Measurement and phenomenology of azimuthal correlations in high transverse momentum multi-jet topologies in CMS at the center-of-mass energy of 13 TeV

Dissertation
zur Erlangung des Doktorgrades
des Fachbereiches Physik
der Universität Hamburg

vorgelegt von
Armando Bermúdez Martínez
aus Havanna, Kuba

Hamburg
2019
Gutachter der Dissertation:  
PD Dr. Hannes Jung  
Prof. Dr. Elisabetta Gallo  
Prof. Dr. Leszek Motyka

Gutachter der Disputation:  
Prof. Dr. Daniela Pfannkuche  
Prof. Dr. Elisabetta Gallo  
Prof. Dr. Sven-Olaf Moch  
Dr. Paolo Gunnellini  
PD Dr. Hannes Jung

Vorsitzender des Prüfungsausschusses:  
Prof. Dr. Daniela Pfannkuche

Datum der Disputation:  
21. Juni 2019

Vorsitzender des Fach-Promotionsausschusses Physik:  
Prof. Dr. Wolfgang Hansen

Leiter des Fachbereichs Physik:  
Prof. Dr. Michael Potthoff

Dekan der Fakultät für Mathematik,  
Informatik und Naturwissenschaften:  
Prof. Dr. Heinrich Graener
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 7. März 2019
ACKNOWLEDGMENTS

To those who have made the completion of this work possible: many thanks!, vielen Dank!, muchas gracias!

My supervisor PD Dr. Hannes Jung: I was just remembering how this journey started, it was a dark, snowy day in the Winter of 2016. Interestingly, it ended in a slightly different way, on a warm, sunny day three years later. You were there on both those days, and on every day in between making sure, that the sun rose at the end. Thank you for opening my eyes to the beauty of particle physics.

I’m grateful to the examination commission members Prof. Dr. Daniela Pfannkuche, Prof. Dr. Elisabetta Gallo, Prof. Dr. Sven-Olaf Moch, Dr. Paolo Gunnellini, and PD Dr. Hannes Jung, for the time dedicated to evaluate and grade the work. My gratitude also goes to Prof. Dr. Leszek Motyka, Prof. Dr. Elisabetta Gallo, and PD Dr. Hannes Jung, referees of the work, for the effort dedicated to assess the studies contained in the dissertation. Thanks Elisabetta, Radek, Patrick, and specially Hannes for the corrections, remarks and suggestions which were decisive for the manuscript to achieve its final shape.

A very special gratitude goes to Hannes, Paolo, Ben, Radek, Panos Kokkas, Francesco Hautmann, Leif Lönnblad, Torbjörn Sjöstrand, and Johannes Bellm for the invaluable discussions on physics.

Thanks to DESY, specially its CMS group, the secretaries, the International Office, the Human Resources department, and the canteen personnel for providing the best environment not only for research but also for making friends. I’m also grateful to the MCnet network specially the department of Theoretical Physics at the University of Lund and my supervisors at the latter Leif Lönnblad and Torbjörn Sjöstrand, who have been crucial when it came to learn the interplay between experiment and phenomenology.

A very special gratitude to Daniel Codorniu Pujals, Fernando Guzmán Martínez, Oscar Rodríguez Hoyos, and the nuclear physics department at the InSTEC in Havana, from whom I learned how to learn. Thank you so much Mirta and Genaro for paving my path to university.

I would also like to thank my dear friends in Hamburg, Havana and elsewhere who were watching every step, providing feedback and support, making me feel confident and at home,
and for showing that the line separating friends and family is, in many cases, a fictitious one.

I’m deeply grateful to my family, specially my little sister Amanda for making me smile every time we talk and making me feel that Cuba is not that far away, and Madelin for her support and encouragement at every moment. Thanks Alexis for the fruitful discussions. I’m infinitely grateful to my wonderful parents for being so present and their unconditional support, for planting the seed of curiosity in me and water it ever since, for teaching me everything I know and making me who I am. Finally thanks to Anneke, my lifemate, friend, and mother of our child, for demonstrating every day that the beauty of life lies in the simplest things and for pushing me and giving me the “German treatment” in the critical moments.
ABSTRACT

This work addresses the measurements and phenomenology of the azimuthal separation $\Delta \phi_{12}$ between the jets with the largest transverse momenta ($p_T$) in multi-jet event topologies. The measurements are based on proton-proton collision data at the center-of-mass energy of 13 TeV recorded by the Compact Muon Solenoid (CMS) experiment in 2016, and correspond to an integrated luminosity of 35.9 fb$^{-1}$. The measurements are performed in ranges of the leading jet transverse momentum ($p_T^{\text{max}}$), covering the range $p_T^{\text{max}} > 200$ GeV. Special attention is given to the investigation of the Sudakov region $170^\circ < \Delta \phi_{12} < 180^\circ$ in inclusive 2- and 3-jet event topologies. This region characterizes nearly back-to-back jet topologies and is sensitive to contributions from soft partonic radiation. The observables are compared to theoretical predictions based on matrix element (ME) calculation at leading- and next-to-leading-order accuracy in perturbative quantum chromodynamics (pQCD) combined with parton showers (PS). Discrepancies of up to 15% are observed for $\Delta \phi_{12} > 177^\circ$. The measurements are complemented with a wide phenomenological investigation of the PS impact and its subsequent combination with the ME calculation on the $\Delta \phi_{12}$ distribution. The measurements of $\Delta \phi_{12}$ in nearly back-to-back jet topologies for inclusive 2- and 3-jet events are not simultaneously described by any of the theoretical models. Predictions using ME and PS at different accuracies in pQCD provide a similar description of the data.
## 4 Phenomenology of the $\Delta \phi_{12}$ distribution

4.1 Fixed-order calculations and parton shower applied to the $\Delta \phi_{12}$ distribution

4.1.1 Initial- and final-state shower contributions

4.2 Impact of the matching of matrix element to parton shower

4.2.1 MC@NLO matching

4.2.2 POWHEG matching

4.2.3 Relation between POWHEG and MC@NLO. The parameter $h_{\text{damp}}$

4.2.4 LO multi-jet merging

4.3 Transverse momentum dependent parton distribution applied to $\Delta \phi_{12}$

## 5 Experimental setup

5.1 The Large Hadron Collider

5.1.1 Luminosity and data taking

5.2 The Compact Muon Solenoid

5.2.1 Inner tracking system

5.2.2 Electromagnetic calorimeter and preshower detector

5.2.3 Hadronic calorimeter system

5.2.4 Muon detector system

5.2.5 Trigger system

5.2.6 Detector simulation

## 6 Event reconstruction and selection

6.1 Data and simulated samples

6.2 Physics objects and vertex reconstruction

6.2.1 Jet reconstruction

6.2.2 Trigger path

6.2.3 Jet energy calibration and jet energy resolution

6.3 Event selection

6.3.1 Observables

6.4 Bin size and resolution

6.4.1 Coarser binning

6.4.2 Finer binning

6.4.3 Relation between tracks $\phi$ resolution and $\Delta \phi_{12}$ resolution

6.5 Purity, stability, acceptance and background

6.5.1 The relation between resolution and purity/stability

6.6 Pileup

6.7 The measurement at detector level

6.8 Detector effects and unfolding to particle level

6.8.1 Unfolding methods

6.8.2 Response matrices and unfolded distributions

6.8.3 Regularization

6.8.4 D’Agostini vs Tikhonov unfolding

6.8.5 Migration matrices obtained with a new method

6.8.6 Validation of the unfolding
7 Experimental uncertainties
7.1 Jet energy scale ........................................ 107
7.2 Jet energy resolution .................................. 109
7.3 Unfolding ............................................... 110
  7.3.1 $\Delta\phi_{12}$ resolution .............................. 110
  7.3.2 Model dependence .................................. 112
7.4 Statistical uncertainty ................................. 113
7.5 Experimental uncertainties overview ................. 113

8 Theoretical uncertainties ................................. 117
8.1 Scale dependence ...................................... 118
8.2 Parton distributions .................................... 118
8.3 Strong coupling ....................................... 120
8.4 Parton showering ...................................... 120
8.5 Non-perturbative sources .............................. 123
8.6 Comparison between theoretical and experimental uncertainties .... 126

