMDs [19,20], it is natural to develop a scheme, where the initial parton shower follows as close as possible the TMD parton density and where either collinear (on-shell) or k_t-dependent (off-shell) hard process calculations can be included at LO or NLO. In order to be flexible and to use the latest developments in automated matrix element calculations of hard process at higher order in the strong coupling α_s, events available in the Les Houches Event (LHE) file format [22], which contains all the information of the hard process including the color

[a] e-mail: jung@mail.desy.de (corresponding author)

structure, can be further processed for parton shower and hadronization in CASCADE3.

In this report we describe the new developments in CASCADE3 for a full PB-TMD parton shower and the matching of TMD parton densities to collinear hard process calculations. We also mention features of the small-x mode of CASCADE3.

2 The hard process

The cross section for the scattering process of two hadrons A and B can be written in collinear factorization as a convolution of the partonic cross section of partons a and b, $a + b \rightarrow X$, and the densities $f_{a(b)}(x, \mu)$ of partons a (b) inside the hadrons A (B),

$$\sigma(A + B \rightarrow Y) = \int dx_a \int dx_b \, f_a(x_a, \mu) \, f_b(x_b, \mu) \, \sigma(a + b \rightarrow X), \tag{1}$$

where $x_a(x_b)$ are the fractions of the longitudinal momenta of hadrons A, B carried by the partons $a(b)$, $\sigma(a + b \rightarrow X)$ is the partonic cross section, and μ is the factorization scale of the process. The final state Y contains the partonic final state X and the recoils from the parton evolution and hadron remnants.

In CASCADE3 we extend collinear factorization to include transverse momenta in the initial state, either by adding a transverse momentum to an on-shell process or by using off-shell processes directly, as described in detail in Sects. 2.1 and 2.2 . TMD factorization is proven for semi-inclusive deep-inelastic scattering, Drell–Yan production in hadron-hadron collisions and e^+e^- annihilation [23–35]. In the high-energy limit (small-x) k_T-factorization has been formulated also in hadronic collisions for processes like heavy flavor or heavy boson (including Higgs) production [14,36–38], with so-called *unintegrated* parton distribution functions (uPDFs), see e.g. Refs. [39–49].

2.1 On-shell processes

The hard processes in collinear factorization (with on-shell initial partons, without transverse momenta) can be calculated by standard automated methods like MADGRAPH5_AMC@NLO[1] for multileg processes at LO or NLO accuracy. The matrix element processes are calculated with collinear parton densities (PDF), as provided by LHAPDF [50].

We extend the factorization formula given in Eq. (1) by replacing the collinear parton densities $f(x, \mu)$ by TMD densities $\mathcal{A}(x, k_t, \mu)$ with k_t being the transverse momentum of the interacting parton, and integrating over the transverse momenta.

However, when the hard process is to be combined with a TMD parton density, as described later, the integral over k_t of the TMD density must agree with the collinear (k_t-integrated) density; this feature is guaranteed by construction for the PB-TMDs (also available as integrated PDFs in LHAPDF format).

In a LO partonic calculation the TMD or the parton shower can be included respecting energy momentum conservation, as described below. In an NLO calculation based on the MC@NLO method [2–5] the contribution from collinear and soft partons is subtracted, as this is added later with the parton shower. For the use with PB TMDs, the HERWIG6 subtraction terms are best suited as the angular ordering conditions coincide with those applied in the PB-method. The PB TMDs play the same role as a parton shower does, in the sense that a finite transverse momentum is created as a result of the parton evolution [51,52].

When transverse momenta of the initial partons from TMDs are to be included to the hard scattering process, which was originally calculated under the assumption of collinear initial partons, care has to be taken that energy and momentum are still conserved. When the initial state partons have transverse momenta, they also acquire virtual masses. The procedure adopted in CASCADE3 is the following: for each initial parton, a transverse momentum is assigned according to the TMD density, and the parton-parton system is boosted to its center-of-mass frame and rotated such that only the longitudinal and energy components are non-zero. The energy and longitudinal component of the initial momenta $p_{a,b}$ are recalculated taking into account the virtual masses $Q_a^2 = k_{t,a}^2$ and $Q_b^2 = k_{t,b}^2$ [53],

$$E_{a,b} = \frac{1}{2\sqrt{\hat{s}}} \left(\hat{s} \pm (Q_b^2 - Q_a^2) \right) \tag{2}$$

$$p_{z\,a,b} = \pm \frac{1}{2\sqrt{\hat{s}}} \sqrt{(\hat{s} + Q_a^2 + Q_b^2)^2 - 4 Q_a^2 Q_b^2} \tag{3}$$

with $\hat{s} = (p_a + p_b)^2$ with $p_a(p_b)$ being the four-momenta of the interacting partons a and b. The partonic system is then rotated and boosted back to the overall center-of-mass system of the colliding particles. By this procedure, the parton-parton mass $\sqrt{\hat{s}}$ is exactly conserved, while the rapidity of the partonic system is approximately restored, depending on the transverse momenta.

In Fig. 1 a comparison of the Drell-Yan (DY) mass, transverse momentum and rapidity is shown for an NLO calculation of DY production in pp collisions at $\sqrt{s} = 13$ TeV in the mass range $30 < m_{DY} < 2000$ GeV. The curve labelled NLO(LHE) is the calculation of MADGRAPH5_AMC@NLO with the subtraction terms, the curve NLO(LHE+TMD) is the prediction after the transverse momentum is included

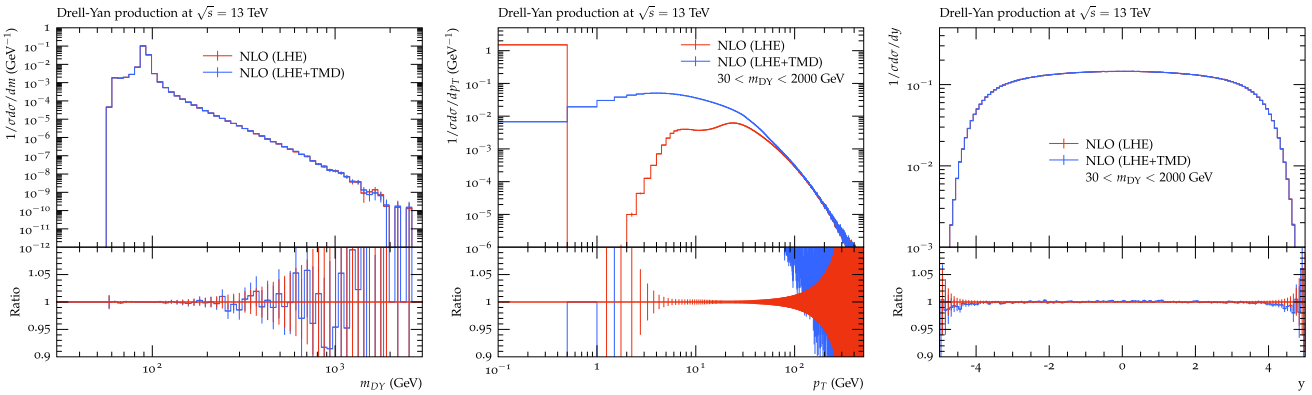


Fig. 1 Distributions of Drell-Yan mass, transverse momentum and rapidity for $pp \to DY + X$ at $\sqrt{s} = 13$ TeV. The hard process is calculated with MADGRAPH5_aMC@NLO. NLO(LHE) is the prediction including subtraction terms, NLO(LHE+TMD) includes transverse momenta of the interacting partons according to the description in the text

according to the procedure described above. In the p_T spectrum one can clearly see the effect of including transverse momenta from the TMD distribution. The DY mass distribution is not changed, and the rapidity distribution is almost exactly reproduced, only at large rapidities small differences are observed.

The transverse momenta k_t are generated according to the TMD density $\mathcal{A}(x, k_t, \mu)$, at the original longitudinal momentum fraction x and the hard process scale μ. In a LO calculation, the full range of k_t is available, but in an NLO calculation via the MC@NLO method a *shower scale* defines the boundary between parton shower and real emissions from the matrix element, limiting the transverse momentum k_t. Technically the factorization scale μ is calculated within CASCADE3 (see parameter `lhescale`) as it is not directly accessible from the LHE file, while the *shower scale* is given by `SCALUP`. The limitation of the transverse momenta coming from the TMD distribution and TMD shower to be smaller than the *shower scale* SCALUP guarantees that the overlap with real emissions from the matrix element is minimized according to the subtraction of counterterms in the MC@NLO method.

The advantage of using TMDs for the complete process is that the kinematics are fixed, independent of simulating explicitly the radiation history from the parton shower. For inclusive processes, for example inclusive Drell-Yan processes, the details of the hadronic final state generated by a parton shower do not matter, and only the net effect of the transverse momentum distribution is essential. However, for processes which involve jets, the details of the parton shower become also important. The parton shower, as described below, follows very closely the transverse momentum distribution of the TMD and thus does not change any kinematic distribution after the transverse momentum of the initial partons are included.

All hard processes, which are available in MADGRAPH5_aMC@NLO can be used within CASCADE3. The treatment of multijet merging is described in Sect. 8.

2.2 Off-shell processes

In a region of phase space, where the longitudinal momentum fractions x become very small, the transverse momentum of the partons cannot be neglected and has to be included already at the matrix element level, leading to so-called *off-shell* processes.

In off-shell processes a natural suppression at large k_t [54] (with $k_t > \mu$) is obtained, shown explicitly in Fig. 2, where the matrix element for $g^*g^* \to Q\bar{Q}$, with Q being a heavy quark, is considered. The process is integrated over the final state phase space [55],

$$\tilde{\sigma}(k_t) = \int \frac{dx_2}{x_2}\, d\phi_{1,2}\, \mathrm{dLips}\, |ME|^2\, (1-x_2)^5\ , \tag{4}$$

where dLips is the Lorentz-invariant phase space of the final state, ME is the matrix-element for the process, $\phi_{1,2}$ is the azimuthal angle between the two initial partons, and a simple scale-independent and k_t-independent gluon density $xG(x) = (1-x)^5$ is included which suppresses large-x contributions. In Fig. 2 we show $\tilde{\sigma}(k_t)$ normalized to its on-shell value $\tilde{\sigma}(0)$ at $\sqrt{s} = 13000$ GeV as a function of the transverse momentum of the incoming gluon $k_{t,2}$ for different values of x_1, which are chosen such that the ratio $k_{t,1}^2/(x_1 s)$ is kept constant.

In Fig. 2 (left) predictions are shown for bottom quarks with mass $m = 5$ GeV and different $k_{t,1}$, in Fig. 2 (right) a comparison is made for different heavy quark masses. Using off-shell matrix elements a suppression at large transverse momenta of the initial partons is obtained, depending on the heavy flavor mass and the transverse momentum. In a collinear approach, with implicit integration over transverse

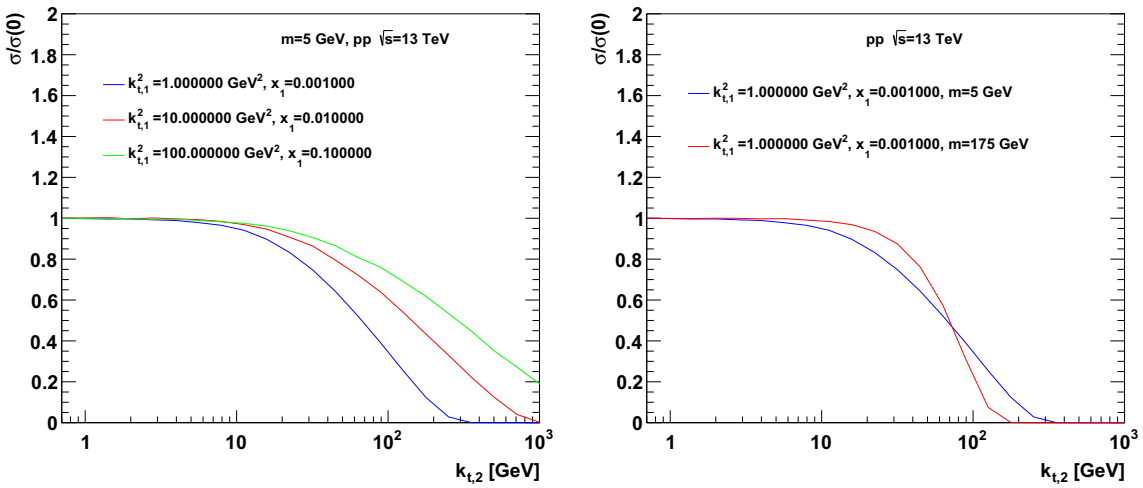


Fig. 2 The reduced cross section $\tilde{\sigma}(k_t)/\tilde{\sigma}(0)$ as a function of the transverse momentum $k_{t,2}$ of the incoming gluon at $\sqrt{s} = 13000$ GeV. (Left) for different values of $k_{t,1}$ and x_1, (right) for different heavy flavor masses and fixed values of $k_{t,1}$ and x_1

momenta of the initial state partons, the transverse momenta are limited by a theta function at the factorization scale, while off-shell matrix elements give a smooth transition to a high k_t tail.

When using off-shell processes, BFKL or CCFM type parton densities should be used to cover the full available phase space in transverse momentum, which can lead to k_t's larger than the transverse momentum of any of the partons of the hard process [56]. Until now, only gluon densities obtained from CCFM [15–18] or BFKL [57–59] are available, thus limiting the advantages of using off-shell matrix elements to gluon induced processes.

Several processes with off-shell matrix elements are implemented in CASCADE3 as listed in Table 1, and described in detail in [60]. However, many more processes are accessible via the automated matrix element calculators for off-shell processes, KATIE [61] and PEGASUS [62]. The events from the hard process are then read with the CASCADE3 package via LHE files. For processes generated with KATIE or PEGASUS no further corrections need to be performed and the event can be directly passed to the showering procedure, described in the next section.

3 Initial state parton shower based on TMDs

The parton shower, which is described here, follows consistently the parton evolution of the TMDs. By this we mean that the splitting functions P_{ab}, the order and the scale in α_s as well as kinematic restrictions are identical to both the parton shower and the evolution of the parton densities (for NLO PB TMD densities, the NLO DGLAP splitting functions [73,74] together with NLO α_s is applied, while for the LO TMD densities the corresponding LO splitting functions [75–77] and LO α_s is used).

3.1 From PB TMD evolution to TMD parton shower

The PB method describes the TMD parton density as (cf Eq. (2.43) in Ref. [19])

$$
\begin{aligned}
x\mathcal{A}_a(x,k_t,\mu) &= \Delta_a(\mu)\, x\mathcal{A}_a(x,k_t,\mu_0) \\
&\quad + \sum_b \int \frac{dq^2}{q^2}\frac{d\phi}{2\pi}\frac{\Delta_a(\mu)}{\Delta_a(q)}\,\Theta(\mu-q)\,\Theta(q-\mu_0) \\
&\quad \times \int_x^{z_M} dz\; P_{ab}^{(R)}(\alpha_s(f(z,q)),z)\,\frac{x}{z}\mathcal{A}_b\left(\frac{x}{z},k_t',q\right)\ ,
\end{aligned}
\tag{5}
$$

with $z_M < 1$ defining resolvable branchings, $\mathbf{k}$ ($\mathbf{q}_c$) being the transverse momentum vector of the propagating (emitted) parton, respectively. The transverse momentum of the parton before branching is defined as $k_t' = |\mathbf{k} + (1-z)\mathbf{q}|$ with $\mathbf{q} = \mathbf{q}_c/(1-z)$ being the rescaled transverse momentum vector of the emitted parton (see Fig. 3, with the notation $k_t = |\mathbf{k}|$ and $q = |\mathbf{q}|$) and ϕ being the azimuthal angle between $\mathbf{q}$ and $\mathbf{k}$. The argument in α_s is in general a function of the evolution scale q. Higher order calculations indicate the transverse momentum of the emitted parton as the preferred scale. The real emission branching probability is denoted by $P_{ab}^{(R)}(\alpha_s(f(z,q)),z)$ including α_s as described in Ref. [19] (in the following we omit α_s in the argument of $P_{ab}^{(R)}$ for easier reading). The Sudakov form factor is given by:

$$
\Delta_a(z_M,\mu,\mu_0) = \exp\left(-\sum_b \int_{\mu_0^2}^{\mu^2}\frac{dq^2}{q^2}\int_0^{z_M} dz\; z\; P_{ba}^{(R)}\right).
\tag{6}
$$

Table 1 Processes included in CASCADE3. Q stands for heavy quarks, q for light quarks

Lepto(photo)production	Process	IPRO	References
	$\gamma^* g^* \to q\bar{q}$	10	[63]
	$\gamma^* g^* \to Q\bar{Q}$	11	[63]
	$\gamma^* g^* \to J/\psi g$	2	[64–67]
Hadroproduction			
	$g^* g^* \to q\bar{q}$	10	[63]
	$g^* g^* \to Q\bar{Q}$	11	[63]
	$g^* g^* \to J/\psi g$	2	[67]
	$g^* g^* \to \Upsilon g$	2	[67]
	$g^* g^* \to \chi_c$	3	[67]
	$g^* g^* \to \chi_b$	3	[67]
	$g^* g^* \to J/\psi J/\psi$	21	[68]
	$g^* g^* \to h^0$	102	[38]
	$g^* g^* \to ZQ\bar{Q}$	504	[69,70]
	$g^* g^* \to Zq\bar{q}$	503	[69,70]
	$g^* g^* \to Wq_i Q_j$	514	[69,70]
	$g^* g^* \to Wq_i q_j$	513	[69,70]
	$qg^* \to Zq$	501	[71]
	$qg^* \to Wq$	511	[71]
	$qg^* \to qg$	10	[72]
	$gg^* \to gg$	10	[72]

Dividing Eq. (5) by $\Delta_a(\mu^2)$ and differentiating with respect to μ^2 gives the differential form of the evolution equation describing the probability for resolving a parton with transverse momentum $\mathbf{k}'$ and momentum fraction x/z into a parton with momentum fraction x and emitting another parton during a small decrease of μ,

$$\mu^2 \frac{d}{d\mu^2}\left(\frac{x\mathcal{A}_a(x,k_t,\mu)}{\Delta_a(\mu)}\right) = \sum_b \int_x^{z_M} dz \frac{d\phi}{2\pi} P_{ab}^{(R)} \frac{x}{z} \frac{\mathcal{A}_b\left(\frac{x}{z},k_t',\mu\right)}{\Delta_a(\mu)} . \tag{7}$$

The normalized probability is then given by

$$\frac{\Delta_a(\mu)}{x\mathcal{A}_a(x,k_t,\mu)} d\left(\frac{x\mathcal{A}_a(x,k_t,\mu)}{\Delta_a(\mu)}\right) = \sum_b \frac{d\mu^2}{\mu^2} \int_x^{z_M} dz \frac{d\phi}{2\pi} P_{ab}^{(R)} \frac{\frac{x}{z}\mathcal{A}_b\left(\frac{x}{z},k_t',\mu\right)}{x\mathcal{A}_a(x,k_t,\mu)} \tag{8}$$

This equation can be integrated between μ_{i-1}^2 and μ^2 to give the no-branching probability (Sudakov form factor) for the backward evolution Δ_{bw},[1]

$$\begin{aligned}\log \Delta_{bw}(x,k_t,\mu,\mu_{i-1}) &= \log\left(\frac{\Delta_a(\mu)}{\Delta_a(\mu_{i-1})}\frac{x\mathcal{A}_a(x,k_t,\mu_{i-1})}{x\mathcal{A}_a(x,k_t,\mu)}\right) \\ &= -\sum_b \int_{\mu_{i-1}^2}^{\mu^2} \frac{dq'^2}{q'^2}\frac{d\phi}{2\pi}\int_x^{z_M} dz\, P_{ab}^{(R)} \frac{x'\mathcal{A}_b\left(x',k_t',q'\right)}{x\mathcal{A}_a(x,k_t,q')},\end{aligned} \tag{9}$$

with $x' = x/z$. This Sudakov form factor is very similar to the Sudakov form factor in ordinary parton shower approaches, with the difference that for the PB TMD shower the ratio of PB TMD densities $[x'\mathcal{A}_b\left(x',k_t',q'\right)]/[x\mathcal{A}_a(x,k_t,q')]$ is applied, which includes a dependence on k_t.

In Eq. (9) a relation between the Sudakov form factor Δ_a used in the evolution equation and the Sudakov form factor Δ_{bw} used for the backward evolution of the parton shower is made explicit. A similar relation was also studied in Refs. [78,79]. In Ref. [78] the z_M limit was identified as a source of systematic uncertainty when using conventional showers with standard collinear pdfs; in the PB approach, the same z_M limit is present in the parton evolution as well as in the PB-shower. The PB approach allows a consistent formulation of the parton shower with the PB TMDs, as in both Sudakov form factors Δ_a and Δ_{bw} the same value of z_M is used.

The splitting functions $P_{ab}^{(R)}$ contain the coupling,

$$P_{ab}(\alpha_s,z) = \sum_{n=1}^{\infty}\left(\frac{\alpha_s(f(z,q))}{2\pi}\right)^n P_{ab}^{(n-1)}(z) , \tag{10}$$

[1] In Eq. (9) ordering in μ is assumed. However, if angular ordering as in CCFM [15–18] is applied then the ratio of parton densities would change to $[x'\mathcal{A}_b(x',k_t',q'/z)]/[x\mathcal{A}_a(x,k_t,q')]$ as discussed in [60].

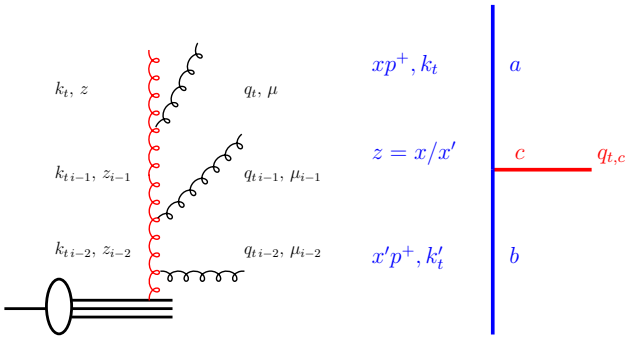


Fig. 3 Left: schematic view of a parton branching process. Right: branching process $b \to a + c$

where the scale $f(z, q)$ in the coupling depends on the ordering condition as discussed later (see Eq. (11)).

The advantage of using a PB TMD shower is that as long as the parameters of the parton shower are set through TMD distributions the parton shower uncertainties can be recast as uncertainties of the TMDs, which in turn can be fitted to experimental data in a systematic global manner.

3.2 Backward evolution for initial state TMD parton shower

A backward evolution method, as now common in Monte Carlo event generators, is applied for the initial state parton shower, evolving from the large scale of the matrix-element process backwards down to the scale of the incoming hadron. However, in contrast to the conventional parton shower, which generates transverse momenta of the initial state partons during the backward evolution, the transverse momenta of the initial partons of the hard scattering process is fixed by the TMD and the parton shower does not change the kinematics. The transverse momenta during the backward cascade follow the behavior of the TMD. The hard scattering process is obtained as described in Sect. 2. The backward evolution of the initial state parton shower follows very closely the description in [60,80,81], which is based on Ref. [53].

The starting value of the evolution scale μ is calculated from the hard scattering process, as described in Sect. 2. In case of on-shell matrix elements at NLO, the transverse momentum of the hardest parton in the parton shower evolution is limited by the *shower-scale*, as described in Sect. 2.1.

Starting at the hard scale $\mu = \mu_i$, the parton shower algorithm searches for the next scale μ_{i-1} at which a resolvable branching occurs (see Fig. 3 left). This scale μ_{i-1} is selected from the Sudakov form factor Δ_{bw} as given in Eq. (9) (see also [60]). In the parton shower language, the selection of the next branching comes from solving $R = \Delta_{bw}(x, k_t, \mu_i, \mu_{i-1})$ for μ_{i-1} using uniformly distributed random numbers R for given x and μ_i. However, to solve the integrals in Eq. (9) numerically for every branching would be too time consuming, instead the veto-algorithm [53,82] is applied.

The splitting function P_{ab} as well as the argument $f(z, q)$ in the calculation of α_s is chosen exactly as used in the evolution of the parton density. In a parton shower one treats "resolvable" branchings, defined via a cut in $z < z_M$ in the splitting function to avoid the singular behavior of the terms $1/(1-z)$, and branchings with $z > z_M$ are regarded as "non-resolvable" and are treated similarly as virtual corrections: they are included in the Sudakov form factor Δ_{bw}. The splitting variable z_{i-1} is obtained from the splitting functions following the standard methods (see Eq.(2.37) in [19]).

The calculation of the transverse momentum k_t is sketched in Fig. 3 (right). The transverse momentum $q_{t\,c}$ can be calculated in case of angular ordering (where the scale q of each branching is associated with the angle of the emission) in terms of the angle Θ of the emitted parton with respect to the beam directions $q_{t,c} = (1 - z)E_b \sin\Theta$,

$$\mathbf{q}_c^2 = (1 - z)^2 q^2 \quad . \tag{11}$$

Once the transverse momentum of the emitted parton $\mathbf{q}_c$ is known, the transverse momentum of the propagating parton can be calculated from

$$\mathbf{k}' = \mathbf{k} + \mathbf{q}_c \tag{12}$$

with a uniformly distributed azimuthal angle ϕ assumed for the vector components of $\mathbf{k}$ and $\mathbf{q}_c$. The generation of the parton momenta is performed in the center-of-mass frame of the collision (in contrast to conventional parton showers, which are generated in different partonic frames).

The whole procedure is iterated until one reaches a scale $\mu_{i-1} < q_0$ with q_0 being a cut-off parameter, which can be chosen to be the starting evolution scale of the TMD. It is of advantage to continue the parton shower evolution to lower scales $q_0 \sim \Lambda_{qcd} \sim 0.3$ GeV.

The final transverse momentum of the propagating parton $\mathbf{k}$ is the sum of all transverse momenta $\mathbf{q}_c$ (see Fig. 3 right):

$$\mathbf{k} = \mathbf{k}_0 - \sum_c \mathbf{q}_c \quad . \tag{13}$$

with $\mathbf{k}_0$ being the intrinsic transverse momentum.

The PB TMD parton shower is selected with `Parton Evolution=2` (or `ICCF=2`).

3.3 CCFM parton evolution and parton shower

The CCFM parton evolution and corresponding parton shower follows a similar approach as described in the previous section and in detail also in Refs. [60,80,81,83]. The main difference to the PB-TMD shower are the splitting functions with the non-Sudakov form factor Δ_{ns} and the allowed

phase space for emission. The original CCFM splitting function $\tilde{P}_g(z, q, k_t)$ for branching $g \to gg$ is given by[2]

$$\tilde{P}_g(z, q, k_t) = \frac{\bar{\alpha}_s(q(1-z))}{1-z} + \frac{\bar{\alpha}_s(k_t)}{z}\Delta_{ns}(z, q, k_t), \quad (14)$$

where the non-Sudakov form factor Δ_{ns} is defined as

$$\log \Delta_{ns} = -\bar{\alpha}_s(k_t) \int_0^1 \frac{dz'}{z'} \times \int \frac{dq^2}{q^2}\Theta(k_t - q)\Theta(q - z'q_t)\,, \quad (15)$$

with $q_t = \sqrt{\mathbf{q}_t^2}$ being the magnitude of the transverse vector defined in Eq. (11) and k_t the magnitude of the transverse vector in Eq. (12).

The CCFM parton shower is selected with `ICCF=1` (`PartonEvolution=1`).[3]

4 The TMD parton densities

In the previous versions of CASCADE the TMD densities were part of the program. With the development of TMDLIB [84,85] there is easy access to all available TMDs, including parton densities for photons (as well as Z, W and H densities, if available).

These parton densities can be selected via `Parton Density` with a value $>$ 100,000. For example the TMDs from the parton branching method [19,20] are selected via `PartonDensity=102100 (102200)` for PB-NLO-HERAI+II-2018-set1 (set2).

Note that the features of the TMD parton shower are only fully available for the PB-TMD sets and the CCFM shower clearly needs CCFM parton densities (like for instance [86]). PB-TMD parton densities are determined in Ref. [87] from fits to HERA DIS F_2 measurements for $Q^2 > 3\,\mathrm{GeV}^2$, giving very good χ^2 values. In Refs. [88,89] the transverse momentum distribution of Drell-Yan pairs at low and high masses, obtained from PB-TMD densities, are compared with experimental measurements in a wide variety of kinematic regions, from low-energy fixed target experiments to high-energy collider experiments. Good agreement is found between predictions and measurements without the need for tuning of non-perturbative parameters, which illustrates the validity of the approach over a broad kinematic range in energy and mass scales.

5 Final state parton showers

The final state parton shower uses the parton shower routine `PYSHOW` of PYTHIA. Leptons in the final state, coming for example from Drell-Yan decays, can radiate photons, which are also treated in the final state parton shower. Here the method from `PYADSH` of PYTHIA is applied, with the scale for the QED shower being fixed at the virtuality of the decaying particle (for example the mass of the Z-boson).

The default scale for the QCD final state shower is $\mu^2 = 2 \cdot (m_{1\perp}^2 + m_{2\perp}^2)$ (`ScaleTimeShower=1`), with $m_{1(2)\perp}$ being the transverse mass of the hard parton 1(2). Other choices are possible: $\mu^2 = \hat{s}$ (`ScaleTimeShower=2`) and $\mu^2 = 2\cdot(m_1^2+m_2^2)$ (`ScaleTimeShower=3`). In addition a scale factor can be applied: `ScaleFactorFinalShower`$\times\mu^2$ (default: `ScaleFactorFinalShower=1`).

6 Hadronization

The hadronization (fragmentation of the partons in colorless systems) is done exclusively by PYTHIA. Hadronization (fragmentation) is switched off by `Hadronization = 0` (or `NFRA = 0` for the older steering cards). All parameters of the hadronization model can be changed via the steering cards.

7 Uncertainties

Uncertainties of QCD calculations mainly arise from missing higher order corrections, which are estimated by varying the factorization and renormalization scales up and down by typically a factor of 2. The scale variations are performed when calculating the matrix elements and are stored as additional weights in the LHE file, which are then passed directly via CASCADE3 to the HEPMC [90] output file for further processing.

The uncertainties coming from the PDFs can also be calculated as additional weight factors during the matrix element calculation. However, when using TMDs, additional uncertainties arise from the transverse momentum distribution of the TMD. The PB-TMDs come with uncertainties from the experimental uncertainties as well as from model uncertainties, as discussed in Ref. [87]. These uncertainties can be treated and applied as additional weight factors with the parameter `Uncertainty_TMD=1`.

[2] Finite terms are neglected as they are not obtained in CCFM at the leading infrared accuracy (cf p. 72 in [17]).

[3] A one loop parton shower (DGLAP like) with $\Delta_{ns} = 1$, one loop α_s and strict ordering in q can be selected with `ICCF=0`.

8 Multi-jet merging

Showered multijet LO matrix element calculations can be merged using the prescription discussed in Ref. [91]. The merging performance is controlled by the three parameters `Rclus`, `Etclus`, `Etaclmax`. Final-state partons with pseudorapidity η <`Etaclmax` present in the event record after the shower step but before hadronization are passed to the merging machinery if `Imerge = 1`. Partons are clustered using the kt-jet algorithm with a cone radius `Rclus` and matched to the PB evolved matrix element partons if the distance between the parton and the jet is $R < 1.5 \times$`Rclus`. The hardness of the reconstructed jets is controlled by its minimum transverse energy `Etclus` (merging scale).

The number of light flavor partons is defined by the `NqmaxMerge` parameter. Heavy flavor partons and their corresponding radiation are not passed to the merging algorithm. All jet multiplicities are treated in exclusive mode except for the highest multiplicity `MaxJetsMerge` which is treated in inclusive mode.

9 Program description

In CASCADE3 all variables are declared as `Double Precision`. With CASCADE3 the source of PYTHIA 6.428 is included to avoid difficulties in linking.

9.1 Random numbers

CASCADE3 uses the `RANLUX` random number generator, with luxory level `LUX = 4`. The random number seed can be set via the environment variable `CASEED`, the default value is `CASEED=12345`.

9.2 Event output

When `HEPMC` is included, generated events are written out in HEPMC [90] format for further processing. The environment variable `+HEPMCOUT+` is used to specify the file name, by default this variable is set to `HEPMCOUT=/dev/null`.

The HEPMC events can be further processed, for example with Rivet [92].

9.3 Input parameters

The input parameters are steered via steering files. The new format of steering is discussed in Sect. 9.3.1 and should be used when reading LHE files, while the other format, which is appropriate for the internal off-shell processes, is discussed in Sect. 9.3.2.