9 Results and discussion .................................... 129
9.1 Measurement of $\Delta\phi_{12}$ in inclusive 2-jet topologies .... 130
9.2 Studies on the matching of matrix element to parton shower ...... 135
  9.2.1 MC@NLO matching .................................. 135
  9.2.2 POWHEG matching .................................. 136
  9.2.3 Relation between the POWHEG and MC@NLO .............. 138
  9.2.4 LO multi-jet merging ................................ 139
  9.2.5 Matrix element correction to parton shower ................ 141
  9.2.6 Application of TMDs to the calculation of the $\Delta\phi_{12}$ distribution .... 141
9.3 Measurement of $\Delta\phi_{12}$ for nearly back-to-back jet topologies in inclusive 2- and 3-jet events ........................................ 144
9.4 Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies .............................. 151
  9.4.1 MC@NLO matching and impact on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies ............... 151
  9.4.2 POWHEG matching and impact on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies ............... 151
  9.4.3 Relation between POWHEG and MC@NLO, and the impact on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies .... 154
  9.4.4 Impact of LO multi-jet merging on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies ...................... 157
  9.4.5 Impact of the parton shower on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies ...................... 157
  9.4.6 Application of TMDs to the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies ...................... 159
9.5 Prospects and extension of the studies .................. 162

10 Summary and conclusions ............................... 167
Appendices

A Azimuthal correlations for inclusive 2-jet, 3-jet and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV 173

B Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV 187

C Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method 205

D New method for migration matrix determination and resolution effects estimation 229

E Matching/merging and the inclusive jets transverse momentum cross section 239

F Matching/merging and azimuthal correlations in 4-jet events 247

G Matching NLO matrix element with initial transverse momentum from TMD evolution to parton shower 251

H Resolution, purity, stability, acceptance and background 255

I Response matrices, fractional errors and correlation matrices 261

J Tests for the unfolding of the inclusive 3-jet $\Delta\phi_{12}$ distributions 269

Bibliography 275
CHAPTER

1

INTRODUCTION

Nowadays, the Large Hadron Collider (LHC) has pushed the frontier for the observation of physical phenomena to typical distances of a per mille of the proton size. At the LHC this has been achieved by colliding protons (and heavier ions) at very high center-of-mass energies. The LHC is therefore a machine with an unprecedented chance of measuring the properties of the fundamental constituents of nature. The data that have been obtained with the LHC has helped to consolidate the Standard Model as the theory that describes within the same framework the electromagnetic, weak and strong interactions. Since the protons are composite objects made up of partons (quarks and gluons), it is not surprising that the bulk of the interactions occurring at the LHC are of strong nature. Even though the strong interaction described by quantum chromodynamics (QCD) has been tested and confirmed by the recorded data, many questions remain partially or even unanswered.

The calculation and interpretation of the cross section for proton-proton (pp) collisions at high momentum transfer relies on the factorization of the non-perturbative physics at hadronic energy scales and the physics that can be perturbatively calculated at higher energy scales. At high momentum transfer pp collisions are characterized by the production of jets that carry the information on the underlying partonic process. At leading order in perturbative QCD (pQCD) two jets (partons) are produced back-to-back in the transverse plane. As the partons evolve in energy down to typical hadronic scales they emit radiation, producing a recoil in transverse momentum that boosts the original 2-jet system. The harder the emitted radiation is, the smaller the resulting azimuthal separation of the leading 2-jet system ($\Delta\phi_{12}$) becomes. When mostly soft radiation is emitted instead, the 2-jet system is less perturbed and remains nearly back-to-back.

In modern Monte Carlo (MC) event generators the description of hard radiation relies on fixed higher-order calculations in pQCD. However, fixed-order matrix elements become unreliable when the emitted partons are soft because of large logarithmic contributions to the cross section. Meaningful predictions involving soft partonic emissions rely on the resummation at all orders of the large logarithmic contributions, which in MC event generators is realized in
Chapter 1. Introduction

the form of parton showers (PS). In addition, non-trivial correlations due to the interference between colored initial- and final-states through the emission of soft partonic radiation can occur. Modern MC event generators combine ME calculations with the resummed prediction from PS. While the leading-order (LO) multi-jet merging method is aimed to improve the theoretical predictions by using higher LO ME, methods like Powheg and MC@NLO are used to combine next-to-LO (NLO) ME with PS. Although these methods should preserve the given accuracy of the integrated cross section, the resummed contributions and correlations in exclusive distributions, like the azimuthal separation $\Delta\phi_{12}$ between the two leading jets in multi-jet events, differ.

In this way $\Delta\phi_{12}$ is a unique and important observable to test pQCD including the nearly back-to-back region where it becomes crucial to take into account the resummation of soft contributions to all orders and the correlations between the initial- and final-state partons. This observable has been previously measured by several collaborations (Refs. [1–6]). In this thesis I present the measurement of $\Delta\phi_{12}$ with the Compact Muon Solenoid (CMS) detector in pp collisions at the highest achieved center-of-mass energy of 13 TeV. Additionally, I present the novel measurement of $\Delta\phi_{12}$ in nearly back-to-back jet topologies which is based on the excellent tracker performance of the CMS detector. The two aforementioned measurements are complemented with a wide phenomenological discussion using the theoretical predictions from higher-order ME and PS calculations.

The thesis is structured in the following way: in Chapter 2 the Standard Model, the theoretical framework which the thesis relies on, is discussed. In Chapter 3, an overview of the QCD sector of the Standard Model as well as a discussion on important concepts which are necessary for calculating QCD processes in MC event generators are presented, with a special emphasis on PS generation and the subsequent combination with ME. Chapter 4 is dedicated to the phenomenological implications of the application of QCD calculations to the $\Delta\phi_{12}$ distribution, with special attention given to the PS and its subsequent matching to ME, as well as to the impact of using transverse momentum dependent (TMD) parton distributions for the evolution of the initial-state partons and the generation of initial-state shower (ISR). In Chapter 5 the LHC is introduced as well as the CMS experiment and its different sub-detectors used for the measurements discussed in this thesis. The event reconstruction and selection corresponding to the measurements, as well as the studies on the resolution of the observables, bin migration due to detector effects, and the unfolding procedure to correct for these effects are discussed in detail in Chapter 6. Subsequently, in Chapter 7 the experimental uncertainty sources and their contribution to the total experimental uncertainty of the measurements are presented. Chapter 8 is focused on the theoretical uncertainty contributions, from both perturbative and non-perturbative origins, that affect the calculation of the theoretical predictions. The results of the measurement of the $\Delta\phi_{12}$ distribution as well as the comparison of the data to predictions from LO and NLO event generators matched to PS are presented in Chapter 9. The impact on the results of the matching of ME to PS, the effects of the use of TMD parton distributions for the initial-state evolution, and the relation between the observations at low and high $\Delta\phi_{12}$ are also discussed in Chapter 9. At the end of Chapter 9 possible extensions of the investigations discussed in this thesis are presented which include the measurement of $\Delta\phi_{12}$ in bins of rapidity and the comparison to the analogous azimuthal separation in processes involving non-colored objects in the final state. Chapter 10 is dedicated to overview and summarize the studies presented in the thesis as well as to assess the impact on future investigations.
The investigations reported in the previously mentioned Chapters are supplemented with studies which are organized in ten appendices. Appendix A reports the measurement of the minimum azimuthal separation between any two of the four hardest jets in inclusive 3- and 4-jet event topologies, in addition to the measurement of $\Delta \phi_{12}$ in multi-jet events. The measurement of $\Delta \phi_{12}$ in nearly back-to-back jet topologies in inclusive 2- and 3-jet events is reported in App. B. Appendix C is dedicated to the determination and application of TMD parton distributions using the parton branching method. A novel method is presented in App. D for the determination of response matrices. The method is used to estimate the systematic uncertainty due to the resolution of the observable that is discussed in Chapter 7. Appendices E, F, and G are dedicated to the phenomenological investigations on the proper merging of LO ME multiplicities with PS, as well as to the matching of ISR to NLO ME where the initial-state is evolved according to TMD parton distributions. These studies complement the phenomenological investigations discussed in Chapter 4, and are important for the interpretation of the results presented in Chapter 9. Resolution, purity, stability, acceptance, and background distributions, which supplement the studies presented in Chapter 6 on bin migration and its relation to the resolution of the observable, are reported in App. H. Similarly, the plots supplementing the studies shown in Chapter 6 on response matrices, fractional errors, and bin correlations, are presented in App. I. Finally, the tests to the unfolding of the $\Delta \phi_{12}$ distribution for inclusive 3-jet event topologies is reported in App. J.
Chapter 1. Introduction
Particle physics is the branch of physical phenomena research dedicated to the understanding of the fundamental constituents of the Universe (particles) and the interactions between them.