9.3.1 Input parameters: new format

Examples for steering files are under `$install_path/share/cascade/LHE`.

```
&CASCADE_input
NrEvents = -1                   ! Nr of events to process
Process_Id = -1                 ! Read LHE file
Hadronisation = 0               ! Hadronisation (on =1, off = 0)
SpaceShower = 1                 ! Space-like Parton Shower
SpaceShowerOrderAlphas=2        ! Order alphas in Space Shower
TimeShower = 1                  ! Time-like Parton Shower
ScaleTimeShower = 4             ! Scale choice for Time-like Shower
!                                  1: 2(m^2_1t+m^2_2t)
!                                  2: shat
!                                  3: 2(m^2_1+m^2_2)
!                                  4: 2*scalup (from lhe file)
!ScaleFactorFinalShower = 1.    ! scale factor for Final State Parton Shower
PartonEvolution = 2             ! type of parton evolution in Space-like Shower
!                                  1: CCFM
!                                  2: full all flavor TMD evolution
! EnergyShareRemnant = 4        ! energy sharing in proton remnant
!                                  1: (a+1)(1-z)**a, <z>=1/(a+2)=1/3
!                                  2: (a+1)(1-z)**a, <z>=1/(a+2)=mq/(mq+mQ
!                                  3: N/(z(1-1/z-c/(1-z))**2), c=(mq/mQ)**2
!                                  4: PYZDIS: KFL1=1
! Remnant = 0                   ! =0 no remnant treatment
PartonDensity = 102200            ! use TMDlib: PB-TMDNLO-set2
! PartonDensity = 102100          ! use TMDlib: PB-TMDNLO-set1
! TMDDensityPath= './share'     ! Path to TMD density for internal files
Uncertainty_TMD = 0                        ! calculate and store uncertainty TMD pdfs
lheInput='MCatNLO-example.lhe' ! LHE input file
lheHasOnShellPartons = 1        ! = 0 LHE file has off-shell parton configuration
lheReweightTMD = 0              ! Reweight with new TMD  given in PartonDensity
lheScale = 2                    ! Scale defintion for TMD

!                                  0: use scalup
!                                  1: use shat
!                                  2: use 1/2 Sum pt^2 of final parton/particles
!                                  3: use shat for Born and 1/2 Sum pt^2 of final parton(particle)
!                                  4: use shat for Born and max pt of most forward/backward
!                                     parton(particle)
lheNBornpart = 2                ! Nr of hard partons (particles) (Born process)
ScaleFactorMatchingScale = 2.   ! Scale factor for matching scale when including TMDs
&End

&PYTHIA6_input
P6_Itune = 370                  ! Retune of Perugia 2011 w CTEQ6L1 (Oct 2012)
! P6_MSTJ(41) = 1               ! (D = 2) type of branchings allowed in shower.
!                                  1: only QCD
!                                  2: QCD and photons off quarks and leptons
P6_MSTJ(45) = 4                 ! Nr of flavors in final state shower: g->qqbar
P6_PMAS(4,1)= 1.6               ! charm mass
P6_PMAS(5,1)= 4.75              ! bottom mass
P6_MSTJ(48) =   1               ! (D=0), 0=no max. angle, 1=max angle def. in PARJ(85)
! P6_MSTU(111) =  1             ! = 0 : alpha_s is fixed, =1 first order; =2 2nd order;
! P6_PARU(112) = 0.2            ! lambda QCD
P6_MSTU(112)=   4               ! nr of flavours wrt lambda_QCD
P6_MSTU(113)=                   ! min nr of flavours for alphas
P6_MSTU(114)=   5               ! max nr of flavours for alphas
&End
```

9.3.2 Input parameters: off-shell processes

Examples for steering files are under `$install_path/share/cascade/HERA` and `$install_path/share/cascade/PP`.

```
* OLD STEERING FOR CASCADE
*
* number of events to be generated
*
```

```
NEVENT 100
*
* +++++++++++++++++ Kinematic parameters +++++++++++++++
*
'PBE1'    1    0    -7000.    ! Beam energy
'KBE1'    1    0     2212     ! -11: positron, 22: photon  2212: proton
'IRE1'    1    0       1      ! 0: beam 1 has no structure
*                             ! 1: beam 1 has structure
'PBE2'     1   0     7000.    ! Beam energy
'KBE2'     1   0     2212     ! 11: electron, 22: photon 2212: proton
'IRE2'     1   0        1     ! 0:  beam 3 has no structure
*                             ! 1:  beam 2 has structure
'NFLA'     1   0        4     ! (D=5) nr of flavours used in str.fct
* +++++++++++++++ Hard subprocess selection ++++++++++++++++++
'IPRO'     1   0        2     ! (D=1)
*                             !  2: J/psi g
*                             !  3: chi_c
'I23S'      1    0       0    ! (D=0) select 2S or 3S state
'IPOL'      1    0       0    ! (D=0)  VM->ll (polarization study)
'IHFL'      1    0       4    ! (D=4) produced flavour for IPRO=11
*                             ! 4: charm
*                             ! 5: bottom
'PTCU'      1    0       1.   ! (D=0) p_t **2 cut for process
* ++++++++++++ Parton shower and fragmentation ++++++++++++
'NFRA'     1     0       1    ! (D=1) Fragmentation on=1 off=0
'IFPS'     1     0       3    ! (D=3) Parton shower
*                             ! 0: off
*                             ! 1: initial state PS
*                             ! 2: final state PS
*                             ! 3: initial and final state PS
'IFIN'     1     0       1    ! (D=1) scale switch for FPS
*                             ! 1: 2(m^2_1t+m^2_2t)
*                             ! 2: shat
*                             ! 3: 2(m^2_1+m^2_2)
'SCAF'     1     0       1.   ! (D=1) scale factor for FPS
'ITIM'     1     0       0    ! 0: timelike partons may not shower
*                             ! 1: timelike partons may shower
'ICCF'     1     0       1    ! (D=1) Evolution equation
*                             ! 0: DGLAP
*                             ! 1: CCFM
*                             ! 2: PB TMD evolution
* ++++++++++++ Structure functions and scales ++++++++++++
'IRAM'     1     0       0    ! (D=0) Running of alpha_em(Q2)
*                             ! 0: fixed
*                             ! 1: running
'IRAS'     1     0       1    ! (D=1) Running of alpha_s(MU2)
*                             ! 0: fixed alpha_s=0.3
*                             ! 1: running
'IQ2S'     1     0       3    ! (D=1) Scale MU2 of alpha_s
*                             !  1: MU2= 4*m**2 (only for heavy quarks)
*                             !  2: MU2 = shat(only for heavy quarks)
*                             !  3: MU2= 4*m**2 + pt**2
*                             !  4: MU2 = Q2
*                             !  5: MU2 = Q2 + pt**2
*                             !  6: MU2 = k_t**2
'SCAL'     1    0     1.0     !  scale factor for renormalisation scale
'SCAF'     1    0     1.0     !  scale factor for factorisation scale*
*'IGLU'    1    0     1201    ! (D=1010)Unintegrated gluon density
*                             ! > 10000 use TMDlib (i.e. 101201 for JH-2013-set1)
*                             !  1201: CCFM set JH-2013-set1 (1201 - 1213)
*                             !  1301: CCFM set JH-2013-set2 (1301 - 1313)
*                             !  1001: CCFM J2003 set 1
*                             !  1002: CCFM J2003 set 2
*                             !  1003: CCFM J2003 set 3
*                             !  1010: CCFM set A0
*                             !  1011: CCFM set A0+
```

```
*                              !  1012: CCFM set A0-
*                              !  1013: CCFM set A1
*                              !  1020: CCFM set B0
*                              !  1021: CCFM set B0+
*                              !  1022: CCFM set B0-
*                              !  1023: CCFM set B1
*                              !  1: CCFM old set JS2001
*                              !  2: derivative of collinear gluon (GRV)
*                              !  3: Bluemlein
*                              !  4: KMS
*                              !  5: GBW (saturation model)
*                              !  6: KMR
*                              !  7: Ryskin,Shabelski
* ++++++++++++ BASES/SPRING Integration procedure ++++++++++++
'NCAL'     1    0    50000    ! (D=20000) Nr of calls per iteration for bases
'ACC1'     1    0     1.0     ! (D=1) relative prec.(
'ACC2'     1    0     0.5     ! (0.5) relative prec.(
* ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
*'INTE'    1    0      0      ! Interaction type (D=0)
*                              ! = 0 electromagnetic interaction
*'KT1 '    1    0      0.44   ! (D=0.0) intrinsic kt for beam 1
*'KT2 '    1    0      0.44   ! (D=0.0) intrinsic kt for beam 2
*'KTRE'    1    0      0.35   ! (D=0.35) primordial kt when non-trivial
*                              ! target remnant is split into two particles
* Les Houches Accord Interface
'ILHA'     1    0      0      ! (D=10) Les Houches Accord
*                              ! = 0  use internal CASCADE
*                              ! = 1  write event file
*                              ! = 10 call PYTHIA for final state PS and remnant frag
* path for updf files
* 'UPDF'   './share'
```

10 Program installation

CASCADE3 now follows the standard AUTOMAKE convention. To install the program, do the following

```
1) Get the source from http://www.desy.de/~jung/cascade

tar xvfz cascade-XXXX.tar.gz
cd cascade-XXXX

2) Generate the Makefiles (do not use shared libraries)
./configure --disable-shared --prefix=install-path --with-lhapdf="lhapdflib_path"
--with-tmdlib="TMDlib_path"  --with-hepmc="hepmc_path"

with (as example):
lhapdflib_path=/Users/jung/MCgenerators/lhapdf/6.2.1/local
TMDlib_path=/Users/jung/jung/cvs/TMDlib/TMDlib2/local
hepmc_path/Users/jung/MCgenerators/hepmc/HepMC-2.06.09/local
3) Compile the binary
make

4) Install the executable and PDF files
make install

4) The executable is in bin
run it with:
export CASEED=1242425
export HEPMCOUT=outfile.hepmc

cd $install-path/bin

./cascade < $install-path/share/cascade/LHE/steering-DY-MCatNLO.txt
```

Acknowledgements FG acknowledges the support and hospitality of DESY, Hamburg, where part of this work started. FH acknowledges the hospitality and support of DESY, Hamburg and of CERN, Theory Division while parts of this work were being done. SB, ALi and MM are grateful the DESY Directorate for the support in the framework of Cooperation Agreement between MSU and DESY on phenomenology of the LHC processes and TMD parton densities. MM was supported by a grant of the foundation for the advancement of theoretical physics and mathematics "Basis" 20-1-3-11-1. STM thanks the Humboldt Foundation for the Georg Forster research fellowship and gratefully acknowledges support from IPM. ALe acknowledges funding by Research Foundation-Flanders (FWO) (application number: 1272421N). QW and HY acknowledge the support by the Ministry of Science and Technology under Grant no. 2018YFA040390 and by the National Natural Science Foundation of China under Grant no. 11661141008.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The source code of the program can be downloaded from https://www.desy.de/\protect\unhbox\voidb@x\penalty\@M\jung/cascade/.]

Funded by SCOAP[3].

Appendix A: Program summary

Title of Program: CASCADE3 3.1.0
Computer for which the program is designed and others on which it is operable: any with standard Fortran 77 (gfortran)
Programming Language used: FORTRAN 77
High-speed storage required: No
Separate documentation available: No
Keywords: QCD, TMD parton distributions.
Method of solution Since measurements involve complex cuts and multi-particle final states, the ideal tool for any theoretical description of the data is a Monte Carlo event generator which generates initial state parton showers according to Transverse Momentum Dependent (TMD) parton densities, in a backward evolution, which follows the evolution equation as used for the determination of the TMD.
Restrictions on the complexity of the problem Any LHE file (with on-shell or off-shell) initial state partons can be processed.
Other Program used PYTHIA (*version* > 6.4) for final state parton shower and hadronization, BASES/SPRING 5.1 for integration (both supplied with the program package), TMDLIB as a library for TMD parton densities.
Download of the program https://www.desy.de/~jung/cascade/
Unusual features of the program None

References

1. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni et al., JHEP **1407**, 079 (2014). arXiv:1405.0301
2. S. Frixione, B.R. Webber (2006) arXiv:hep-ph/0612272
3. S. Frixione, P. Nason, B.R. Webber, JHEP **08**, 007 (2003). arXiv:hep-ph/0305252
4. S. Frixione, B.R. Webber (2002) arXiv:hep-ph/0207182
5. S. Frixione, B.R. Webber, JHEP **06**, 029 (2002). arXiv:hep-ph/0204244
6. S. Alioli, K. Hamilton, P. Nason, C. Oleari, E. Re, JHEP **04**, 081 (2011). arXiv:1012.3380
7. S. Frixione, P. Nason, C. Oleari, JHEP **0711**, 070 (2007). arXiv:0709.2092
8. M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton et al., Eur. Phys. J. C **58**, 639 (2008). arXiv:0803.0883
9. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, Comput. Phys. Commun. **191**, 159 (2015). arXiv:1410.3012
10. J. Alwall, S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad et al., Eur. Phys. J. C **53**, 473 (2008). arXiv:0706.2569
11. S. Catani, F. Krauss, R. Kuhn, B.R. Webber, JHEP **11**, 063 (2001). arXiv:hep-ph/0109231
12. R. Frederix, S. Frixione, JHEP **12**, 061 (2012). arXiv:1209.6215
13. K. Hamilton, P. Nason, G. Zanderighi, JHEP **10**, 155 (2012). arXiv:1206.3572
14. S. Catani, M. Ciafaloni, F. Hautmann, Phys. Lett. B **242**, 97 (1990)
15. M. Ciafaloni, Nucl. Phys. B **296**, 49 (1988)
16. S. Catani, F. Fiorani, G. Marchesini, Phys. Lett. B **234**, 339 (1990)
17. S. Catani, F. Fiorani, G. Marchesini, Nucl. Phys. B **336**, 18 (1990)
18. G. Marchesini, Nucl. Phys. B **445**, 49 (1995). arXiv:hep-ph/9412327
19. F. Hautmann, H. Jung, A. Lelek, V. Radescu, R. Zlebcik, JHEP **01**, 070 (2018). arXiv:1708.03279
20. F. Hautmann, H. Jung, A. Lelek, V. Radescu, R. Zlebcik, Phys. Lett. B **772**, 446 (2017). arXiv:1704.01757
21. R. Angeles-Martinez et al., Acta Phys. Pol. B **46**, 2501 (2015). arXiv:1507.05267
22. J. Alwall et al., Comput. Phys. Commun. **176**, 300 (2007). arXiv:hep-ph/0609017
23. J.C. Collins, D.E. Soper, Nucl. Phys. B **193**, 381 (1981). (**Erratum: Nucl. Phys. B 213, 545 (1983)**)
24. J.C. Collins, D.E. Soper, Nucl. Phys. B **194**, 445 (1982)
25. J.C. Collins, D.E. Soper, G.F. Sterman, Nucl. Phys. B **223**, 381 (1983)
26. J.C. Collins, D.E. Soper, G.F. Sterman, Phys. Lett. B **109**, 388 (1982)
27. J.C. Collins, D.E. Soper, G.F. Sterman, Nucl. Phys. B **250**, 199 (1985)
28. J. Collins, *Foundations of Perturbative QCD. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*, vol. 32 (2011)
29. R. Meng, F.I. Olness, D.E. Soper, Phys. Rev. D **54**, 1919 (1996). arXiv:hep-ph/9511311
30. P.M. Nadolsky, D.R. Stump, C.P. Yuan, Phys. Rev. D **61**, 014003 (2000). (**Erratum: Phys. Rev. D 64, 059903 (2001)**). arXiv:hep-ph/9906280
31. P.M. Nadolsky, D.R. Stump, C.P. Yuan, Phys. Rev. D **64**, 114011 (2001). arXiv:hep-ph/0012261