A particle physicist in the ancient Greece (also ancient cultures in Egypt, Babylonia, Japan, Tibet, and India), would have understood as fundamental constituents, the four elements air, water, fire and earth [7,8]. This relatively simple view of the universe radically changed as scientists became able to observe the phenomena happening at subsequently smaller distances. By the late 1860s our Standard Model could have been regarded as Mendeleev’s periodic table, carrying 56 known elements at the moment. Few decades later, by the 1920, the photon, the electron and the proton had already been discovered and promoted as the building blocks of the atomic theory.

The subject of this chapter nevertheless, is none of the aforementioned "Standard Models", but today’s Standard Model of particle physics, built up from 48 matter fields, 12 force carriers plus a Higgs boson. These sum up to 61 fundamental constituents (the chosen counting will be clarified at the end of section 2.4). It is perhaps interesting to notice the apparent randomness of the number of fundamental constituents as a function of time.
Chapter 2. The Standard Model of Particle Physics

The Standard Model of particle physics describes the electromagnetic, weak and strong forces. One could track back its early birth in the unification of the electromagnetic and weak forces by Sheldon Glashow in 1961 [9], together with the addition of the Higgs symmetry breaking mechanism [10, 11] to Glashow’s theory by Steven Weinberg and Abdus Salam in 1967 [12, 13]. Its present form was achieved after incorporating Quantum Chromodynamics (QCD) as the theory for strong interactions, once its asymptotic freedom [14, 15] was deduced in 1973.

2.1 The Standard Model group and its fundamental representation

The Standard Model of particle physics (SM) is a non-abelian gauge theory [16] obeying the following group structure:

$$SU(3) \times SU(2) \times U(1),$$

where $SU(n)$ stands for Special Unitary group of order $n$. By looking at its group structure one can build up the SM and associate it to the matter and force content we observe in the experiments.

The irreducible representation of the SM group arise from the expression 2.1 by means of the fundamental representations of the $SU(3)$, $SU(2)$ and $U(1)$ groups as shown in Tab. 2.1:

<table>
<thead>
<tr>
<th>SM irreducible representation</th>
<th>$SU(3) \times SU(2) \times U(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C^3 \otimes C^2 \otimes C$</td>
<td></td>
</tr>
<tr>
<td>$C^3 \otimes C \otimes C$</td>
<td></td>
</tr>
<tr>
<td>$C \otimes C^2 \otimes C$</td>
<td></td>
</tr>
<tr>
<td>$C \otimes C \otimes C$</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: SM irreducible representation arising from the fundamental representations of the $SU(3)$, $SU(2)$ and $U(1)$ groups.

In Tab. 2.1 the notation $C^n$ refers to a representation spanning a vector space of dimension $n$. For instance $C^3$ could only be spanned by the fundamental representation of $SU(3)$, whereas $C$ can be spanned by the trivial representations of $SU(3)$, $SU(2)$ and $U(1)$.

We have experimental evidence that the quarks have a color degree of freedom [17] that runs from 1 to 3, so they are triplets under $SU(3)$. We also learnt from the experiments that only the left-handed fermions take part in weak interactions [18] meaning that they are doublets under $SU(2)$, whereas the right-handed fermions correspond to singlets under $SU(2)$ transformations. As a result we could associate the first and second representations in Tab. 2.1 with the left- and right-handed quark content of the universe respectively.

Similarly, there is evidence that the leptons do not participate in strong interactions, indicating that they are singlets under $SU(3)$ (span the trivial representation of $SU(3)$).
Additionally, only the left-handed leptons undergo weak interactions meaning that they are doublets under $SU(2)$, whereas on the other hand the right-handed leptons are singlets under $SU(2)$. As a consequence, the third and fourth representations in Tab. 2.1 can be associated with the left- and right-handed lepton content respectively.

We have also evidence that the SM particles span an extra one dimensional space (last factor in the representations depicted in Tab. 2.1), which is indicated by the conservation of electric charge and weak isospin in electroweak interactions.

Interestingly enough the conservation of electric charge $Q$ and weak isospin $I$ in electroweak interactions does not imply a conservation of these quantities for the particles undergoing the interaction. One can imagine a weak decay of a nucleus, for which a down-quark transforms into an up-quark. In such a case there is clearly a (+1) change in charge, but also a (+1) change in weak isospin. What is conserved is then the difference between the two quantities, and it is proportional to a quantum number denoted as the hypercharge ($Y$):

$$Q - I = c \cdot Y,$$

where clearly $c \neq 0$ since for example, a left-handed electron has an integer electric charge and a half-integer weak isospin. The Eq. 2.2 implies that $Y$ is arbitrary in the SM, a freedom that in theories beyond the SM (BSM) has been successfully eliminated [19–21]. When $c = 1/2$ one recovers the Gell-Mann-Nishijima formula [22, 23]. In the following the convention $c = 1/2$ will be used, which leads to:

$$Q - I = \frac{Y}{2}$$

When we unfold the representations in Tab. 2.1 by adding the hypercharge quantum number we get the full irreducible representation of the SM:

<table>
<thead>
<tr>
<th>matter fields</th>
<th>SM representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_L$</td>
<td>$\mathbb{C}^3 \otimes \mathbb{C}^2 \otimes \mathbb{C}_{1/3}$</td>
</tr>
<tr>
<td>$u_R$</td>
<td>$\mathbb{C}^3 \otimes \mathbb{C} \otimes \mathbb{C}_{4/3}$</td>
</tr>
<tr>
<td>$d_R$</td>
<td>$\mathbb{C}^3 \otimes \mathbb{C} \otimes \mathbb{C}_{-2/3}$</td>
</tr>
<tr>
<td>$L$</td>
<td>$\mathbb{C} \otimes \mathbb{C}^2 \otimes \mathbb{C}_{-1}$</td>
</tr>
<tr>
<td>$\nu_R$</td>
<td>$\mathbb{C} \otimes \mathbb{C} \otimes \mathbb{C}_0$</td>
</tr>
<tr>
<td>$e^-_R$</td>
<td>$\mathbb{C} \otimes \mathbb{C} \otimes \mathbb{C}_{-2}$</td>
</tr>
</tbody>
</table>

Table 2.2: Matter content of the SM and the corresponding SM irreducible representation.

where the hypercharge quantum number values are indicated by the $U(1)$ sub-indices in Tab. 2.2. On the one hand the left-handed quarks are represented by $Q_L$, whereas the left-handed leptons are represented by $L$. On the other hand, the up and down right-handed quark families are represented by $u_R, d_R$, whereas $\nu_R$ and $e^-_R$ denote the right-handed neutrinos and the negative charged right-handed leptons respectively.

2.1. The Standard Model group and its fundamental representation
In order to make explicit the sets of fields which the SM group acts on, the expression 2.1 can be rewritten as:

\[ SU(3)_c \times SU(2)_L \times U(1)_Y, \]  

(2.4)

where \( c \) refers to color, \( L \) refers to left-handed and \( Y \) to hypercharge. In the following section the fundamental force fields will be addressed.

### 2.2 Gauging the Standard Model

Countless literature exists on the theoretical and experimental foundations of the Standard Model of particle physics, i.e \[24–26\]. Here are only discussed the key features, which are important for the work presented in this thesis, are discussed.