32. X.-D. Ji, J.-P. Ma, F. Yuan, Phys. Rev. D **71**, 034005 (2005). arXiv:hep-ph/0404183
33. X.-D. Ji, J.-P. Ma, F. Yuan, Phys. Lett. B **597**, 299 (2004). arXiv:hep-ph/0405085
34. M.G. Echevarria, A. Idilbi, I. Scimemi, JHEP **07**, 002 (2012). arXiv:1111.4996
35. J.-Y. Chiu, A. Jain, D. Neill, I.Z. Rothstein, Phys. Rev. Lett. **108**, 151601 (2012). arXiv:1104.0881
36. E.M. Levin, M.G. Ryskin, Y.M. Shabelski, A.G. Shuvaev, Sov. J. Nucl. Phys. **53**, 657 (1991)
37. J.C. Collins, R.K. Ellis, Nucl. Phys. B **360**, 3 (1991)
38. F. Hautmann, Phys. Lett. B **535**, 159 (2002). arXiv:hep-ph/0203140
39. E. Avsar (2012) arXiv:1203.1916
40. E. Avsar, Int. J. Mod. Phys. Conf. Ser. **04**, 74 (2011). arXiv:1108.1181
41. S. Jadach, M. Skrzypek, Acta Phys. Pol. B **40**, 2071 (2009). arXiv:0905.1399
42. F. Dominguez, *Unintegrated Gluon Distributions at Small-x*. Ph.D. Thesis, Columbia U (2011)
43. F. Dominguez, J.-W. Qiu, B.-W. Xiao, F. Yuan, Phys. Rev. D **85**, 045003 (2012). arXiv:1109.6293
44. F. Dominguez, A. Mueller, S. Munier, B.-W. Xiao, Phys. Lett. B **705**, 106 (2011). arXiv:1108.1752
45. F. Hautmann, Acta Phys. Pol. B **40**, 2139 (2009)
46. F. Hautmann, M. Hentschinski, H. Jung (2012) arXiv:1205.6358
47. F. Hautmann, H. Jung, Nucl. Phys. Proc. Suppl. **184**, 64 (2008). arXiv:0712.0568
48. S. Catani, M. Ciafaloni, F. Hautmann, Phys. Lett. B **307**, 147 (1993)
49. S. Catani, F. Hautmann, Nucl. Phys. B **427**, 475 (1994). arXiv:hep-ph/9405388
50. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, G. Watt, Eur. Phys. J. C **75**, 132 (2015). arXiv:1412.7420
51. S. Dooling, P. Gunnellini, F. Hautmann, H. Jung, Phys. Rev. D **87**, 094009 (2013). arXiv:1212.6164
52. F. Hautmann, H. Jung, Eur. Phys. J. C **72**, 2254 (2012). arXiv:1209.6549
53. M. Bengtsson, T. Sjostrand, M. van Zijl, Z. Phys. C **32**, 67 (1986)
54. S. Catani, M. Ciafaloni, F. Hautmann, Nucl. Phys. B Proc. Suppl. **29**, 182 (1992)
55. G. Marchesini, B.R. Webber, Nucl. Phys. B **386**, 215 (1992)
56. F. Hautmann, H. Jung, S.T. Monfared, Eur. Phys. J. C **74**, 3082 (2014). arXiv:1407.5935
57. E.A. Kuraev, L.N. Lipatov, V.S. Fadin, Sov. Phys. JETP **44**, 443 (1976)
58. E.A. Kuraev, L.N. Lipatov, V.S. Fadin, Sov. Phys. JETP **45**, 199 (1977)
59. I.I. Balitsky, L.N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978)
60. H. Jung, S. Baranov, M. Deak, A. Grebenyuk, F. Hautmann et al., Eur. Phys. J. C **70**, 1237 (2010). arXiv:1008.0152
61. A. van Hameren, Comput. Phys. Commun. **224**, 371 (2018). arXiv:1611.00680
62. A. Lipatov, M. Malyshev, S. Baranov, Eur. Phys. J. C **80**, 330 (2020). arXiv:1912.04204
63. S. Catani, M. Ciafaloni, F. Hautmann, Nucl. Phys. B **366**, 135 (1991)
64. V. Saleev, N. Zotov, Mod. Phys. Lett. A **9**, 151 (1994). (**Erratum: Mod. Phys. Lett. A 9, 1517–1518 (1994)**)
65. A. Lipatov, N. Zotov, Eur. Phys. J. C **27**, 87 (2003). arXiv:hep-ph/0210310
66. S. Baranov, N. Zotov, J. Phys. G **29**, 1395 (2003). arXiv:hep-ph/0302022
67. S.P. Baranov, Phys. Rev. D **66**, 114003 (2002)
68. S. Baranov, Phys. Rev. D **84**, 054012 (2011)
69. S.P. Baranov, A.V. Lipatov, N.P. Zotov, Phys. Rev. D **78**, 014025 (2008). arXiv:0805.4821
70. M. Deak, F. Schwennsen, JHEP **09**, 035 (2008). arXiv:0805.3763
71. S. Marzani, R.D. Ball, Nucl. Phys. B **814**, 246 (2009). arXiv:0812.3602
72. M. Deak, F. Hautmann, H. Jung, K. Kutak, JHEP **09**, 121 (2009). arXiv:0908.0538
73. W. Furmanski, R. Petronzio, Z. Phys. C **11**, 293 (1982)
74. G. Curci, W. Furmanski, R. Petronzio, Nucl. Phys. B **175**, 27 (1980)
75. Y.L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977). (**Zh. Eksp. Teor. Fiz. 73, 1216 (1977)**)
76. G. Altarelli, G. Parisi, Nucl. Phys. B **126**, 298 (1977)
77. V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972). (**Yad. Fiz. 15, 781 (1972)**)
78. Z. Nagy, D.E. Soper, Phys. Rev. D **102**, 014025 (2020). arXiv:2002.04125
79. L. Gellersen, D. Napoletano, S. Prestel, Monte Carlo studies, in *11th Les Houches Workshop on Physics at TeV Colliders: Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*, p. 131 (2020). Also in preprint 2003.01700
80. H. Jung, Comput. Phys. Commun. **143**, 100 (2002). arXiv:hep-ph/0109102
81. H. Jung, G.P. Salam, Eur. Phys. J. C **19**, 351 (2001). arXiv:hep-ph/0012143
82. S. Platzer, M. Sjodahl, Eur. Phys. J. Plus **127**, 26 (2012). arXiv:1108.6180
83. F. Hautmann, H. Jung, JHEP **10**, 113 (2008). arXiv:0805.1049
84. F. Hautmann, H. Jung, M. Krämer, P. Mulders, E. Nocera et al., Eur. Phys. J. C **74**, 3220 (2014). arXiv:1408.3015
85. N.A. Abdulov et al., TMDlib2 and TMDplotter: a platform for 3D hadron structure studies. arXiv:2103.09741
86. F. Hautmann, H. Jung, Nucl. Phys. B **883**, 1 (2014). arXiv:1312.7875
87. A.B. Martinez, P. Connor, F. Hautmann, H. Jung, A. Lelek, V. Radescu, R. Zlebcik, Phys. Rev. D **99**, 074008 (2019). arXiv:1804.11152
88. A.B. Martinez et al., Eur. Phys. J. C **80**, 598 (2020). arXiv:2001.06488
89. A.B. Martinez et al., Phys. Rev. D **100**, 074027 (2019). arXiv:1906.00919
90. M. Dobbs, J.B. Hansen, Comput. Phys. Commun. **134**, 41 (2001)
91. A.B. Martinez et al., *Jet merging with TMD parton branching* (To be published) (2021)
92. A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk, H. Schulz, F. Siegert, Comput. Phys. Commun. **184**, 2803 (2013). arXiv:1003.0694

Eur. Phys. J. C (2022) 82:36
https://doi.org/10.1140/epjc/s10052-022-09997-1

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method

M. I. Abdulhamid[1], M. A. Al-Mashad[2], A. Bermudez Martinez[3], G. Bonomelli[4], I. Bubanja[5,6], N. Crnković[7], F. Colombina[3], B. D'Anzi[8], S. Cerci[9], M. Davydov[10], L. I. Estevez Banos[3], N. Forzano[4], F. Hautmann[11,12], H. Jung[3,a], S. Kim[12], A. León Quirós[13], D. E. Martins[14], M. Mendizabal[3], K. Moral Figueroa[15], S. Prestel[16], S. Taheri Monfared[3], C. Süslü[17], D. Sunar Cerci[9], A. M. van Kampen[11], P. Van Mechelen[11], A. Verbytskyi[18], Q. Wang[3,19], H. Yang[3,19], Y. Zhou[20]

[1] Department of Physics, Tanta University, Tanta, Egypt
[2] Faculty of Science, Center for High Energy Physics (CHEP-FU), Fayoum University, Faiyum, Egypt
[3] Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[4] Università degli Studi Milano-Bicocca, Milan, Italy
[5] Interuniversity Institute for High Energies (IIHE), Université libre de Bruxelles, Brussels, Belgium
[6] Faculty of Natural Sciences and Mathematics, University of Montenegro, Podgorica, Montenegro
[7] Ruđer Bošković Institute, Zagreb, Croatia
[8] National Institute for Nuclear Physics INFN and University of Bari, Bari, Italy
[9] Adiyaman University, Adiyaman, Turkey
[10] Moscow State University, Moscow, Russia
[11] Elementary Particle Physics, University of Antwerp, Antwerp, Belgium
[12] University of Oxford, Oxford, UK
[13] Universidad Autónoma de Madrid, Madrid, Spain
[14] Universidade Federal de Pelotas, Pelotas, Brazil
[15] University of Edinburgh, Edinburgh, UK
[16] Department of Astronomy and Theoretical Physics, Lund University, Lund, Sweden
[17] Department of Physics, Ihsan Dogramaci Bilkent University, Ankara, Turkey
[18] Max-Planck Institut für Physik, Munich, Germany
[19] School of Physics, Peking University, Beijing, China
[20] University of Cambridge, Cambridge, UK

Received: 22 December 2021 / Accepted: 4 January 2022

Abstract The azimuthal correlation, $\Delta\phi_{12}$, of high transverse momentum jets in pp collisions at $\sqrt{s} = 13$ TeV is studied by applying PB-TMD distributions to NLO calculations via MCatNLO together with the PB-TMD parton shower. A very good description of the cross section as a function of $\Delta\phi_{12}$ is observed. In the back-to-back region of $\Delta\phi_{12} \to \pi$, a very good agreement is observed with the PB-TMD Set 2 distributions while significant deviations are obtained with the PB-TMD Set 1 distributions. Set 1 uses the evolution scale while Set 2 uses transverse momentum as an argument in α_s, and the above observation therefore confirms the importance of an appropriate soft-gluon coupling in angular ordered parton evolution. The total uncertainties of the predictions are dominated by the scale uncertainties of the matrix element, while the uncertainties coming from the PB-TMDs and the corresponding PB-TMD shower are very small. The $\Delta\phi_{12}$ measurements are also compared with predictions using MCatNLO together PYTHIA8, illustrating the importance of details of the parton shower evolution.

1 Introduction

The description of the cross section of high p_T jets in proton-proton (pp) collisions is one of the most important tests of predictions obtained in Quantum Chromodynamics (QCD), and much progress has been achieved in the description of inclusive jets [1–12] by applying next-to-leading (NLO) [13–16] and next-to-next-to-leading-order (NNLO) calculations [17–20]. In multijet production, the azimuthal angle $\Delta\phi_{12}$ between the two highest transverse momentum p_T-jets is an inclusive measurement of additional jet radiation. At leading order (LO) in strong coupling α_s, where only two jets are present, the jets are produced back-to-back, with $\Delta\phi_{12} = \pi$, while a deviation from this back-to-back configuration indi-

[a] e-mail: hannes.jung@desy.de (corresponding author)

Published online: 13 January 2022

cates the presence of additional jets, and only higher-order calculations can describe the observations. The azimuthal correlation between two jets has been measured in $p\bar{p}$ collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV by the D0 collaboration [21,22] and in pp collisions by the ATLAS Collaboration at $\sqrt{s} = 7$ TeV [23] and by the CMS Collaboration at $\sqrt{s} = 7, 8$, and 13 TeV [24–27]. When measurements of azimuthal correlations of dijets are compared with LO or NLO computations supplemented by parton showers, deviations of 50% are observed in the medium $\Delta\phi_{12}$ region even at NLO (see e.g. [25,26]), which requires a more detailed understanding. In the $\Delta\phi_{12} \to \pi$ region, deviations of up to 10 % are observed [27], significantly larger than the experimental uncertainties.

Since initial state parton radiation moves the jets away from the $\Delta\phi_{12} = \pi$ region, it is appropriate to investigate the implications of transverse momentum dependent parton densities (TMDs [28]) in the description of the $\Delta\phi_{12}$ measurements. Kinematic effects of the initial-state transverse momenta in the interpretation of jet measurements were pointed out in [29,30]. The region $\Delta\phi_{12} \to \pi$ is especially sensitive to soft multi-gluon emissions, for which QCD resummation is needed. Calculations at leading-logarithm have been obtained in Ref. [31]. A calculation based on TMD distributions is found in Refs. [32,33] and further investigated in [34,35]. However, in the region $\Delta\phi_{12} \to \pi$ soft-gluon effects are expected which lead to so-called *factorization-breaking* [36–38]. An indirect strategy to explore the potential impact of these effects is to compare calculations which assume factorization with high-precision measurements.

The Parton Branching (PB)-method [39,40] allows one to determine TMD parton distributions. With these PB-TMD distributions a very good description of the Drell–Yan process [41] is achieved at the LHC [42] as well as at lower energies [43]. Drell–Yan lepton pair production in association with jet final states is also well described using the TMD jet merging [44]. In Ref. [45] it is shown that Z + b production is also well described. TMD parton distributions have been used together with off-shell matrix elements at the lowest order in Refs. [46–48] showing a reasonably good description of the measurements.

In this article we investigate in detail high-p_T dijet production by applying the PB formulation of TMD evolution together with NLO calculations of the hard scattering process in the MadGraph5_aMC@NLO [49] framework. We first give a very brief recap of the PB distributions in Sect. 2. In Sect. 3 we describe how TMDs and TMD parton showers are included in the Monte Carlo generator Cascade3 [50]. We discuss predictions obtained by applying fixed-order NLO perturbative calculations and study the region where soft gluon resummation becomes important. We show predictions using PB-TMDs together with TMD parton shower in Sect. 4. We compare these predictions with the one using the Pythia8 parton shower. We finally give conclusions in Sect. 5.

2 PB TMDs

The PB method [39,40] provides a solution of evolution equations for collinear and TMD parton distributions. The equations are solved by applying the concept of resolvable and non-resolvable branchings with Sudakov form factors providing the probability to evolve from one scale to another without resolvable branching. The method is described in Refs. [40,51].

For the numerical calculations we use the NLO parton distribution sets, PB-NLO-2018-Set 1 and PB-NLO-2018-Set 2, as obtained in Ref. [51] from a fit to inclusive deep inelastic scattering precision measurements at HERA [52]. Both the collinear and TMD distributions are available in TMDlib[53,54], including uncertainty bands. PB-NLO-2018-Set 1 corresponds at collinear level to HERAPDF 2.0 NLO [52], while PB-NLO-2018-Set 2 uses transverse momentum (instead of the evolution scale in Set1) for the scale in the running coupling α_s which corresponds to the angular ordering of soft gluon emissions in the initial-state parton evolution [55–58].

In Fig. 1 the distributions of the collinear densities from Set 1 and Set 2 are shown for up-quarks at evolution scales of $\mu = 100$ and 1000 GeV , typical for multi-jet production described below. The collinear densities are also available in a format compatible with LHAPDF [59], and can be used in calculations of physical processes at NLO. In Fig. 2 we show the TMD distributions for up-quarks at $x = 0.01$ and $\mu = 10$ and 100 GeV . The differences between Set 1 and Set 2 are clearly visible in the small k_T-region.

The uncertainties of the TMD distributions include both experimental and model uncertainties, as determined in Ref. [51]. In general, it is observed that those uncertainties are small; for $k_T > 1$ GeV they are of the order of 2–3 %.

3 Multijet production

The predictions for multijet production at NLO are obtained using the MadGraph5_aMC@NLO [49] framework. We used MadGraph5_aMC@NLO in two different modes: one is the fixed NLO mode, in which only partonic events are produced, without parton shower and hadronization, and the other one is the real MC@NLO mode, in which infrared subtraction terms are included to avoid double counting of parton emissions between matrix-element and parton-shower calculations, so that events need to be supplemented with a parton shower (or with PB TMD evolution) in order to produce a physical cross section.

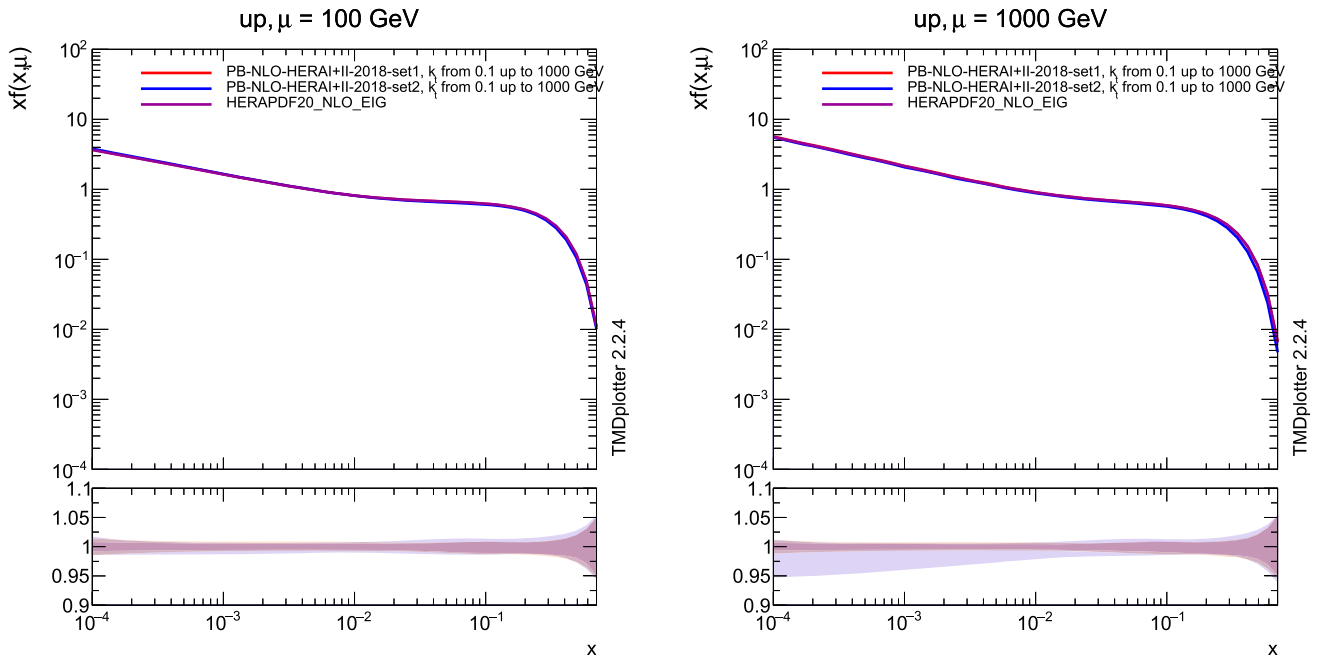


Fig. 1 Collinear parton density distributions for up quarks (PB-NLO-2018-Set 1, PB-NLO-2018-Set 2 and HERAPDF2.0) as a function of x at $\mu = 100$ and 1000 GeV. In the lower panel the uncertainties are shown

Fixed NLO dijet production is calculated within the MadGraph5_aMC@NLO framework. Technically, in the fixed NLO mode, MadGraph5_aMC@NLO (version 2.9.3) produces event files with the partonic configuration in LHE format [60] which can be processed through Cascade3 [50] combining events and counter events (due to infrared subtraction) so that they are treated as one event for a proper calculation of statistical uncertainties. In the fixed NLO mode, the MadGraph5_aMC@NLO event record is kept without any modification. Processing through Cascade3 has a significant advantage that a fixed NLO calculation can be obtained making use of all the analyses coded in Rivet [61].

In the MC@NLO mode, subtraction terms are included which depend on the parton shower used. For the PB-TMDs and the PB-TMD parton shower we use Herwig6 [62,63] subtraction terms, as already applied in Z and Drell–Yan analyses [42,43], motivated by the angular ordering in the PB evolution. MadGraph5_aMC@NLO (version 2.6.4, hereafter labeled MCatNLO) [49] together with the NLO PB parton distributions with $\alpha_s(M_Z) = 0.118$ is used for NLO calculation of dijet production. The matching scale μ_m, which limits the contribution from PB-TMDs and TMD showers (μ_m=SCALUP included in the LHE file), guarantees that the overlap with real emissions from the matrix element is minimized according to the subtraction of counterterms in the MC@NLO method. The factorization and renormalization scale in MCatNLO is set to $\mu_{R,F} = \frac{1}{2}\sum_i \sqrt{m_i^2 + p_{t,i}^2}$, where the index i runs over all particles in the matrix element final state. This scale is also used in the PB-TMD parton distribution $\mathcal{A}(x, k_T, \mu)$.