In the previous section the SM group structure and its fundamental representation, in which the matter content lives were introduced. However, we know that more particles exist and they should be taken care of in a consistent theory.

One more ingredient is needed in order to introduce the existence of the force carriers, and that is the principle of local symmetry \[27, 28\].

The fact that the matter fields are spin-1/2 fermion fields implies that they obey the Dirac equation, hence a corresponding Dirac kinetic Lagrangian density (\( \mathcal{L}_K \)) can be constructed:

\[
\mathcal{L}_K = \sum_{\Psi=Q_L,u_R,d_R,L,e_R,\nu_R} \bar{\Psi} i\gamma^\mu \partial_\mu \Psi + h.c, 
\]  

(2.5)

where \( \Psi \) runs over the representations listed in Tab. 2.2. The additional term \( h.c \) stands for hermitian conjugate, and it is needed to ensure a hermitian lagrangian density. The \( \gamma^\mu \)'s correspond to the Dirac matrices. The sum runs over all the matter fields depicted in Tab. 2.2.

The lagrangian density in the Eq. 2.5 is not invariant under local \( SU(3)_c, SU(2)_L \) and \( U(1)_Y \) transformations of the fields. To compensate for this it is enough to add new degrees of freedom that when transformed accordingly, compensate for the local change of phase of the matter fields [26]. In this way one achieves the invariance of Eq. 2.5 under local transformations of the fields, also called gauge invariance. The gauge lagrangian density \( \mathcal{L}_G \) can be written as:

\[
\mathcal{L}_G = \sum_{\Psi=Q_L,u_R,d_R,L,e_R,\nu_R} \bar{\Psi} i\gamma^\mu D_\mu \Psi + h.c, 
\]  

(2.6)

where the factor \( D_\mu = \partial_\mu - ig_1 \frac{Y}{2} B_\mu - ig_2 \frac{\tau^a}{2} W_\mu^a - ig_3 \frac{\lambda^a}{2} G_\mu^a \) and is known as the covariant derivative. The \( \tau^a \)'s are the three Pauli matrices, generators of the \( SU(2) \) group, whereas the \( \lambda^a \) are the eight Gell-Mann matrices, generators of the \( SU(3) \) color group. The coupling strength of the force fields to the matter fields are denoted as \( g_1, g_2 \) and \( g_3 \).

The field \( B_\mu \), the three fields \( W_\mu^a \) and the eight fields \( G_\mu^a \) compensate then for the local \( U(1)_Y, SU(2)_L \) and \( SU(3)_c \) transformations respectively of the matter fields. These new degrees of freedom constitute the fundamental force fields and can be related to the four electroweak bosons (\( \gamma, W^\pm \) and \( Z \) bosons) and to the eight gluons of the strong sector.

One should remember that the right-handed neutrinos are \( SU(3)_c \) and \( SU(2)_L \) singlets,
and have zero hypercharge (see Tab. 2.2), therefore they do not take part in the interactions given by SM local group structure (gauge interactions). Nevertheless, these neutrinos are not totally sterile due to the non-zero neutrino masses (Ref. [29]), so in principle they could interact with the SM particles through their mass term.

For consistency we can add the kinetic terms for the new degrees of freedom to the lagrangian in Eq. 2.6:

$$L_G = \sum_{\Psi=Q_L,u_R,d_R,L,e_R,\nu_R} \bar{\Psi}i\gamma^\mu D_\mu \Psi + h.c - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}, \quad (2.7)$$

where $B_{\mu\nu}$, $W^a_{\mu\nu}$, and $G^a_{\mu\nu}$ are the field strength tensors corresponding to $B_\mu$, $W^a_\mu$ and $G^a_\mu$ respectively [26]. These new terms are also invariant under local transformations of the fields.

Up to now, following rather simple concepts of Group Theory, we have been able to include the known matter and force fields (except gravity) into the same framework. The only "detail" left is that the lagrangian density in Eq. 2.6 does not admit mass terms for the force fields without violating the invariance under local transformations, and definitely not for the matter fields as the right- and left-handed components transform differently under $SU(2)$. As a consequence this theory would only describe a world of massless particles.

The next section is dedicated to the problem of having a SM theory with massive force carriers as well as massive matter fields.

2.3 Higgs mechanism

The discovery of the massive $Z$ and $W^\pm$ bosons in 1983 at the Super Proton Synchrotron at CERN [30–33] confirmed the prediction of the existence of three force carriers by Glashow, Weinberg, Salam [9,12,13], that together with the photon would be responsible for electroweak interactions.

The consistent inclusion of those in Eq. 2.7 made use of the Goldstone’s theorem [34, 35]. One could phrase the theorem as follows: for every broken generator of a continuous symmetry, there is a scalar particle (Goldstone boson). In the context of the SM we have a field potential (Higgs potential) that has a global $SU(2) \times U(1)$ symmetry and infinitely many ground states (vacuums). However, for one specific ground state like the one we live in, this symmetry is not manifest (spontaneously broken by the vacuum state). Once the symmetry is broken one can apply the aforementioned Goldstone’s theorem. This corresponds the so-called Higgs mechanism [10, 11], and it was successfully applied to the electroweak sector of the SM by Weinberg and Salam in 1967.

In the application of the Higgs mechanism, a symmetric potential with infinitely many degenerate minima, dependent on a complex scalar $SU(2)$ doublet field $H$ was introduced, along with the kinetic term for $H$:

$$L_H = |D_\mu H|^2 + V(H) \quad (2.8)$$

When the potential $V(H)$ acquires a non-zero real expectation value in one of the minima, three of the four generators of the $SU(2) \times U(1)$ symmetry are broken, leading to three Goldstone bosons and one real scalar field from the unbroken generator of the symmetry. These sums up to the four degrees of freedom contained in the original $H$ field. The new real
Chapter 2. The Standard Model of Particle Physics

scalar field is the so-called Higgs boson, consistent with the scalar particle discovered in 2012 by the CMS and ATLAS collaborations at CERN [36, 37].

From the kinetic term in Eq. 2.8, after a specific vacuum is chosen (symmetry breaking), the mass terms for the $W^\pm$ and $Z$ bosons arise. And this is nothing else than the three Goldstone bosons becoming the longitudinal mode of the previously massless $W^\pm$ and $Z$ bosons.

2.4 The Standard Model Lagrangian density

After having briefly discussed some of the ingredients of the SM of particle physics, the remaining points and the final form of the SM lagrangian density are presented in the following.

As we want the theory to be decoupled from possible new physics present at much higher scales (i.e Plank Scale), only renormalizable terms are allowed. With this in mind we can try to write down the most general SM lagrangian density in four dimensions.

Additionally, the Higgs mechanism also solves the problem of having massive matter fields in the theory. That is, as the $H$ field transforms under $SU(2)$, terms of the form $(\Psi_L H \Psi_R)$ are allowed by the symmetry of the lagrangian density and can be introduced, for instance:

$$L_Y = \cdots - y_u \overline{Q} L H u_R + \cdots$$

(2.9)

The $y$ are matrices in the generation space and are called Yukawa matrices. These terms contain the mixing between families for the case of the quarks, the interaction of the matter fields with the Higgs, and the masses of the matter fields (after symmetry breaking). They also contain a CP violating phase for the quark sector.

The SM lagrangian density can then be written taking into account the terms discussed so far:

$$L = \sum_{\Psi=Q_L, u_R, d_R, L, e_R} \overline{\Psi} \gamma ^\mu D^\mu \Psi + h.c - \frac{1}{4} B_{\mu \nu} B^{\mu \nu} - \frac{1}{4} W^a_{\mu \nu} W^{a \mu \nu} - \frac{1}{4} G^a_{\mu \nu} G^{a \mu \nu} + \mathcal{L}_H + \mathcal{L}_Y$$

(2.10)

Since there is evidence that the neutrinos are massive (Ref. [29]), Eq. 2.10 should be extended with a neutrino sector. We could also add to Eq. 2.10 the CP violating term $\propto \theta _{\mu \nu \alpha \beta} G^a_{\mu \nu} G^a_{\alpha \beta}$ in the strong sector. However these inclusions would not change the general picture.

Finally, we can count the matter and force content of the SM. This counting depends on how degenerated one considers these fields. Here the matter content is defined such that it follows closely the SM representation introduced in section 2.1:

$$N = 3 \sum_{\Psi=Q_L, u_R, d_R, L, e_R, \nu_R} N^c_\Psi N^I_\Psi = 48,$$

(2.11)

where 3 is the number of matter fields families, $N^c$ is the number of colors and $N^I$ is the isospin multiplicity $(2I + 1)$. As an example, the counting for the left-handed leptons would be: three (families) times one (color) times two (isospin) equals six. In addition to the 48 matter states there are 12 force fields (8 gluons and 4 electroweak bosons). This gives us a
2.4. The Standard Model Lagrangian density

A total of 61 particles when the Higgs boson is taken into account (after symmetry breaking).

2.4.1 QCD lagrangian density

QCD is the theory of strong interactions. It is governed by the $SU(3)$ sector of the lagrangian density expressed in Eq. 2.10:

$$L_{\text{QCD}} = \sum_{\Psi = Q, u_R, d_R} \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi + h.c - \frac{1}{4} G_{\mu\nu}^a G^{\mu\nu a} \quad (2.12)$$

Recalling that $D_\mu = \partial_\mu - ig_3(\lambda^a/2)G^a_\mu$, and remembering the expression for the field strength tensor $G_{\mu\nu}^a = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_3 f^{abc} G^b_\mu G^c_\nu$, one can distinguish between the mass terms, terms only involving the derivatives of the fields (propagation in space-time), as well as a group of terms involving the powers of the fields and the dimensionless coupling $g_3$ (interaction terms).

Depending on the size of $g_3$ one can identify two regimes, the perturbative regime for small values of $g_3$ and the non-perturbative regime for high values of it. The thesis will be mainly concerned with the perturbative regime of QCD.
As introduced in the last chapter, $g_3$ governs the strength of the interaction in the strong sector. From $g_3$ one can define the strong coupling $\alpha_s = g_3^2/4\pi$.

In the perturbative regime, as a result of removing ultraviolet divergences (renormalization), the value of $\alpha_s$ becomes dependent on the energy scale at which the interaction occurs.
The dependence of $\alpha_s$ on the renormalization scale ($\mu_r$) is determined by the renormalization group equations and can be expressed as an expansion in $\alpha_s$ (for an overview see Refs. [38,39]):

$$\mu_r^2 \frac{\partial \alpha_s(\mu_r^2)}{\partial \mu_r^2} = -\frac{\beta_0}{4\pi} \alpha_s^2 + O(\alpha_s^3),$$

(3.1)

where the coefficient $\beta_0 = 11 - \frac{2}{3} n_f$ with $n_f$ corresponding to the number of active flavours at the scale $\mu_r^2$. This implies that $\alpha_s$ decreases with the increase of the scale at which it is evaluated (asymptotic freedom), meaning that a perturbative expansion in $\alpha_s$ is reliable at large energy scales. Conversely, a decrease in the energy scale provokes the increase of $\alpha_s$ and with it the perturbation series becomes unreliable.

The scale at which $\alpha_s$ diverges is called $\Lambda_{\text{QCD}} \sim 250$ MeV (for an overview see Refs. [38,39]). This scale is known as the Landau pole of QCD. One can express the solution of Eq. 3.1 in terms of $\Lambda_{\text{QCD}}$ as:

$$\alpha_s(\mu_r^2) = \frac{1}{\beta_0 \ln \frac{\mu_r^2}{\Lambda_{\text{QCD}}^2}} \frac{\alpha_s^2}{4\pi}$$

(3.2)

The functional form of $\alpha_s(\mu_r^2)$ in Eq. 3.2 implies for instance that in the $\overline{\text{MS}}$ renormalization scheme (refer to Ref. [40] for an overview): $\alpha_s ((1 \text{ GeV})^2) \approx 0.5$, $\alpha_s ((10 \text{ GeV})^2) \approx 0.2$, $\alpha_s ((100 \text{ GeV})^2) \approx 0.1$. The previously mentioned values depend, specifically at small scales, on the renormalization scheme used to remove the ultraviolet divergences (see Ref. [40]).

### 3.1 A colliding high-energetic proton

Protons are bound systems of quarks and gluons, which are also-called generically partons. As a result, the interactions involving protons, and the proton system itself is governed by the QCD lagrangian density in Eq. 2.12. Protons are singlets under $SU(3)_c$ transformations and are therefore referred to as being "colorless". In addition, the proton mass being $m_P \sim 1$ GeV, implies that the proton dynamics at these scales lies in the non-perturbative region of QCD as $\alpha_s$ reaches high values, as discussed in previous sections.

As long as $m_P$ is the only scale present, a perturbative solution is not reliable. However, a second scale can emerge as the proton collides with another energetic system, an electron for instance. If a scale $Q$ is reached, such that $Q \gg m_P$ then, according to the Heisenberg principle $\tau_P \gg \tau_Q$, where $\tau_P$ is the characteristic time of the proton dynamics and $\tau_Q$ the time scale of the interaction of the proton with the electron. This means that the proton undergoes the interaction as a frozen system. In analogy with a camera one can think that the proton picture we obtain when we open our camera lens over times of order $\tau_P$ can be very different from the one we see over times of order $\tau_Q$, since the first is an average over the latter. In a collision in which $\tau_P \gg \tau_Q$ the proton can be described by its partonic structure, with such structure depending then on $\tau_Q$ itself.

In a high-energetic collision involving a proton at least two scales are involved, the proton scale $m_P$ and the highest probed scale $Q$. If $Q \gg m_P$ then the calculation of the process cross section involves a non-perturbative as well as a perturbative component. The separation of those components and the construction of a measurable cross section will be the subject of
3.1. A colliding high-energetic proton

the next sections.

3.1.1 Collinear factorization

Factorization comprises the separation, in perturbation theory, of the dynamics happening at times $\sim \tau_P$ from that happening at much smaller time scales $\tau_Q$. In collinear factorization the collinear divergencies in the initial state are factorized into process independent parton distribution functions. For cases in which a colorless final state is present, like Drell-Yan (DY) and Deep Inelastic Scattering (DIS), factorization has been proven (see Ref. [41] for details). On the other hand, for processes with colorful initial and final states, a factorization proof has not been yet achieved due to the highly non-trivial structure caused by the interference between initial and final-state soft radiation.

As will be discussed in the following sections, the present thesis is focused on observables in a region of phase space which is highly sensitive to soft radiation. The processes involved are purely QCD originated, meaning that both the initial and final states are colorful. As a result the present work can be regarded as a first step towards a dedicated study on non-trivial interference of soft initial and final-state radiation (Ref. [42, 43]).

Nevertheless, the factorization in purely QCD processes is assumed in order to obtain meaningful predictions that can be then compared to the experiment, and perhaps serve as a support or a challenge to the factorization assumption.

The dynamics happening at very long time scales ($\sim \tau_P$) is encoded in non-perturbative objects, the so-called collinear parton distribution functions (PDFs). On the other hand, the short time scale ($\sim \tau_Q$) dynamics of the interaction is described in perturbation theory. In a proton-proton collision factorization can be realized using the following expression for the production cross section of a two-parton final-state $P_1 P_2 \rightarrow cd$:

$$
\sigma_{P_1 P_2 \rightarrow cd} = \sum_{ab} \int_{0}^{1} dx_a \int_{0}^{1} dx_b f_{a/P_1}(x_a, \mu_f) f_{b/P_2}(x_b, \mu_f) \hat{\sigma}_{ab \rightarrow cd}(\mu_f, \alpha_s),
$$

(3.3)

where $x_a, x_b$ are the longitudinal momentum fractions of the respective proton momentum. At leading order, the expression can be interpreted as follows: we can calculate the cross section for the production of two partons $c$ and $d$ by convoluting the partonic cross section $\hat{\sigma}_{ab \rightarrow cd}$ with the probability density of finding a parton $a$ in proton $P_1$ ($f_{a/P_1}$) and a parton $b$ in proton $P_2$ ($f_{b/P_2}$). The factorization scale $\mu_f$ arises as a result of absorbing the collinear singularities of the initial-state partons into a scale dependent PDF. Similarly, the ultraviolet divergencies are absorbed into the implicit dependence of $\alpha_s$ on the renormalization scale $\mu_r$.