In Cascade3, as described in detail in Ref. [50], the transverse momentum of the initial state partons is calculated according to the distribution of k_T provided by the PB-TMD $\mathcal{A}(x, k_T, \mu)$ at given longitudinal-momentum fraction x and evolution scale μ. This transverse momentum is used for the initial state partons provided by MCatNLO, and their longitudinal momentum is adjusted such that the mass and the rapidity of the dijet system is conserved, similar to what has been done in the Drell–Yan case [43]. The initial state TMD parton shower is included in a backward evolution scheme, respecting all parameters and constraints from the PB-TMD. The kinematics of the hard process are not changed by the shower, after the k_T from the TMD is included. The final state parton shower is obtained with the corresponding method implemented in Pythia6 [64], by vetoing emissions which do not satisfy angular ordering (`MSTJ(42)=2`).

In Fig. 3 we show results for the transverse momentum distribution of the dijet system $p_{T,12}$ and the azimuthal correlation $\Delta\phi_{12}$ between the two leading jets as obtained from the MCatNLO calculation at fixed NLO (blue curve), at the level including subtraction terms (LHE level, green curve) and after inclusion of PB-TMDs (red curve). One can clearly observe the rising cross section of the fixed NLO calculation towards small $p_{T,12}$ (or at large $\Delta\phi_{12}$). This is the region in $p_{T,12}$ and $\Delta\phi_{12}$ where the subtraction terms are relevant and a physical prediction is obtained when PB-TMDs and parton showers are included. The jets are defined with the

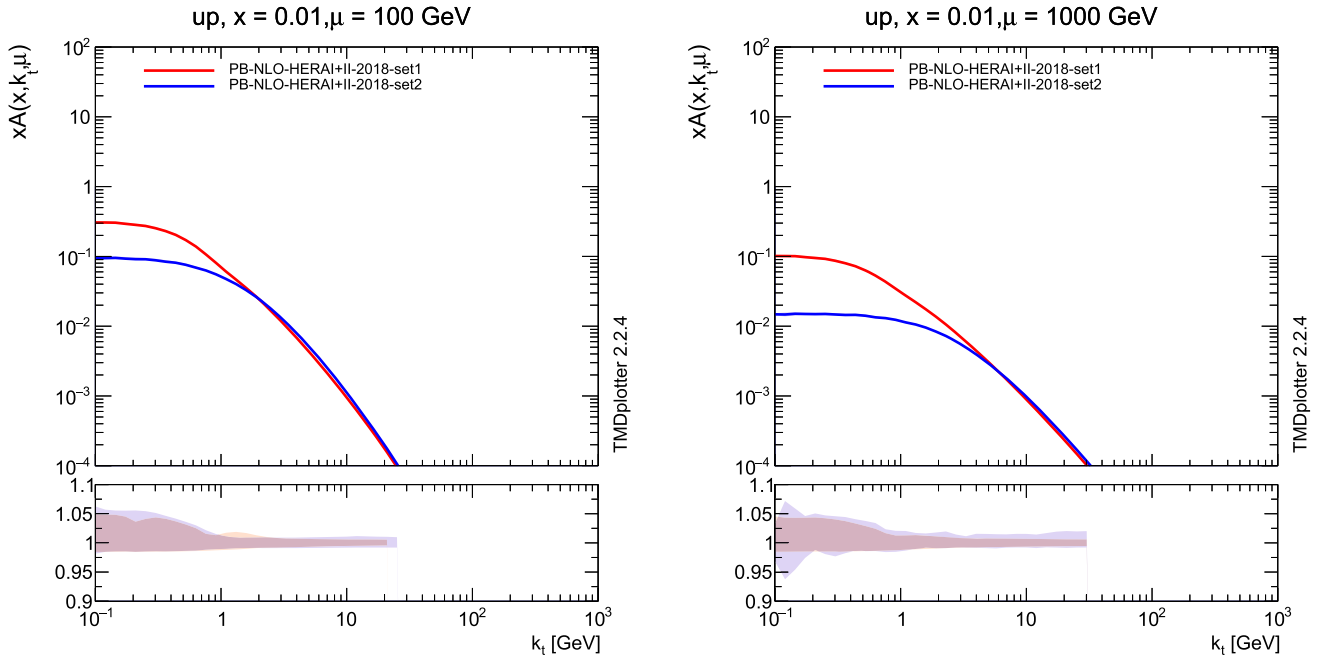


Fig. 2 TMD parton density distributions for up quarks (PB-NLO-2018-Set 1 and PB-NLO-2018-Set 2) as a function of k_T at $\mu = 100$ and 1000 GeV and $x = 0.01$. In the lower panels show the full uncertainty of the TMDs, as obtained from the fits [51]

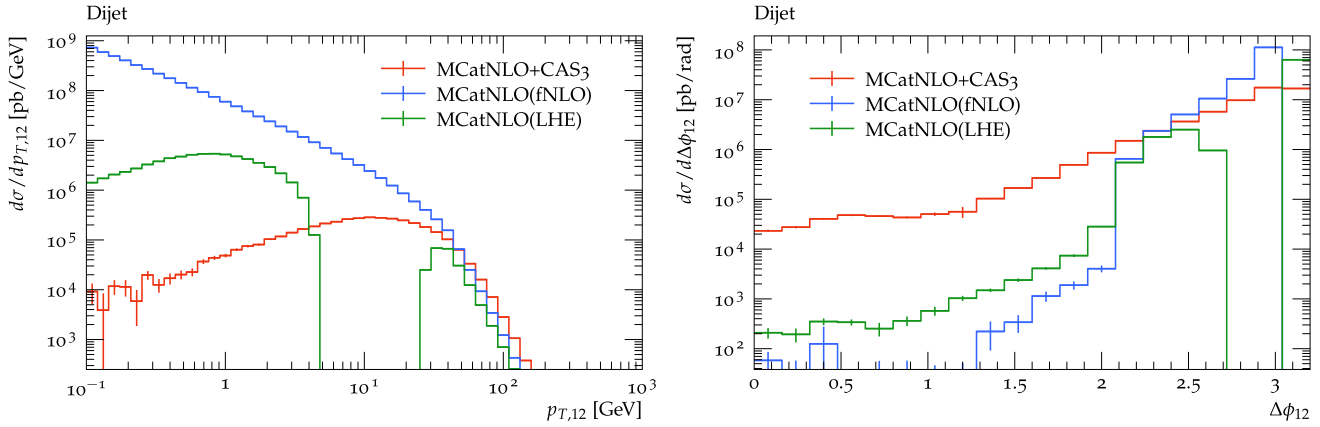


Fig. 3 Transverse momentum spectrum of the dijet system $p_{T,12}$ (left) and $\Delta\phi_{12}$ distribution (right). The predictions are shown for fixed NLO (MCatNLO(fNLO), the (unphysical) LHE level (MCatNLO(LHE)) and after inclusion of PB-TMDs (MCatNLO+CAS3)

anti-k_T jet-algorithm [65], as implemented in the FASTJET package [66], with a distance parameter of R=0.4 and a transverse momentum $p_T > 200$ GeV. The use of jets (instead of partons) is the reason for the tail towards small $\Delta\phi_{12}$ in the MCatNLO(LHE) and MCatNLO(fNLO) calculation.

4 Azimuthal correlations in multijet production

We next apply the framework described in the previous section, based on the matching of PB-TMDs with NLO matrix elements, to describe the measurement of azimuthal correlations $\Delta\phi_{12}$ obtained by CMS at $\sqrt{s} = 13$ TeV [26] and in the back-to-back region ($\Delta\phi_{12} \to \pi$) [27]. Only leading jets with a transverse momentum of $p_T^{\text{leading}} > 200$ GeV are considered. We show distributions of $\Delta\phi_{12}$ for $p_T^{\text{leading}} > 200$ GeV as well as for the very high p_T region of $p_T^{\text{leading}} > 1000$ GeV, where the jets appear very collimated. We apply the collinear and TMD set PB-NLO-2018-Set 2, unless explicitly specified, with running coupling $\alpha_s(m_Z) = 0.118$. We may estimate the theoretical uncertainties on the predictions by considering two kinds of uncertainties: those that come from variation of the arbitrary scales that appear in the various factors that enter the jet cross section, and those that come from the determination of the TMD parton distributions and showers. The former include

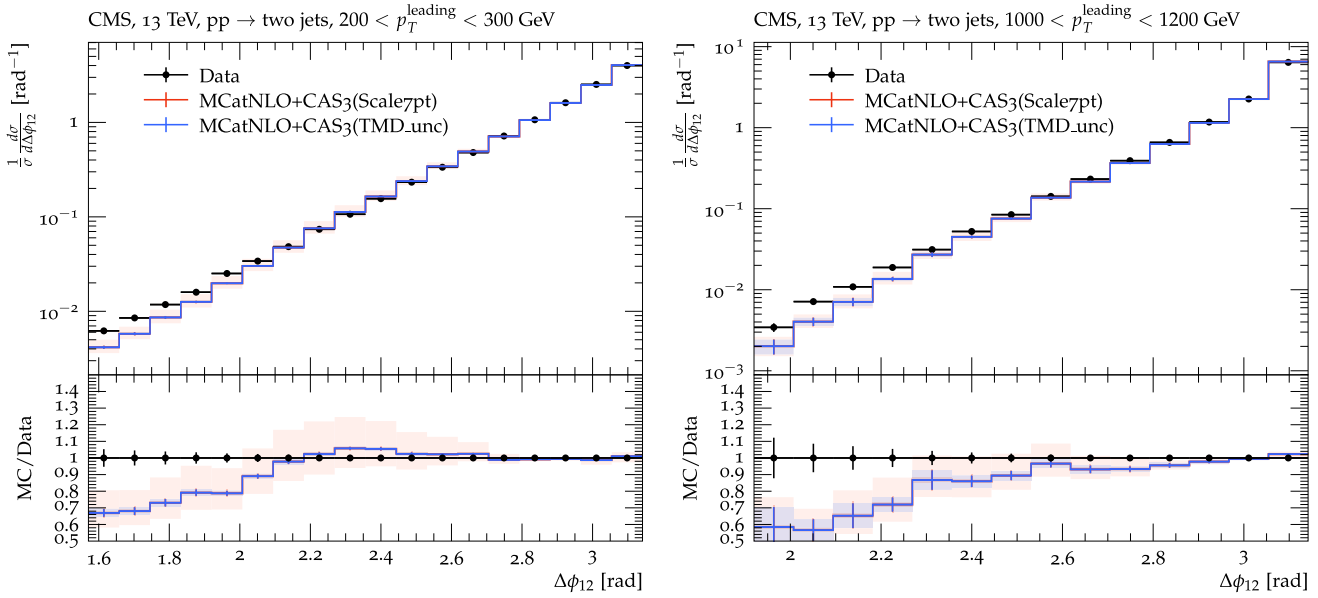


Fig. 4 Azimuthal correlation $\Delta\phi_{12}$ for $p_{\text{T}}^{\text{leading}} > 200$ GeV (left) and $p_{\text{T}}^{\text{leading}} > 1000$ GeV (right) as measured by CMS [26] compared with predictions from MCatNLO+CAS3. Shown are the uncertainties coming from the scale variation (as described in the text) as well as the uncertainties coming from the TMD

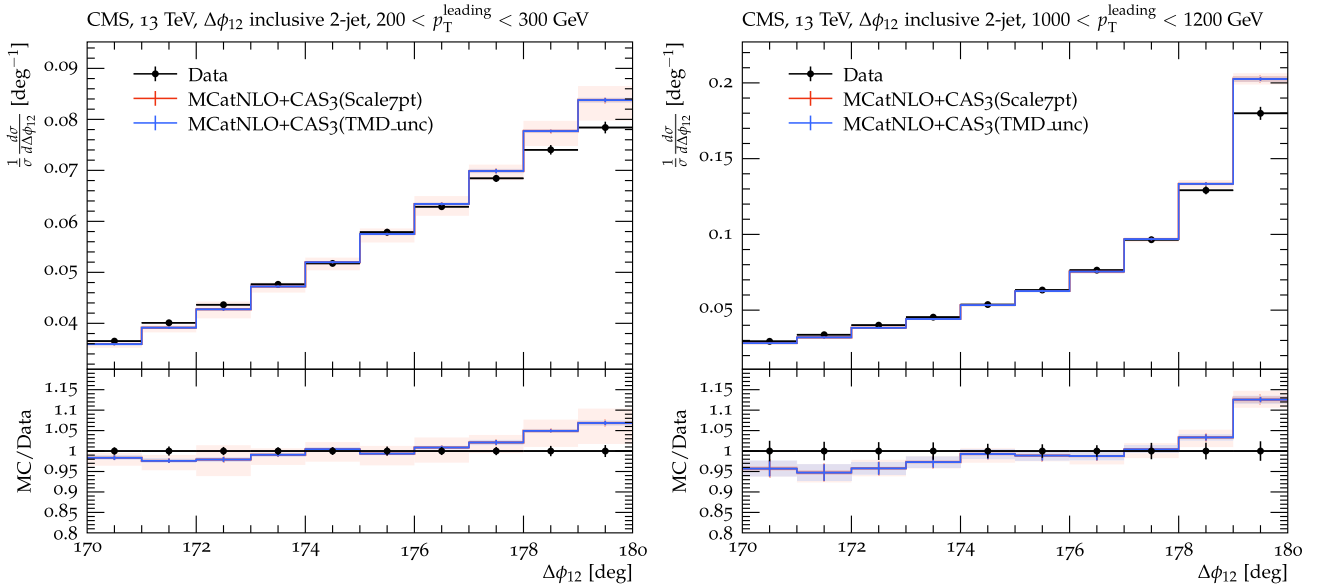


Fig. 5 Azimuthal correlation $\Delta\phi_{12}$ in the back-to-back region for $p_{\text{T}}^{\text{leading}} > 200$ GeV (left) and $p_{\text{T}}^{\text{leading}} > 1000$ GeV (right) as measured by CMS [27] compared with predictions from MCatNLO+CAS3. Shown are the uncertainties coming from the scale variation (as described in the text) as well as the uncertainties coming from the TMD

the renormalization scale in the strong coupling, the factorization scale used in the parton distribution and the matching scale to combine the matrix element and PB TMD. The latter include both experimental and model uncertainties in the TMD extraction. As regards the scale variations, we present results corresponding to the 7-point scheme variation around the central values for the renormalization and factorization scale (avoiding the extreme cases of variation). We have studied the dependence on the matching scale μ_m and found that is within the band of variation of factorization and renormalization scales. The experimental and model uncertainties on the determination of the TMD distributions as described in [51] are included.

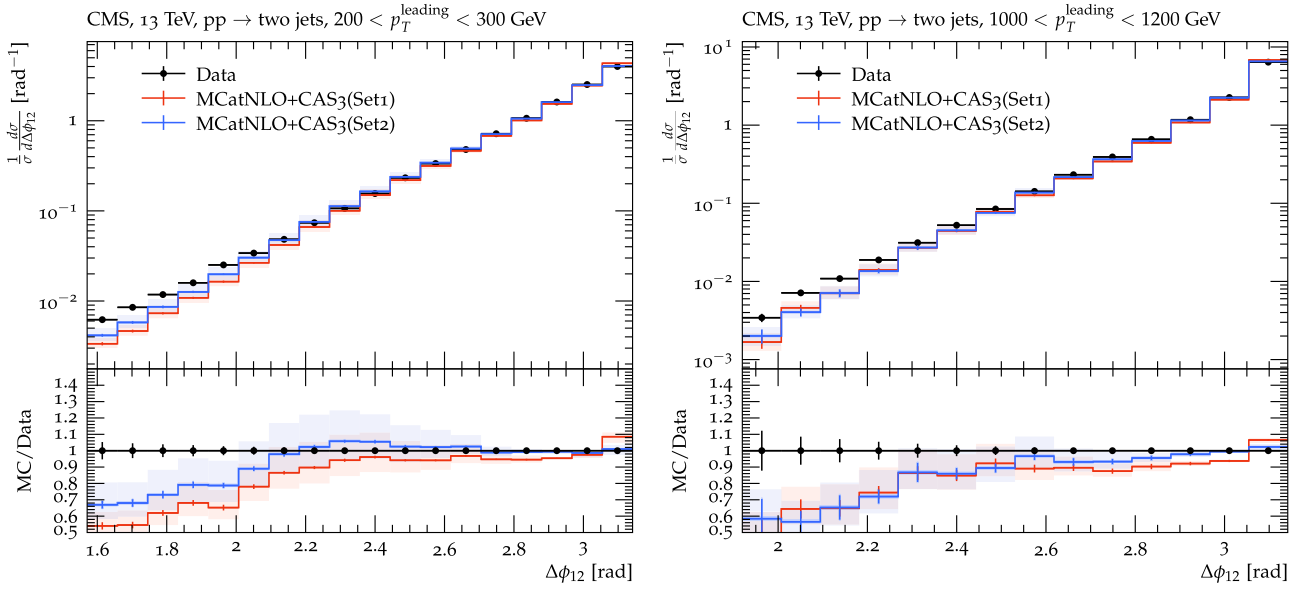


Fig. 6 Azimuthal correlation $\Delta\phi_{12}$ for $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV (left) and $p_{\mathrm{T}}^{\mathrm{leading}} > 1000$ GeV (right) as measured by CMS [26] compared with predictions from MCatNLO+CAS3. Shown are the uncertainties coming from the scale variation (as described in the text) as well as the uncertainties coming from the TMD

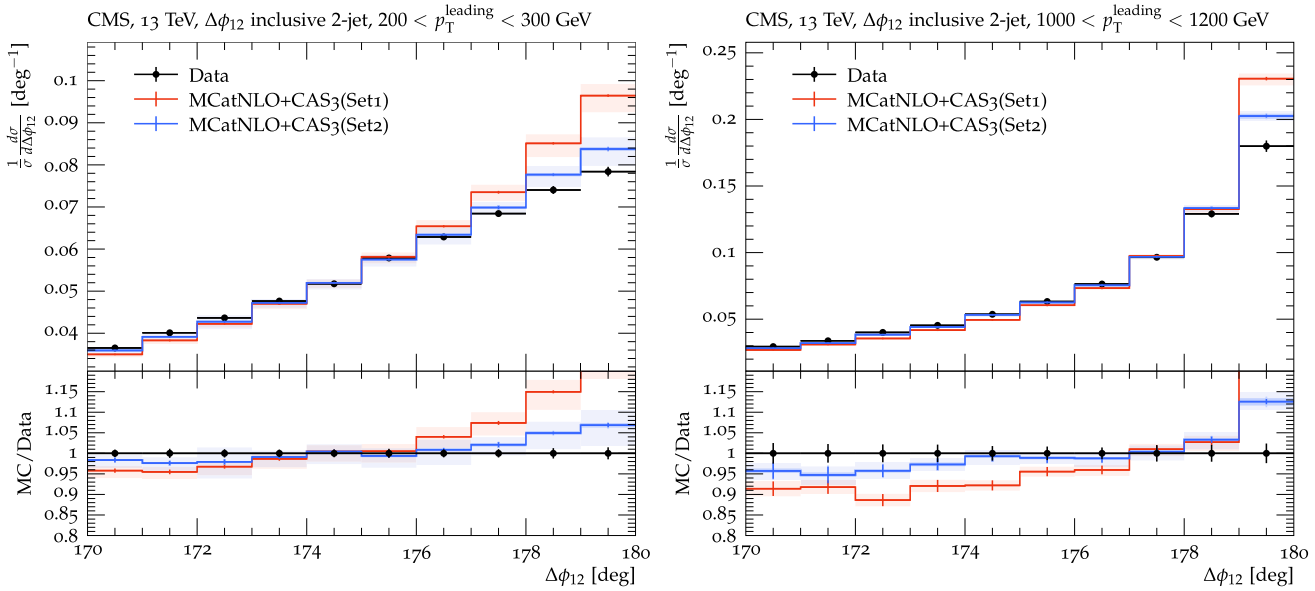


Fig. 7 Azimuthal correlation $\Delta\phi_{12}$ in the back-to-back region for $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV (left) and $p_{\mathrm{T}}^{\mathrm{leading}} > 1000$ GeV (right) as measured by CMS [27] compared with predictions from MCatNLO+CAS3. Shown are the uncertainties coming from the scale variation (as described in the text) as well as the uncertainties coming from the TMD

In Fig. 4 we show a comparison of the measurement by CMS [26] for different values of $p_{\mathrm{T}}^{\mathrm{leading}}$ with the calculation MCatNLO+CAS3 including PB-TMDs, parton shower, and hadronization. The uncertainties from scale variation and TMD determination are shown separately.

In Fig. 5 the measured $\Delta\phi_{12}$ distribution [27] in the back-to-back region is compared with the prediction MCatNLO +CAS3.

In general, the measurements are very well described, especially in the back-to-back region. The scale uncertainty is significantly larger than the TMD uncertainty, especially in the low $p_{\mathrm{T}}^{\mathrm{leading}}$ region. A difference between the measure-

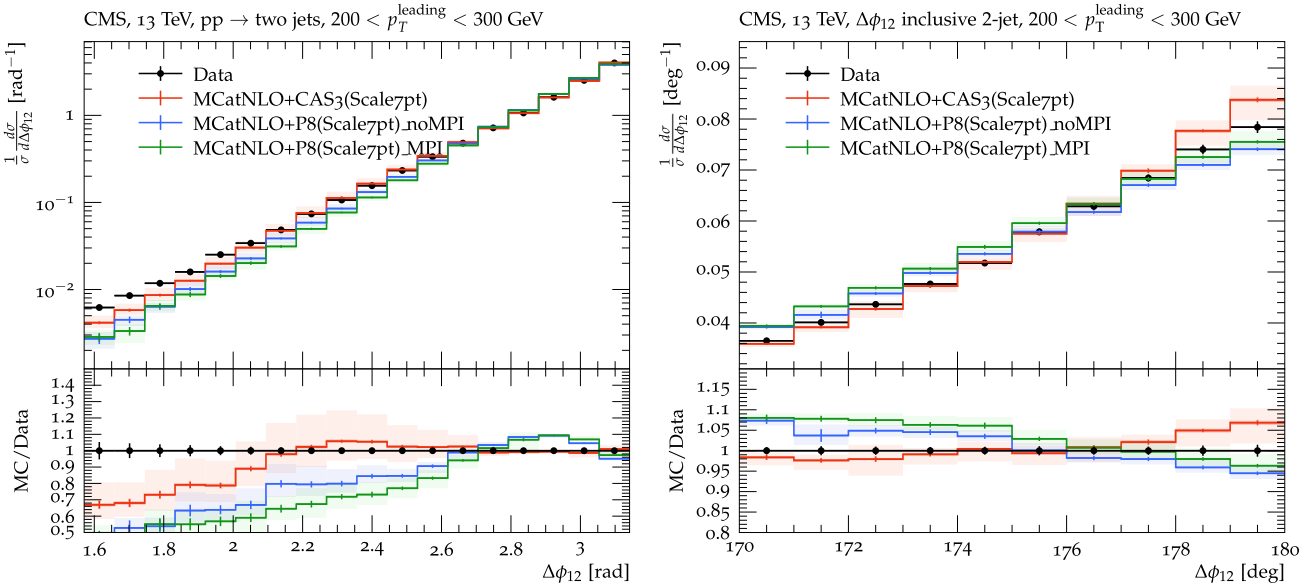


Fig. 8 Azimuthal correlation $\Delta\phi_{12}$ over a wide range and (left) in the back-to-back region (right) for $p_T^{leading} > 200$ GeV compared with predictions from MCatNLO+PYTHIA8 and MCatNLO+CAS3. The uncertainties in the MCatNLO+PYTHIA8 calculation are obtained from scale and associated shower variations, as described in the main text

ment and the prediction is observed for smaller $\Delta\phi_{12}$ which is due to missing higher order corrections in the matrix element calculation. Even at high $p_T^{leading} > 1000$ GeV the prediction is in agreement with the measurements (within uncertainties), while only in the highest $\Delta\phi_{12}$ bin ($\Delta\phi_{12} > 179^o$) a deviation of about 10% is observed.

In Figs. 6 and 7, the predictions using PB-NLO-2018-Set 1 are compared with those from PB-NLO-2018-Set 2 and with the measurements. The difference between Set 1 and Set 2 becomes significant in the back-to-back region, which is sensitive to the low k_T-region of the TMD. As already observed in the case of Z-boson production in Ref. [42], Set 2 with the transverse momentum as a scale for α_s, which is required from angular ordering conditions, allows a much better description of the measurement. It has been explicitly checked that the choice of the collinear parton density function (in contrast to the choice of the TMD densities) does not matter for the $\Delta\phi_{12}$ distributions, since they are normalized. The region of low $\Delta\phi_{12}$ in Figs. 4 and 6 is not well described with an NLO dijet matrix element calculation supplemented with TMD densities and TMD parton shower because in the low $\Delta\phi_{12}$ region higher-order hard emissions play a significant role. It has been shown in [67] that the inclusion of higher order matrix elements with the new TMD merging method of Ref. [44] leads to a very good description of the low $\Delta\phi_{12}$ region.

In Fig. 8 predictions obtained with MCatNLO+PYTHIA8 are compared with MCatNLO+CAS3. In the calculation of MCatNLO+PYTHIA8, the PYTHIA8 subtraction terms are used and the NNPDF3.0 [68] parton density and tune CUETP8M1 [69] are applied. The uncertainties of the PYTHIA prediction are derived by combining the fixed-order scale variation from MCatNLO with renormalization scale variations in the parton shower. We use the method of [70] together with the guidelines of [71] to obtain consistent scale variations where possible. In particular, this means that the renormalization scale variation at fixed order and in the parton shower are fully correlated.[1] The factorization scale variation is only applied at fixed order, as argued in [71]. We observe a significant dependence on the matching scale μ_m, the details of matching in case of dijets needs further investigation.

Shown in Fig. 8 is also the contribution from multiparton interactions, which is very small for jets with $p_T^{leading} > 200$ GeV. The prediction obtained with MCatNLO+PYTHIA8 is in all $\Delta\phi_{12}$ regions different from the measurement and MCatNLO+CAS3, illustrating the role of the treatment of parton showers.

In conclusion, the predictions of MCatNLO+CAS3 are in reasonable agreement with the measurements in the larger $\Delta\phi_{12}$ regions, where the contribution from higher order matrix elements is small. In the back-to-back region ($\Delta\phi_{12} \to \pi$) the predictions obtained with PB-TMDs and

[1] This also ensures that for fixed-order-dominated observables, the cancellation between the expansion of the shower and the subtraction in MC@NLO also occurs for non-central renormalization scales without significant deformation of the – there fully appropriate – fixed-order uncertainties.

parton shower are in good agreement with the measurement. The uncertainties of the predictions are dominated by the scale uncertainties of the matrix element calculations, while the PB-TMD and TMD shower uncertainties are very small, as they are directly coming from the uncertainties of the PB-TMDs. No uncertainties, in addition to those from the PB-TMD, come from the PB-TMD parton shower.

5 Conclusion

We have investigated the azimuthal correlation of high transverse momentum jets in pp collisions at $\sqrt{s} = 13$ TeV by applying PB-TMD distributions to NLO calculations via MCatNLO. We use the same PB-TMDs and MCatNLO calculations as we have used for Z -production at LHC energies in Ref. [42]. A very good description of the cross section as a function of $\Delta\phi_{12}$ is observed. In the back-to-back region of $\Delta\phi_{12} \to \pi$ a very good agreement is observed with PB-TMD Set 2 distributions [51] while significant differences are obtained with PB-TMD Set 1 distributions, which use the evolution scale as an argument in α_s. This observation confirms the importance of consistently handling the soft-gluon coupling in angular ordered parton evolution.

The uncertainties of the predictions are dominated by the scale uncertainties of the matrix element, while uncertainties coming from the PB-TMDs and the corresponding PB-TMD shower are very small. No other uncertainties, in addition to those of the PB-TMD, come from the PB-TMD shower, since it is directly correlated with the PB-TMD density.

We have also investigated predictions using MCatNLO with PYTHIA8 to illustrate the importance of details of the parton shower.

Acknowledgements This is based in part on studies during the "Virtual Monte Carlo school - PB TMDs with CASCADE3" [72] held from 8.-12. November 2021 at DESY, Hamburg. We are grateful to Olivier Mattelaer from the MADGRAPH5_AMC@NLO team for discussions, help and support with the LHE option for fixed NLO calculations in MCatNLO. STM thanks the Humboldt Foundation for the Georg Forster research fellowship and gratefully acknowledges support from IPM. D. E. Martins is supported by CNPQ-Brazil, process 164609/2020-2.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: There are no data to be deposited: the measurements are already on HEPdata.]

Funded by SCOAP[3].

References

1. ATLAS Collaboration, Measurement of the inclusive jet cross section in pp collisions at sqrt(s) = 2.76 TeV and comparison to the inclusive jet cross section at sqrt(s) = 7 TeV using the ATLAS detector. Eur. Phys. J. C **73**(8), 2509 (2013). arXiv:1304.4739
2. CMS Collaboration, Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV. Eur. Phys. J. C **76**(5), 265 (2016). arXiv:1512.06212
3. CMS Collaboration, Measurements of differential jet cross sections in proton–proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector. Phys. Rev. D **87**(11), 112002 (2013). arXiv:1212.6660. [Erratum: Phys. Rev. D **87**, 119902 (2013)]
4. CMS Collaboration, Measurement of the ratio of inclusive jet cross sections using the anti-k_T algorithm with radius parameters R = 0.5 and 0.7 in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. D **90**, 072006 (2014). arXiv:1406.0324
5. ATLAS Collaboration, Measurement of the inclusive jet cross-sections in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **09**, 020 (2017). arXiv:1706.03192
6. CMS Collaboration, Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV. JHEP **03**, 156 (2017). arXiv:1609.05331
7. ATLAS Collaboration, Measurement of inclusive jet and dijet cross-sections in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP **05**, 195 (2018). arXiv:1711.02692
8. CMS Collaboration, Measurement of the double-differential inclusive jet cross section in proton–proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **76**, 451 (2016). arXiv:1605.04436
9. CMS Collaboration, Dependence of inclusive jet production on the anti-k_T distance parameter in pp collisions at $\sqrt{s} = 13$ TeV. JHEP **12**, 082 (2020). arXiv:2005.05159
10. ALICE Collaboration, Measurements of inclusive jet spectra in pp and central Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Phys. Rev. C **101**, 034911 (2020). arXiv:1909.09718
11. ATLAS Collaboration, Measurement of inclusive jet and dijet production in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector. Phys. Rev. D **86**, 014022 (2012). arXiv:1112.6297
12. ATLAS Collaboration, Measurement of inclusive jet and dijet cross sections in proton–proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector. Eur. Phys. J. C **71**, 1512 (2011). arXiv:1009.5908
13. S. Alioli et al., Jet pair production in POWHEG. JHEP **04**, 081 (2011). arXiv:1012.3380
14. Z. Nagy, Three jet cross-sections in hadron hadron collisions at next-to-leading order. Phys. Rev. Lett. **88**, 122003 (2002). arXiv:hep-ph/0110315
15. W.T. Giele, E.W.N. Glover, D.A. Kosower, The two-jet differential cross section at $\mathcal{O}(\alpha_s^3)$ in hadron collisions. Phys. Rev. Lett. **73**, 2019–2022 (1994). arXiv:hep-ph/9403347
16. S.D. Ellis, Z. Kunszt, D.E. Soper, Two jet production in hadron collisions at order α_s^3 in QCD. Phys. Rev. Lett. **69**, 1496 (1992)
17. M. Czakon, A. van Hameren, A. Mitov, R. Poncelet, Single-jet inclusive rates with exact color at $\mathcal{O}(\alpha_s^4)$. JHEP **10**, 262 (2019). arXiv:1907.12911
18. A. Gehrmann-De Ridder et al., Triple differential dijet cross section at the LHC. Phys. Rev. Lett. **123**(10), 102001 (2019). arXiv:1905.09047