The PDFs have non-perturbative initial conditions that have to be determined experimentally. Nevertheless, attempts are being made in order to calculate PDFs from first principles by means of lattice QCD (Ref. [44]). However, the computations are quite cumbersome and not yet competitive with PDFs extracted from measurements. The factorization formulas imply that the PDFs are universal objects, meaning that once obtained they can be used for different processes.
3.1.2 Transverse momentum dependent PDFs and $k_t$ factorization

There are situations where the transverse momentum of the partons inside the protons cannot be neglected, for instance a proton-proton collision happening at small longitudinal momentum fractions $x$ or a collision in which the momentum transfer is small. Under such circumstances the validity of the factorization formula in Eq. 3.3 gets compromised. In that case a generalized factorization formula ansatz requires the dependence on the transverse momentum of the incoming partons:

$$
\sigma_{P_1 P_2 \rightarrow cd} = \sum_{ab} \int_0^1 dx_a \int_0^1 dx_b \int d^2k_{t1} \int d^2k_{t2} A_{a/P_1}(x_a, k_{t1}; \mu_f) A_{b/P_2}(x_b, k_{t2}; \mu_f) \hat{\sigma}_{ab \rightarrow cd}(\mu_f, \alpha_s),
$$

where the partonic cross section now includes off-shell partons in the initial state. We call the functions $A_i(x, k_t)$ in Eq. 3.4 unintegrated PDFs (uPDFs) or transverse momentum dependent PDFs (TMDs). There is an increasing interest regarding the determination and application of TMDs (see overview in Ref. [45]).

The contributions from small-$x$ gluons becomes increasingly important at very high center-of-mass energy of the collisions. These contributions appear in the cross section as terms proportional to $\alpha_s^n \ln^n(1/x)$. One can observe for instance that if $\alpha_s \ln(1/x) \sim O(1)$ then $x \sim O(10^{-5})$. In this kinematic region an expansion in $\alpha_s$ is unreliable since the large factor $\ln(1/x)$ compensates the smallness of $\alpha_s$. The resummation of these low-$x$ logarithmic contributions is based on the Balitsky-Fadin-Kuraev-Lipatov (BFKL) evolution equation [46–50].

A different TMD evolution equation, valid not only at low-$x$ but also at large-$x$, was introduced in Refs. [51, 52] and will be discussed in App. C. The evolution of a TMD $A(x, k_t; Q^2)$ from a scale $Q_0^2$ to a scale $Q^2$ obeys the following equation:

$$
A_a(x, k_t; Q^2) = \Delta_a(Q^2, Q_0^2) A_a(x, k_t; Q_0^2) \nonumber
\sum_b \int_{Q_0^2}^{Q^2} \frac{d^2Q'}{Q'^2} \int_{Q_0^2}^{\max} \frac{d^2P}{P^2} \Delta_a(Q^2, Q_0^2) \Delta_b(Q^2, Q_0^2) A_b \left( \frac{x}{z}, k_t + (1-z)Q'; Q'^2 \right)
$$

The solution of Eq. 3.5 using the Parton Branching (PB) method introduced in Refs [51–53] is implemented in the xFITTER package [54]. In the App. C, the use of the PB method [51, 53] for the solution of Eq. 3.5 and for the determination of TMDs is discussed. In the specific case of TMDs determined using the PB method one recovers the collinear PDFs by integrating over the transverse component:

$$
\int d^2k_t A_a(x, k_t)
$$

Once the TMD $A(x, k_t)$ is determined, the corresponding off-shell matrix elements $\hat{\sigma}(x, k_t)$ have to be calculated in order to use the $k_t$ factorization formula in Eq. 3.4. Off-shell matrix elements can be calculated using the KATIE package [55], and then convoluted with the TMDs available in TMDlib [56] using the CASCADE package [57, 58].
3.1. A colliding high-energetic proton

Even though the factorization in Eq. 3.4 uses off-shell matrix elements, approximate solutions using on-shell matrix elements can be also calculated (see App. C). As a result of the TMD evolution given by the PB implementation of Eq. 3.5, a $k_t$ can be added to the on-shell matrix elements.

In the parton shower implementations that will be discussed in the next sections, the on-shell, collinear parton approximation together with the energy-momentum conservation requirement result in kinematics shifts in the longitudinal momentum distributions (Ref. [59]). An important feature of the $k_t$ factorization approach discussed earlier and the PB solution of Eq. 3.5 implemented in CASCADE, is that energy-momentum in the events is conserved without further kinematic reshuffling.

3.1.3 DGLAP evolution equations

In a highly energetic proton-proton collision at least two scales $m_P \ll Q$ are present as discussed earlier. As a result, even if the PDF at the proton scale is a pure non-perturbative object, the evolution of it up to the scale $Q$ lives in the perturbative regime.

By the emission of partons the PDF can evolve from one scale to another different scale. Using the Feynman rules of QCD (see Ref. [60] for an overview), one can obtain the matrix elements governing the emission of a parton off another parton. Fig. 3.1 depicts the diagrams for these transition probabilities at leading-order in $\alpha_s$.

![Diagram](image)

Figure 3.1: Diagrams depicting the leading-order QCD emission of a parton off another parton.

The diagrams depict a parton (emitter) with four-momentum $p$ emitting a second parton (emitted) with four-momentum $q$. Defining the splitting variable $z$ as the ratio between the light-cone momenta of the emitted and emitter partons, the leading-order regularized splitting kernels for the processes in the diagrams above result in (see Refs. [38, 39] for an overview):
where $C_F$, $C_A$ and $T_R$ are color factors contributing to the coupling strength between partons, and implying for instance that gluons radiate more than quarks ($C_A > C_F$). The "+" sign indicates that the plus prescription of the function between the corresponding brackets is taken (see overview in Ref. [38]). This prescription regulates the behaviour at $z = 1$ where the virtual corrections live, and therefore ensures the proper cancellation of virtual and real contributions once the integration over $z$ is performed, for instance $\int_0^1 dz P^{(0)}_{qq}(z) = 0$ holds.

Using the splitting kernels, which govern the emission amplitude squared of a parton off another parton, the evolution equations for the PDFs can be expressed as (see Refs. [38, 39] for an overview):

$$\frac{\partial}{\partial \ln Q^2} \left( \frac{f_q(x, Q^2)}{f_g(x, Q^2)} \right) = \alpha_s(Q^2) \int_x^1 \frac{dx'}{x'} \left( \frac{P^{(0)}_{qq}(x)}{x} \frac{P^{(0)}_{gq}(x)}{x} \frac{P^{(0)}_{gg}(x)}{x} \right) \left( f_q(x', Q^2) f_q(x', Q^2) \right)$$

The expression Eq. 3.11 is the so-called Dokshitser-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equation [61–64]. Due to its importance its meaning will be further discussed.

What is the change in the quark content, from "looking" at the proton with a resolution $Q^2 + \Delta Q^2$ compared to its content at $Q^2$? A possible answer to this question is sketched in Fig. 3.2.

![Figure 3.2: Sketch showing the change in the quark PDF when a gluon is emitted.](image)

We have a quark content $f_q(x', Q^2)$ at the scale $Q^2$ and after resolving a gluon the quark content 'looks like' $f_q(x, Q^2 + \Delta Q^2)$. The transition is driven by the vertex whose explicit amplitude squared is $(\alpha_s(Q^2)/2\pi) P^{(0)}_{qq}(z)$.