19. J. Currie, E.W.N. Glover, J. Pires, Next-to-next-to leading order QCD predictions for single jet inclusive production at the LHC. Phys. Rev. Lett. **118**(7), 072002 (2017). arXiv:1611.01460
20. J. Currie et al., Single jet inclusive production for the individual jet p_{T} scale choice at the LHC. Acta Phys. Pol. B **48**, 955–967 (2017). arXiv:1704.00923
21. D0 Collaboration, Measurement of dijet azimuthal decorrelations at central rapidities in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Phys. Rev. Lett. **94**, 221801 (2005). arXiv:hep-ex/0409040
22. D0 Collaboration, Measurement of the combined rapidity and p_T dependence of dijet azimuthal decorrelations in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Phys. Lett. B **721**, 212–219 (2013). arXiv:1212.1842
23. ATLAS Collaboration, Measurement of dijet azimuthal decorrelations in pp collisions at sqrt(s)= 7 TeV. Phys. Rev. Lett. **106**, 172002 (2011). arXiv:1102.2696
24. CMS Collaboration, Dijet azimuthal decorrelations in pp collisions at sqrt(s) = 7 TeV. Phys. Rev. Lett. **106**, 122003 (2011). arXiv:1101.5029
25. CMS Collaboration, Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C **76**, 536 (2016). arXiv:1602.04384
26. CMS Collaboration, Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **78**, 566 (2018). arXiv:1712.05471
27. CMS Collaboration, Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **79**, 773 (2019). arXiv:1902.04374
28. R. Angeles-Martinez et al., Transverse momentum dependent (TMD) parton distribution functions: status and prospects. Acta Phys. Pol. B **46**(12), 2501 (2015). arXiv:1507.05267
29. S. Dooling, P. Gunnellini, F. Hautmann, H. Jung, Longitudinal momentum shifts, showering and nonperturbative corrections in matched NLO-shower event generators. Phys. Rev. D **87**, 094009 (2013). arXiv:1212.6164
30. F. Hautmann, H. Jung, Collinearity approximations and kinematic shifts in partonic shower algorithms. Eur. Phys. J. C **72**, 2254 (2012). arXiv:1209.6549
31. A. Banfi, M. Dasgupta, Y. Delenda, Azimuthal decorrelations between QCD jets at all orders. Phys. Lett. B **665**, 86–91 (2008). arXiv:0804.3786
32. P. Sun, C.P. Yuan, F. Yuan, Soft gluon resummations in dijet azimuthal angular correlations in hadronic collisions. Phys. Rev. Lett. **113**(23), 232001 (2014). arXiv:1405.1105
33. P. Sun, C.P. Yuan, F. Yuan, Transverse momentum resummation for dijet correlation in hadronic collisions. Phys. Rev. D **92**(9), 094007 (2015). arXiv:1506.06170
34. Y. Hatta, B.-W. Xiao, F. Yuan, J. Zhou, Azimuthal angular asymmetry of soft gluon radiation in jet production. Phys. Rev. D **104**(5), 054037 (2021). arXiv:2106.05307
35. Y. Hatta, B.-W. Xiao, F. Yuan, J. Zhou, Anisotropy in dijet production in exclusive and inclusive processes. Phys. Rev. Lett. **126**(14), 142001 (2021). arXiv:2010.10774
36. J. Collins, J.-W. Qiu, k_T factorization is violated in production of high-transverse-momentum particles in hadron–hadron collisions. Phys. Rev. D **75**, 114014 (2007). arXiv:0705.2141
37. W. Vogelsang, F. Yuan, Hadronic dijet imbalance and transverse-momentum dependent parton distributions. Phys. Rev. D **76**, 094013 (2007). arXiv:0708.4398
38. T.C. Rogers, P.J. Mulders, No generalized TMD-factorization in hadro-production of high transverse momentum hadrons. Phys. Rev. D **81**, 094006 (2010). arXiv:1001.2977
39. F. Hautmann et al., Soft-gluon resolution scale in QCD evolution equations. Phys. Lett. B **772**, 446 (2017). arXiv:1704.01757
40. F. Hautmann et al., Collinear and TMD quark and gluon densities from Parton Branching solution of QCD evolution equations. JHEP **01**, 070 (2018). arXiv:1708.03279
41. S. Drell, T.-M. Yan, Massive lepton pair production in hadron–hadron collisions at high-energies. Phys. Rev. Lett. **25**, 316–320 (1970)
42. A. Bermudez Martinez et al., Production of Z-bosons in the parton branching method. Phys. Rev. D **100**, 074027 (2019). arXiv:1906.00919
43. A. Bermudez Martinez et al., The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method. Eur. Phys. J. C **80**, 598 (2020). arXiv:2001.06488
44. A. Bermudez Martinez, F. Hautmann, M.L. Mangano, TMD evolution and multi-jet merging. Phys. Lett. B **822**, 136700 (2021). arXiv:2107.01224
45. S.P. Baranov et al., Discriminating the heavy jet production mechanisms in associated Z + heavy flavor events at the LHC. arXiv:2111.04521
46. F. Hautmann, H. Jung, Angular correlations in multi-jet final states from kt-dependent parton showers. JHEP **10**, 113 (2008). arXiv:0805.1049
47. S. Dooling, F. Hautmann, H. Jung, Hadroproduction of electroweak gauge boson plus jets and TMD parton density functions. Phys. Lett. B **736**, 293 (2014). arXiv:1406.2994
48. M. Bury et al., Calculations with off-shell matrix elements, TMD parton densities and TMD parton showers. Eur. Phys. J. C **78**, 137 (2018). arXiv:1712.05932
49. J. Alwall et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP **1407**, 079 (2014). arXiv:1405.0301
50. S. Baranov et al., CASCADE3 A Monte Carlo event generator based on TMDs. Eur. Phys. J. C **81**, 425 (2021). arXiv:2101.10221
51. A. BermudezMartinez et al., Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method. Phys. Rev. D **99**, 074008 (2019). arXiv:1804.11152
52. ZEUS, H1 Collaboration, Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C **75**, 580 (2015). arXiv:1506.06042
53. N.A. Abdulov et al., TMDlib2 and TMDplotter: a platform for 3D hadron structure studies. Eur. Phys. J. C **81**, 752 (2021). arXiv:2103.09741
54. F. Hautmann et al., TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions. Eur. Phys. J. C **74**(12), 3220 (2014). arXiv:1408.3015
55. A. Bassetto, M. Ciafaloni, G. Marchesini, Jet structure and infrared sensitive quantities in perturbative QCD. Phys. Rep. **100**, 201–272 (1983)
56. Y.L. Dokshitzer, V.A. Khoze, S.I. Troian, A.H. Mueller, QCD coherence in high-energy reactions. Rev. Mod. Phys. **60**, 373 (1988)
57. S. Catani, B.R. Webber, G. Marchesini, QCD coherent branching and semiinclusive processes at large x. Nucl. Phys. B **349**, 635–654 (1991)
58. F. Hautmann, L. Keersmaekers, A. Lelek, A.M. Van Kampen, Dynamical resolution scale in transverse momentum distributions at the LHC. Nucl. Phys. B **949**, 114795 (2019). arXiv:1908.08524
59. A. Buckley et al., LHAPDF6: parton density access in the LHC precision era. Eur. Phys. J. C **75**, 132 (2015). arXiv:1412.7420
60. J. Alwall et al., A standard format for Les Houches event files. Comput. Phys. Commun. **176**, 300 (2007). arXiv:hep-ph/0609017
61. A. Buckley et al., Rivet user manual. Comput. Phys. Commun. **184**, 2803–2819 (2013). arXiv:1003.0694
62. G. Corcella et al., HERWIG 6.5 release note. arXiv:hep-ph/0210213

63. G. Marchesini et al., HERWIG: a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1-April 1991. Comput. Phys. Commun. **67**, 465–508 (1992)
64. T. Sjöstrand, S. Mrenna, P. Skands, PYTHIA 6.4 physics and manual. JHEP **05**, 026 (2006). arXiv:hep-ph/0603175
65. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). arXiv:0802.1189
66. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097
67. A. Bermudez Martinez et al., Jet production and TMD evolution. Presented at the workshop on Resummation, Evolution, Factorization (REF2021), DESY (2021)
68. NNPDF Collaboration, Parton distributions for the LHC Run II. JHEP **04**, 040 (2015). arXiv:1410.8849
69. CMS Collaboration, Event generator tunes obtained from underlying event and multiparton scattering measurements. Eur. Phys. J. C **76**, 155 (2016). arXiv:1512.00815
70. S. Mrenna, P. Skands, Automated parton-shower variations in Pythia 8. Phys. Rev. D **94**(7), 074005 (2016). arXiv:1605.08352
71. L. Gellersen, S. Prestel, Scale and scheme variations in unitarized NLO merging. Phys. Rev. D **101**(11), 114007 (2020). arXiv:2001.10746
72. A. Bermudez Martinez, H. Jung, S. Taheri Monfared, Q. Wang, Virtual Monte Carlo school-PB TMDs with CASCADE (2021). https://indico.desy.de/event/31877/

BIBLIOGRAPHY

[1] D0 Collaboration, "Measurement of dijet azimuthal decorrelations at central rapidities in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **94** (2005) 221801, doi:10.1103/PhysRevLett.94.221801, arXiv:hep-ex/0409040. 2

[2] D0 Collaboration, "Measurement of the combined rapidity and p_{T} dependence of dijet azimuthal decorrelations in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Lett. B* **721** (2013) 212, doi:10.1016/j.physletb.2013.03.029, arXiv:1212.1842. 2

[3] ATLAS Collaboration, "Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 172002, doi:10.1103/PhysRevLett.106.172002, arXiv:1102.2696. 2

[4] CMS Collaboration, "Dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 122003, doi:10.1103/PhysRevLett.106.122003, arXiv:1101.5029. 2

[5] CMS Collaboration, "Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **76** (2016) 536, doi:10.1140/epjc/s10052-016-4346-8, arXiv:1602.04384. 2

[6] CMS Collaboration, "Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **78** (2018) 566, doi:10.1140/epjc/s10052-018-6033-4, arXiv:1712.05471. 2, 99, 101

[7] CMS Collaboration, "Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **79** (2019) 773, doi:10.1140/epjc/s10052-019-7276-4, arXiv:1902.04374. 2, 100

[8] S. Chaturvedi and S. Biswas, "Fermi-Dirac statistics: Derivation and Consequences", *Resonance* **19** (01, 2014) doi:10.1007/s12045-014-0006-1. 4

[9] I. G. Kaplan, "Pauli Exclusion Principle and its theoretical foundation", 2019. 4

[10] Y. Yokoi and S. Abe, "Derivation of Bose–Einstein and Fermi–Dirac statistics from quantum mechanics: gauge-theoretical structure", *Journal of Statistical Mechanics: Theory and Experiment* **2018** (Feb, 2018) 023112, doi:10.1088/1742-5468/aaa794. 4

[11] "Standard Model". https://en.wikipedia.org/wiki/Standard_Model, Accessed: 2021-10-15. 4

[12] B. S. Acharya, "The Standard Model of Particle Physics". Lecture at the African School of Fundamental Physics and Applications 2016, Kigali, Rwanda, accessed: 2022-01-10. 5

[13] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508. 5

[14] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235. 5, 36

[15] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214. 5

[16] W. Buchmüller and C. Lüdeling, "Field Theory and Standard Model", doi:10.5170/CERN-2006-014.1. 5

[17] F. Sbisà, "Classical and quantum ghosts", *Eur. J. Phys.* **36** (2015) 015009, doi:10.1088/0143-0807/36/1/015009, arXiv:1406.4550. 6

[18] L. D. Faddeev and V. N. Popov, "Feynman Diagrams for the Yang-Mills Field", *Phys. Lett. B* **25** (1967) 29–30, doi:10.1016/0370-2693(67)90067-6. 6

[19] B. Foster, "Deep inelastic scattering at HERA", *Int. J. Mod. Phys. A* **13** (1998) 1543–1622, doi:10.1142/S0217751X9800069X, arXiv:hep-ex/9712030. 6

[20] A. Signer, "Low-energy Precision Physics and the High-energy Frontier", *Physics Procedia* **51** (12, 2014) 25–30, doi:10.1016/j.phpro.2013.12.007. 7

[21] Chekelian, Vladimir, "Proton structure and parton distribution functions from HERA", *EPJ Web Conf.* **126** (2016) 02005, doi:10.1051/epjconf/201612602005. 6, 7

[22] H.-W. Lin et al., "Parton distributions and lattice QCD calculations: a community white paper", *Prog. Part. Nucl. Phys.* **100** (2018) 107–160, doi:10.1016/j.ppnp.2018.01.007, arXiv:1711.07916. 6

[23] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, doi:10.1142/9789814503266_0001, arXiv:hep-ph/0409313. 7

[24] M. Bonvini and F. Giuli, "A new simple PDF parametrization: improved description of the HERA data", *Eur. Phys. J. Plus* **134** (2019), no. 10, 531, doi:10.1140/epjp/i2019-12872-x, arXiv:1902.11125. 8

[25] T.-J. Hou et al., "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC", *Phys. Rev. D* **103** (2021), no. 1, 014013, doi:10.1103/PhysRevD.103.014013, arXiv:1912.10053. 8

[26] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017), no. 10, 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428. 8

[27] M. Ciafaloni, "Coherence effects in initial jets at small Q2/s", *Nuclear Physics B* **296** (1988), no. 1, 49–74, doi:https://doi.org/10.1016/0550-3213(88)90380-X. 8

[28] S. Catani, F. Fiorani, and G. Marchesini, "QCD coherence in initial state radiation", *Physics Letters B* **234** (1990), no. 3, 339–345. 8

[29] G. Marchesini, "QCD coherence in the structure function and associated distributions at small x", *Nucl. Phys. B* **445** (1995) 49–80, doi:10.1016/0550-3213(95)00149-M, arXiv:hep-ph/9412327. 8

[30] R. Pasechnik and M. Šumbera, "Different Faces of Confinement", *Universe* **7** (2021), no. 9, doi:10.3390/universe7090330. 8

[31] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbative Theory in QCD (in Russian)", *Sov. Phys. JETP* **46** (1977) 641. 8

[32] V. Gribov and L. Lipatov, "Deep Inelastic ep Scattering in Perturbation Theory", *Sov. J. Nucl. Phys.* **15** (1972) 438. 8

[33] L. Lipatov, "The Parton Model and Perturbation Theory", *Sov. J. Nucl. Phys.* **20** (1975) 96. 8

[34] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys.* **B 126** (1977) 298. 8

[35] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and collider physics". Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, Cambridge, 2003. Photography by S. Vascotto. 9, 15

[36] F. Hautmann et al., "Collinear and TMD Quark and Gluon Densities from Parton Branching Solution of QCD Evolution Equations", *JHEP* **01** (2018) 070, doi:10.1007/JHEP01(2018)070, arXiv:1708.03279. 9, 11, 16

[37] R. Angeles-Martinez et al., "Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects", *Acta Phys. Polon. B* **46** (2015), no. 12, 2501–2534, doi:10.5506/APhysPolB.46.2501, arXiv:1507.05267. 11

[38] F. Hautmann and H. Jung, "Transverse momentum dependent gluon density from DIS precision data", *Nucl. Phys. B* **883** (2014) 1–19, doi:10.1016/j.nuclphysb.2014.03.014, arXiv:1312.7875. 11

[39] A. Bermudez Martinez et al., "Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method", *Phys. Rev. D* **99** (2019) 074008, doi:10.1103/PhysRevD.99.074008, arXiv:1804.11152. 11, 12, 98

[40] H. Jung, S. Taheri Monfared, and T. Wening, "Determination of collinear and TMD photon densities using the Parton Branching method", *Physics Letters B* **817** (2021) 136299, doi:https://doi.org/10.1016/j.physletb.2021.136299. 11

[41] F. Hautmann and H. Jung, "Collinearity approximations and kinematic shifts in partonic shower algorithms", *Eur. Phys. J. C* **72** (2012) 2254, doi:10.1140/epjc/s10052-012-2254-0, arXiv:1209.6549. 12, 113

[42] S. Baranov et al., "CASCADE3 A Monte Carlo event generator based on TMDs", *Eur. Phys. J. C* **81** (2021) 425, doi:10.1140/epjc/s10052-021-09203-8, arXiv:2101.10221. 12, 16, 18, 97, 98, 163

[43] A. Bermudez Martinez et al., "Production of Z-bosons in the parton branching method", *Phys. Rev. D* **100** (2019) 074027, doi:10.1103/PhysRevD.100.074027, arXiv:1906.00919. 12

[44] A. Bermudez Martinez et al., "The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method", *Eur. Phys. J. C* **80** (2020), no. 7, 598, doi:10.1140/epjc/s10052-020-8136-y, arXiv:2001.06488. 12, 113

[45] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301. 12

[46] H. Jung and S. T. Monfared, "TMD parton densities and corresponding parton showers: the advantage of four- and five-flavour schemes", arXiv:2106.09791. 12

[47] M. I. Abdulhamid et al., "Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method", *The European Physical Journal C* **82** (2022), no. 1, 36, doi:10.1140/epjc/s10052-022-09997-1. 12, 100, 163

[48] N. A. Abdulov et al., "TMDlib2 and TMDplotter: a platform for 3D hadron structure studies", arXiv:2103.09741. 12

[49] S. Catani and M. H. Seymour, "A General algorithm for calculating jet cross-sections in NLO QCD", *Nucl. Phys. B* **485** (1997) 291–419, doi:10.1016/S0550-3213(96)00589-5, arXiv:hep-ph/9605323. [Erratum: Nucl.Phys.B 510, 503–504 (1998)]. 14

[50] S. Frixione, Z. Kunszt, and A. Signer, "Three jet cross-sections to next-to-leading order", *Nucl. Phys. B* **467** (1996) 399–442, doi:10.1016/0550-3213(96)00110-1, arXiv:hep-ph/9512328. 14

[51] S. K. Kundu, T. Sarkar, and M. Maity, "Optimizing the parton shower model in PYTHIA with pp collision data at $\sqrt{s} = 13$ TeV", *Int. J. Mod. Phys. A* **34** (2019), no. 33, 1950219, doi:10.1142/S0217751X19502191, arXiv:1906.08472. 16

[52] S. Gieseke, P. Stephens, and B. Webber, "New formalism for QCD parton showers", *JHEP* **12** (2003) 045, doi:10.1088/1126-6708/2003/12/045, arXiv:hep-ph/0310083. 16

[53] J. Alwall et al., "A Standard format for Les Houches event files", *Comput. Phys. Commun.* **176** (2007) 300–304, doi:10.1016/j.cpc.2006.11.010, arXiv:hep-ph/0609017. 18, 101

[54] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244. 18

[55] T. Sjostrand and P. Z. Skands, "Multiple interactions and the structure of beam remnants", *JHEP* **03** (2004) 053, doi:10.1088/1126-6708/2004/03/053, arXiv:hep-ph/0402078. 19

[56] S. Bansal et al., "Progress in Double Parton Scattering Studies", in *5th International Workshop on Multiple Partonic Interactions at the LHC*. 10, 2014. arXiv:1410.6664. 18, 19

[57] M. Rinaldi, S. Scopetta, M. Traini, and V. Vento, "Double parton scattering: a study of the effective cross section within a Light-Front quark model", *Phys. Lett. B* **752** (2016) 40–45, doi:10.1016/j.physletb.2015.11.031, arXiv:1506.05742. 18, 19

[58] M. Diehl and J. R. Gaunt, "Double parton scattering theory overview", *Adv. Ser. Direct. High Energy Phys.* **29** (2018) 7–28, doi:10.1142/9789813227767_0002, arXiv:1710.04408. 19

[59] E. A. Zepeda Garcia and A. Ortiz, "Multiparton Interactions in pp collisions from Machine Learning", *PoS* **LHCP2021** (2021) 347, doi:10.22323/1.397.0347, arXiv:2110.01748. 19, 20