Taking into account the previous discussion we arrive at the expression for the quark
content variation with respect to the resolution scale:

\[ Q^2 \frac{\partial f_q(x, Q^2)}{\partial Q^2} \sim \frac{\alpha_s(Q^2)}{2\pi} P_{qq}^{(0)}(z) f_q(x', Q^2) \]  

(3.12)

where \( x = zx' \) and \( 0 < z < 1 \). The \( Q^2 \) factor is needed for ensuring that the left-hand side is dimensionless. From Fig. 3.2 it is clear that \( x' > x \), and since it is a fraction of the momentum of the proton, \( x' \) can go up to 1. We can rewrite Eq. 3.12 by integrating over the allowed \( z \) and \( x' \) values:

\[ Q^2 \frac{\partial f_q(x, Q^2)}{\partial Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 dx' \int_0^1 dz P_{qq}^{(0)}(z) f_q(x', Q^2) \delta(x - zx') \]  

(3.13)

By including the analogous term coming from the splitting of a gluon into a \( q\bar{q} \) pair (Fig. 3.1), and performing the integral over \( z \) we obtain precisely one of the DGLAP equations in Eq. 3.11.

### 3.2 Monte Carlo methods and collider physics

Monte Carlo (MC) is the name given to a computational technique for solving problems based on the generation of random numbers. It has a variety of applications in high energy physics, among them the simulation of the detector response to particle passing through, the phase space integration in the matrix elements (ME) calculations, the simulation of the remaining contributions to the event like the parton shower, multi-parton interactions, and hadronization. In the following section some of the applications will be reviewed.

#### 3.2.1 Parton branching and space-like showers

The solution of the DGLAP evolution equation can be built up from a sequence of parton splittings happening at different scales. In the following the iterative solution of the DGLAP equation based on the PB method will be discussed. The discussion will follow closely the ideas presented in App. C, and in Refs. [51–53]. For simplicity one can consider a one flavor quark distribution that evolves by only emitting gluons. Additionally, only leading-order splitting functions will be used. The full flavor treatment and the use of higher-order splitting functions is discussed in the previously mentioned references. The evolution of such a quark density will be driven by Eq. 3.13. After integrating over \( x' \) and conveniently splitting the integration interval over \( z \) we obtain:

\[ Q^2 \frac{\partial f_q(x, Q^2)}{\partial Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^{1-\epsilon} \frac{dz}{z} \hat{P}_{qq}^{(0)}(z) f_q \left( \frac{x}{z}, Q^2 \right) + \frac{\alpha_s(Q^2)}{2\pi} \int_{1-\epsilon}^{1} \frac{dz}{z} P_{qq}^{(0)}(z) f_q \left( \frac{x}{z}, Q^2 \right), \]  

(3.14)

where \( \hat{P}_{qq}^{(0)} \) is the unregularized splitting function, meaning that because we are avoiding the pole \( z = 1 \) we can drop both the \( ^*+^* \) and the \( \delta \)-function in Eq. 3.10. The price to pay is that the propagation of the quark without splitting and the virtual corrections live at \( z = 1 \). As
a consequence, the unregularized splitting function is not a well-defined distribution since its integral diverges for \( \epsilon \to 0 \).

The integration variable \( z \) is close to 1 when \( \epsilon \) is small in the second term on the right-hand side of Eq. 3.14. In addition, \( \int_0^1 dz P_{qq}^{(0)}(z) = 0 \) implies that also \( \int_{z_{\text{max}}}^1 dz \hat{P}_{qq}^{(0)}(z) = -\int_0^{z_{\text{max}}} dz \hat{P}_{qq}^{(0)}(z) \) holds. As a result one can rewrite Eq. 3.14 as:

\[
Q^2 \frac{\partial f_q(x, Q^2)}{\partial Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^{z_{\text{max}}} \frac{dz}{z} \hat{P}_{qq}^{(0)}(z) f_q \left( \frac{x}{z}, Q^2 \right) - \frac{\alpha_s(Q^2)}{2\pi} f_q(x, Q^2) \int_0^{z_{\text{max}}} dz \hat{P}_{qq}^{(0)}(z), \tag{3.15}
\]

where \( z_{\text{max}} = 1 - \epsilon \). It is interesting to notice that the second term now accounts for the case in which the quark either does not split, or splits loosing a small fraction of its momentum in the process. In that case we say that a non-resolvable emission has taken place with a resolution scale \( z_{\text{max}} \). Physical observables must not depend on the choice of an arbitrary separation scale, and this can be used to test the validity of the solution.

The partial solution which is obtained by considering only the second term in Eq. 3.15 represents the evolution without branching. In such a case, the result of the integration reads:

\[
f_q^{(0)}(x, Q^2) = \Delta(Q^2, Q_0^2) f_q(x, Q_0^2),
\]

where:

\[
\Delta(Q^2, Q_0^2) \equiv \exp \left[ - \int_{Q_0^2}^{Q^2} \frac{dQ'^2}{Q'^2} \int_0^{z_{\text{max}}} dz \frac{z}{2\pi} \hat{P}_{qq}(z) f_q \left( \frac{x}{z}, Q^2 \right) \right] \tag{3.16}
\]

The term \( \Delta(Q^2, Q_0^2) \) corresponds to the so-called Sudakov form factor (see Ref. [38]). One can rewrite the evolution equation Eq. 3.15 using \( \Delta(Q^2, Q_0^2) \) as:

\[
Q^2 \frac{\partial}{\partial Q^2} \Delta(Q^2, Q_0^2) = \frac{1}{\Delta(Q^2, Q_0^2)} \int_x^{z_{\text{max}}} \frac{dz}{z} \frac{z}{2\pi} \hat{P}_{qq}(z) f_q \left( \frac{x}{z}, Q^2 \right) \tag{3.17}
\]

By integrating Eq. 3.17 we obtain:

\[
f_q(x, Q^2) = \Delta(Q^2, Q_0^2) f_q(x, Q_0^2) + \int_{Q_0^2}^{Q^2} \frac{dQ'^2}{Q'^2} \Delta(Q^2, Q_0^2) \int_x^{z_{\text{max}}} \frac{dz}{z} \frac{z}{2\pi} \hat{P}_{qq}(z) f_q \left( \frac{x}{z}, Q^2 \right) \tag{3.18}
\]

We can interpret Eq. 3.18 as follows: during the evolution from a scale \( Q_0^2 \) to \( Q^2 \) the parton will not emit any resolvable radiation with a probability \( \Delta(Q^2, Q_0^2) \) (first term on the right-hand side). The exponential nature of the Sudakov form factor implies that \( \Delta(Q^2, Q_0^2) = \Delta(Q^2, Q_0^2) \), and in turn this implies that the second term has the following interpretation: after splitting at the scale \( Q^2 \) the quark will evolve without splitting from \( Q^2 \) to \( Q^2 \).

The integral equation Eq. 3.18 has an iterative solution. The procedure is the following: Eq. 3.18 has a first approximate solution \( f_q^{(0)}(x, Q^2) = \Delta(Q^2, Q_0^2) f_q(x, Q_0^2) \) which does not involve any branching. We can continue and obtain the next approximate solution \( f_q^{(1)} \) by substituting \( f_q^{(0)} \) into Eq. 3.18. The solution \( f_q^{(1)} \) involves one branching. In the same way \( f_q^{(2)} \) can be obtained from \( f_q^{(1)} \), and it will involve two branchings.
In order to obtain the kinematical variables in each splitting one can make use of the Sudakov form factor. One generates a random $z$ of the splitting according to $P^{(0)}_{qq}(z)$. A scale $Q'^2$ is then generated according to $\Delta(Q^2, Q'^2)$, if $Q'^2 > Q^2$ then the evolution is stopped, if not, the branching is generated and the evolution continues. One proceeds and generates a new $z$ and $Q''^2$ starting from $Q'^2$, and in the same way the evolution will continue only if $Q''^2 < Q^2$. The azimuthal angle of the branchings is generated uniformly between $0$ and $2\pi$.

In this way, a sequence of additional particles to the hard process are produced, and an event with high multiplicity in the final state can be obtained. The chain of partons produced in this way is called parton shower (PS).

The mapping between the kinematical variables of the splitting and the evolution variables should be made. In the App. C a study on the physics behind different choices is presented.