[60] S. Gieseke, C. Rohr, and A. Siodmok, "Multiparton interactions in Herwig++", in *3rd International Workshop on Multiple Partonic Interactions at the LHC*, pp. 51–56. 8, 2012. doi:10.3204/DESY-PROC-2012-03/45. 19, 21

[61] T. Sjöstrand, "The Development of MPI Modeling in Pythia", *Adv. Ser. Direct. High Energy Phys.* **29** (2018) 191–225, doi:10.1142/9789813227767_0010, arXiv:1706.02166. 20

[62] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815. 21

[63] G. S. Bali and K. Schilling, "Static quark-antiquark potential: Scaling behavior and finite-size effects in SU(3) lattice gauge theory", *Phys. Rev. D* **46** (Sep, 1992) 2636–2646, doi:10.1103/PhysRevD.46.2636. 22

[64] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics", *Physics Reports* **97** (1983), no. 2, 31–145, doi:https://doi.org/10.1016/0370-1573(83)90080-7. 22

[65] S. Ferreres-Solé and T. Sjöstrand, "The space–time structure of hadronization in the Lund model", *The European Physical Journal C* **78** (11, 2018) doi:10.1140/epjc/s10052-018-6459-8. 22

[66] F. Boge and C. Zeitnitz, "Polycratic hierarchies and networks: what simulation-modeling at the LHC can teach us about the epistemology of simulation", *Synthese* **199** (12, 2021) doi:10.1007/s11229-020-02667-3. 22

[67] T. Sjöstrand, "Old ideas in Hadronization: The Lund String". Lecture at Durham UK 15-17 April 2009, accessed: 2022-01-01. 23

[68] A. Kupco, "Cluster hadronization in HERWIG 5.9", in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, pp. 292–300. 4, 1998. arXiv:hep-ph/9906412. 23

[69] D. Amati and G. Veneziano, "Preconfinement as a Property of Perturbative QCD", *Phys. Lett. B* **83** (1979) 87–92, doi:10.1016/0370-2693(79)90896-7. 23

[70] G. Corcella et al., "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *JHEP* **01** (2001) 010, doi:10.1088/1126-6708/2001/01/010, arXiv:hep-ph/0011363. 23

[71] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189. 24

[72] M. Rangel, "Experimental aspects of jet physics at LHC", in *New Trends in High-Energy Physics and QCD*, pp. 230–234. 2016, 2016. arXiv:1608.00057. 24

[73] O. S. Brüning et al., "LHC Design Report". CERN Yellow Reports: Monographs. CERN, Geneva, 2004. 27

[74] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (aug, 2008) S08004–S08004, doi:10.1088/1748-0221/3/08/s08004. 27, 32

[75] K. Hübner, "The CERN intersecting storage rings (ISR)", *The European Physical Journal H* **36** (feb, 2012) 509–522, doi:10.1140/epjh/e2011-20058-8. 28

[76] N. A. Tahir et al., "The CERN Super Proton Synchrotron as a tool to study high energy density physics", *New J. Phys.* **10** (2008) 073028, doi:10.1088/1367-2630/10/7/073028. 28

[77] S. Myers, "The LEP Collider, from design to approval and commissioning". John Adams' memorial lecture. CERN, Geneva, 1991. Delivered at CERN, 26 Nov 1990. 28

[78] "Finding the W and Z". https://cerncourier.com/a/finding-the-w-and-z/, Accessed: 2021-05-15. 28

[79] L. Di Lella and C. Rubbia, "The Discovery of the W and Z Particles", *Adv. Ser. Direct. High Energy Phys.* **23** (2015) 137–163, doi:10.1142/9789814644150_0006. 29

[80] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN", (Dec, 2008). Taken from CERN CDS. 29

[81] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (aug, 2008) S08005–S08005, doi:10.1088/1748-0221/3/08/s08005. 28

[82] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (aug, 2008) S08003–S08003, doi:10.1088/1748-0221/3/08/s08003. 28

[83] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (aug, 2008) S08002–S08002, doi:10.1088/1748-0221/3/08/s08002. 28

[84] W. Herr and B. Muratori, "Concept of luminosity", doi:10.5170/CERN-2006-002.361. 30

[85] H. Burkhardt and P. Grafström, "Absolute Luminosity from Machine Parameters", technical report, 2007. 30

[86] S. van der Meer, "Calibration of the effective beam height in the ISR", technical report, CERN, Geneva, 1968. 31

[87] O. Karacheban, "Luminosity Measurement at the Compact Muon Solenoid Experiment of the LHC". PhD thesis, Brandenburg Tech. U., 2017. doi:10.3204/PUBDB-2017-11439. 31

[88] "CMS Luminosity-Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, Accessed: 2021-09-15. 31

[89] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032, arXiv:1311.4942. 32

[90] CMS Collaboration, "Bending particles". https://cms.cern/detector/bending-particles, Accessed: 2021-08-30. 34

[91] CMS Collaboration, "Alignment of the CMS Silicon Strip Tracker during stand-alone Commissioning", *JINST* **4** (Apr, 2009) T07001. 41 p, doi:10.1088/1748-0221/4/07/T07001, arXiv:0904.1220. Comments: 41 pages, 63 postscript figures, submitted to JINST. 35

[92] CMS Collaboration, "The Phase-2 Upgrade of the CMS Barrel Calorimeters", technical report, CERN, Geneva, Sep, 2017. This is the final version, approved by the LHCC. 36

[93] CMS Collaboration, "The CMS ECAL performance with examples", technical report, CERN, Geneva, Nov, 2013. doi:10.1088/1748-0221/9/02/C02008. 36, 37

[94] K. W. Bell et al., "Vacuum phototriodes for the CMS electromagnetic calorimeter endcap", *IEEE Trans. Nucl. Sci.* **51** (2004) 2284–2287, doi:10.1109/TNS.2004.836053. 37

[95] CMS ECAL Collaboration, "The CMS preshower construction and commissioning", *Nucl. Instrum. Meth. A* **617** (2010) 103–104, doi:10.1016/j.nima.2009.09.116. 37

[96] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", doi:10.2172/1151651. 38, 39

[97] S. Abdullin et al., "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges", *The European Physical Journal C* **55** (2008), no. 1, 159–171, doi:10.1140/epjc/s10052-008-0573-y. 38

[98] S. Abdullin et al., "Design, performance, and calibration of the CMS hadron-outer calorimeter", *The European Physical Journal C* **57** (2008), no. 3, 653–663, doi:10.1140/epjc/s10052-008-0756-6. 38

[99] S. Abdullin et al., "Design, performance, and calibration of CMS forward calorimeter wedges", *The European Physical Journal C* **53** (2008), no. 1, 139–166, doi:10.1140/epjc/s10052-007-0459-4. 38

[100] CMS Collaboration, G. L. Bayatian et al., "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical design report. CMS. CERN, Geneva, 2006. 40

[101] CMS Collaboration, J. G. Layter, "The CMS muon project: Technical Design Report". Technical design report. CMS. CERN, Geneva, 1997. 38

[102] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366. 39

[103] A. Sirunyan et al., "Particle-flow reconstruction and global event description with the CMS detector", *Journal of Instrumentation* **12** (oct, 2017) P10003–P10003, doi:10.1088/1748-0221/12/10/p10003. 41, 42

[104] T. Speer et al., "Track reconstruction in the CMS tracker", *Nucl. Instrum. Meth. A* **559** (2006) 143–147, doi:10.1016/j.nima.2005.11.207. 42

[105] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, Geneva, Apr, 2009. 43

[106] CMS Collaboration, "Jet performance in CMS", technical report, CERN, Geneva, Oct, 2013. 43

[107] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, Geneva, 2017. 43, 44, 56

[108] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503. 43

[109] CMS Collaboration, "How CMS weeds out particles that pile up". CMS Pysiscs briefing, accessed: 2022-01-05. 43

[110] T. Carli, K. Rabbertz, and S. Schumann, "Studies of Quantum Chromodynamics at the LHC", pp. 139–194. Springer Berlin Heidelberg, 2015. arXiv:1506.03239. doi:10.1007/978-3-319-15001-7_5. 44

[111] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS Physics Analysis Summary CMS-PAS-TRK-10-005, CERN, Geneva, 2010. 45

[112] CMS Collaboration and McCauley, Thomas, "Display of an event from cross section measurements of jet multiplicity and jet transverse momenta in multijet events at 13 TeV", (Aug, 2021). 45

[113] CMS Collaboration, "Measurements of jet multiplicity and differential production cross sections of $Z+$ jets events in proton-proton collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. D* **91** (2015), no. 5, 052008, doi:10.1103/PhysRevD.91.052008, arXiv:1408.3104. 46

[114] CMS Collaboration, "Measurement of the production cross section for Z + b jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:2112.09659. 46

[115] ATLAS Collaboration, "Measurements of multijet production cross sections in proton-proton collisions at 7 TeV center-of-mass energy with the ATLAS Detector",. 46

[116] CMS Collaboration, "Cross section measurements of jet multiplicity and jet transverse momenta in multijet events at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-SMP-21-006, CERN, Geneva, 2021. 47, 163

[117] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8. 50

[118] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820. 50

[119] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7. 50

[120] T. Sjostrand and P. Z. Skands, "Transverse-momentum-ordered showers and interleaved multiple interactions", *Eur. Phys. J.* **C39** (2005) 129–154, doi:10.1140/epjc/s2004-02084-y, arXiv:hep-ph/0408302. 50

[121] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815. 50, 100

[122] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J.* **C74** (2014), no. 8, 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630. 50

[123] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301. 50, 97, 98

[124] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129. 50

[125] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883. 50

[126] G. Marchesini and B. R. Webber, "Simulation of QCD Jets Including Soft Gluon Interference", *Nucl. Phys.* **B238** (1984) 1–29, doi:10.1016/0550-3213(84)90463-2. 50

[127] B. R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference", *Nucl. Phys.* **B238** (1984) 492–528, doi:10.1016/0550-3213(84)90333-X. 51

[128] CMS Collaboration and Mc Cauley, Thomas, "Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill", (Nov, 2016). 52

[129] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (2011), no. 11, P11002. 52, 58

[130] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021), no. 9, 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927. 52, 91

[131] CMS Collaboration, "Measurement of the inelastic *pp* cross section at sqrts = 7 TeV", CMS Physics Analysis Summary CMS-PAS-QCD-11-002, CERN, Geneva, 2011. 52

[132] J. M. Grados Luyando, "Charged particle spectra in different final states at $\sqrt{s} = 13$ TeV with the CMS Experiment". PhD thesis, Hamburg U., Hamburg, 2017. doi:10.3204/PUBDB-2018-00045. 54

[133] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663. 56, 57, 58

[134] CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018",. 58

[135] J. Pekkanen, "Jet Energy Resolution". `https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution`. CMS Private area, Accessed: 2021-08-21. 59, 61

[136] V. Lendermann et al., "Combining Triggers in HEP data analysis", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **604** (2009), no. 3, 707–718, `doi:https://doi.org/10.1016/j.nima.2009.03.173`. 61

[137] A. Sirunyan et al., "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$TeV using the CMS detector", *Journal of Instrumentation* **14** (Jul, 2019) P07004–P07004, `doi:10.1088/1748-0221/14/07/p07004`. 66

[138] CMS Collaboration, "MET Optional Filters for Run1". CMS twiki page, accessed: 2022-01-05. 67

[139] CMS Collaboration, "MET Filter Recomendations for Run II". CMS twiki page, accessed: 2022-01-05. 67

[140] V. Lefébure, S. Banerjee, and I. González, "CMS Simulation Software Using Geant4", Technical Report CMS-NOTE-1999-072, CERN, Geneva, Dec, 1999. 75

[141] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *JINST* **7** (2012) T10003, `doi:10.1088/1748-0221/7/10/T10003`, `arXiv:1205.6201`. 76

[142] CMS Collaboration, V. I. Behnke O., Kuusela M., "Recommendations on Unfolding". CMS twiki page, accessed: 2022-01-05. 76

[143] Wikipedia. `https://en.wikipedia.org/wiki/Moore-Penrose_inverse`. Accessed: 2020-01-06. 76

[144] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, `doi:10.1007/JHEP04(2015)040`, `arXiv:1410.8849`. 97

[145] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, `arXiv:hep-ph/0603175`. 98

[146] CMS Collaboration, "Measurement of the double-differential inclusive jet cross section in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J.* **C76** (2016), no. 8, 451, `doi:10.1140/epjc/s10052-016-4286-3`, `arXiv:1605.04436`. 98

[147] A. Buckley et al., "The HepMC3 event record library for Monte Carlo event generators", *Comput. Phys. Commun.* **260** (2021) 107310, `doi:10.1016/j.cpc.2020.107310`, `arXiv:1912.08005`. 105

[148] C. Bierlich et al., "Robust Independent Validation of Experiment and Theory: Rivet version 3", *SciPost Phys.* **8** (2020) 026, `doi:10.21468/SciPostPhys.8.2.026`, `arXiv:1912.05451`. 105

[149] ATLAS Collaboration Collaboration, "A new method to distinguish hadronically decaying boosted Z bosons from W bosons using the ATLAS detector. A new method to distinguish hadronically decaying boosted Z bosons from W bosons using the ATLAS detector", *Eur. Phys. J. C* **76** (Sep, 2015) 238. 46 p, doi:10.1140/epjc/s10052-016-4065-1, arXiv:1509.04939. 32 pages plus author list (49 pages total), 19 figures, published version, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2015-02/. 117

[150] ATLAS Collaboration, "Measurements of top-quark pair single- and double-differential cross-sections in the all-hadronic channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector", *JHEP* **01** (2021) 033, doi:10.1007/JHEP01(2021)033, arXiv:2006.09274. 117

[151] CMS Collaboration, "Search for light pseudoscalar boson pairs produced from decays of the 125 GeV Higgs boson in final states with two muons and two nearby tracks in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **800** (2020) 135087, doi:10.1016/j.physletb.2019.135087, arXiv:1907.07235. 117

[152] CMS Collaboration, "Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **04** (2019) 031, doi:10.1007/JHEP04(2019)031, arXiv:1810.05905. 117

[153] CMS Collaboration, "Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **77** (2017), no. 9, 636, doi:10.1140/epjc/s10052-017-5192-z, arXiv:1707.01303. 117

[154] CMS Collaboration, "Search for heavy resonances in the all-hadronic vector-boson pair final state with a multi-dimensional fit", CMS Physics Analysis Summary CMS-PAS-B2G-18-002, CERN, Geneva, 2019. 117

[155] T. Carli, "Small x QCD effects in particle collisions at high energies", *Int. J. Mod. Phys. A* **17** (2002) 3185–3203, doi:10.1142/S0217751X02012685, arXiv:hep-ph/0111232. 117

[156] J. Collins and J.-W. Qiu, "k_T factorization is violated in production of high-transverse-momentum particles in hadron-hadron collisions", *Phys. Rev. D* **75** (2007) 114014, doi:10.1103/PhysRevD.75.114014, arXiv:0705.2141. 117

[157] W. Vogelsang and F. Yuan, "Hadronic Dijet Imbalance and Transverse-Momentum Dependent Parton Distributions", *Phys. Rev. D* **76** (2007) 094013, doi:10.1103/PhysRevD.76.094013, arXiv:0708.4398. 117

[158] T. C. Rogers and P. J. Mulders, "No Generalized TMD-Factorization in Hadro-Production of High Transverse Momentum Hadrons", *Phys. Rev. D* **81** (2010) 094006, doi:10.1103/PhysRevD.81.094006, arXiv:1001.2977. 117

[159] A. Dainese et al., "Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC", technical report, Geneva, Switzerland, 2019. doi:10.23731/CYRM-2019-007. 117

[160] G. Bohm and G. Zech, "Introduction to statistics and data analysis for physicists". DESY, Hamburg, 2010. 121

[161] F. James, "Monte Carlo theory and practice", *Reports on Progress in Physics - REP PROGR PHYS* **43** (09, 1980) 1145–1189, doi:10.1088/0034-4885/43/9/002. 121

[162] S. Weinzierl, "Introduction to Monte Carlo methods", arXiv:hep-ph/0006269. 122, 123

ACKNOWLEDGMENTS

First of all, I would like to give my eternal gratitude to Dr. Hannes Jung, my main supervisor, who guided me throughout this project. Thank you very much for everything I have learned from you in these four years. I am very grateful for all the invaluable comments you gave me to improve this thesis.

Thanks to DESY CMS group for awarding me a position as PhD student in Particle Physics. Also to SMP CMS group for all the insightful feedback received during the working meetings. All my gratitude to Dr. Patrick Connor, who supported me through the realization of this work. Also to Dr. Armando Bermudez who was always nice to discuss and gave me his feedback.

To all my piers and friends at DESY, and the " Hannes' QCD group" thanks for the time we shared during lunch and Friday Coffee. To the "Cuban Community from InSTEC" here in Hamburg, I hope we can share more grills and friendly meetings together.

I would also like to give thanks to Dr. Fernando Guzman, who supported my application to the summer student program at DESY, which was the very start of my journey in the world of Particle Physics. Dear Fernando, thanks a lot for the nice discussions about philosophy and physics. Also to all the great professors I had at InSTEC during my studies in my beloved Cuba, my most sincere gratitude.

Finaly thanks to my wife, family and numerous friends who have been supporting me through this long way, always offering their love and good wishes to my future career.