Specifically, the PB method gives access to the transverse momentum $k_t$ of the subsequent splittings when solving Eq. 3.18. This can be connected to the TMDs discussed briefly in Sec. 3.1.2. Such a connection is discussed in App. C, and in Refs. [51–53].

### 3.2.2 Further comments on parton showers

The parton evolution in the previous section involved the PDFs, meaning that it is supposed to handle partons in the initial state. Parton showers of this type are therefore called initial-state radiation (ISR) and also space-like showers due to the fact that the parton which undergoes the splittings has a negative virtuality. In parton shower MC event generators like PYTHIA8, HERWIG++ and CASCADE the initial-state evolution starts at the scale of the process and evolves backwards to the proton. This is the so-called "backwards evolution" procedure (Ref. [65]).

In order to see how the backwards evolution of the partons is implemented in generators like PYTHIA8 we can start from Eq. 3.13. For simplicity one can consider a quark only emitting gluons. After integrating over $x'$ and dividing by $f_q(x, Q^2)$ the resulting expression is:

$$\frac{df_q(x, Q^2)}{f_q(x, Q^2)} = \frac{dQ^2}{Q^2} \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} \frac{P^{(0)}_{qq}(z)}{f_q(z, Q^2)}$$

By integrating the Eq. 3.19 between a scale $Q$ and a maximum scale $Q_{\text{max}}$ we obtain:

$$S_{\text{ISR}}(Q^2) = \frac{f_q(x, Q^2)}{f_q(x, Q_{\text{max}}^2)} = \exp \left[ -\int_{Q^2}^{Q_{\text{max}}^2} \frac{dQ'^2}{Q'^2} \frac{\alpha_s(Q'^2)}{2\pi} \int_x^1 \frac{dz}{z} \frac{P^{(0)}_{qq}(z)}{f_q(z, Q'^2)} \right]$$

The result in Eq. 3.20 corresponds to the probability of no emission above the scale $Q$. For instance if $Q = Q_{\text{max}}$ then the probability of no emission is 1 as one would expect. The expression in Eq. 3.20 corresponds to the Sudakov form factor $S_{\text{ISR}}$ which is used for ISR. One starts from a given scale $Q_{\text{max}}$ and evolves backwards to lower values according to the Sudakov factor.

The resulting amplitude squared of a quark in the initial state undergoing a splitting at a scale $Q$ would be weighted by the corresponding probability that no emission occurred above
Chapter 3. QCD and collider physics

this scale $S_{\text{ISR}}(Q^2)$:

$$dP = \frac{\alpha_s(Q^2)}{2\pi} \frac{dQ^2}{Q^2} P_{qq}^{(0)}(z) dz \times S_{\text{ISR}}(Q^2)$$

(3.21)

If the parton undergoing the splitting corresponds to a final-state parton then an analogous relation for the probability of the emission to occur can be written as:

$$dP = \frac{\alpha_s(Q^2)}{2\pi} \frac{dQ^2}{Q^2} P_{qq}^{(0)}(z) dz \times S_{\text{FSR}}(Q^2)$$

(3.22)

where in this case $S_{\text{FSR}}$ is the corresponding Sudakov form factor, which unlike $S_{\text{ISR}}$, does not include the PDFs ratio weight in the exponent. A sequence of branchings of this type is called final-state radiation (FSR), or time-like parton shower due to the fact that the parton which undergoes the splittings carries a positive virtuality. The final-state parton evolution is implemented forward in time, from the hard scale down to a cut-off scale that avoids the non-perturbative region.

The previous, simplified model of quarks only emitting gluons during the evolution can be generalized to include the remaining parton flavors.

In practical implementations of the PS algorithms a proper choice of the evolution variable becomes important. The evolution variable choices that are commonly used in modern MC generators are the transverse momentum ($p_t$-ordered PS) and the angle $\theta$ (angular-ordered PS) of the branching. Both choices reproduce the soft and collinear limit, with the main difference arising in the region of phase space away from this limit, where the coherence of the emissions have to be handled. The angular-ordered PS are intrinsically coherent, whereas in $p_t$-ordered showers the coherence can be enforced by a veto on emissions (Ref. [66]).

The $1 \rightarrow 2$ splitting realization for PS is nowadays complemented with the dipole-antenna picture in modern MC generators, in which two color-connected partons are considered as the source of a $2 \rightarrow 3$ splitting. This approach achieves color coherence similarly to angular-ordered showers. As an example, we will observe that there are three color dipoles in the leading-color term in Eq. 3.25 that can be used to generate further radiation. In Sec. 3.4 some of the MC generators that use this approach will be mentioned.

Various ways of improving the accuracy of a PS exist. One possibility is to correct the shower emissions with matrix elements calculations (discussed in Sec. 3.3). An additional option is to use higher-order splitting kernels in the evolution, which is, for instance implemented in the DIRE [67, 68] shower.

Various sources of uncertainties that stem from a PS algorithm are present (see Ref. [69]) like the sub-leading perturbation terms in the splitting functions, the recoil scheme, the parameters entering the shower algorithm (for example the cut-off scale at which the PS is truncated). In Sec. 8.4 the perturbative source of PS uncertainty will be discussed.

3.2.3 Color flow and large $N_c$ limit

Up to now the only clear connection to the non-abelian nature of QCD has been the triple gluon diagram shown in Fig. 3.1. The other important difference comes from the color degree of freedom. For instance 3.23 depicts a Feynman diagram representing a simple tree level
amplitude $\mathcal{A}$,

$$\mathcal{A} \sim$$

where $a, b = 1..8$ are color indices for the gluons and $i, j, k, l, m, n, r = 1..3$ are color indices for the quarks and anti-quarks. By only looking at the terms governing the possible color configurations of the amplitude we obtain a factor $T_{ij}^a T_{kl}^b T_{ln}^c T_{nr}^d$ (one generator matrix $T$ per vertex). By means of the Fierz identity $T_{ij}^a T_{kl}^b = \frac{1}{2} (\delta_{il} \delta_{jk} - \frac{1}{N_c} \delta_{ij} \delta_{kl})$ we obtain for $\mathcal{A}$:

$$\mathcal{A} \sim \delta_{il} \delta_{jk} \delta_{ln} \delta_{nr} - \frac{1}{N_c} (\delta_{ij} \delta_{kl} \delta_{mn} + \delta_{il} \delta_{jk} \delta_{ln} \delta_{nr}) + \frac{1}{N_c^2} \delta_{ij} \delta_{kl} \delta_{mn} \delta_{nr}$$

It is convenient to interpret Eq 3.24 in terms of the different color flows in Eq 3.24. The $\delta$-functions express which color indices get connected, so that one could write:

$$\mathcal{A} \sim \frac{1}{N_c} \left( -\sum_{\text{color flows}} + \sum_{\text{color flows}} \right) + \frac{1}{N_c^2}$$

Equation 3.25

One should notice that in Eq. 3.25 the first term is the only one that is not suppressed by a power of $1/N_c$. Whenever a gluon acts like a singlet (dashed line), a suppression factor $1/N_c$ arises. The first term is called the leading color approximation. When only terms of this type are considered we are in the so-called large $N_c$ limit. This approximation simplifies considerably the calculation. In this limit a gluon is basically replaced with a quark-antiquark pair regarding the color indices. Parton shower MC event generators like PYTHIA8 and HERWIG++ make use of this approximation, although efforts towards PS generators which take into account sub-leading color contributions are being made (Ref. [70]).

### 3.2.4 Fixed-order QCD

One can extend the cross section in Eq. 3.3 to the case of an $n$-parton final-state:

$$\sigma_{P_1 P_2 \rightarrow n} = \sum_{ab} \int_0^1 \int_0^1 dx_a dx_b f_{a/P_1} \left( x_a, \mu_t \right) f_{b/P_2} \left( x_b, \mu_t \right) \tilde{\sigma}_{ab \rightarrow n} \left( \mu_t, \alpha_s(\mu_t) \right)$$

Equation 3.26

In an explicit calculation of the partonic cross section $\tilde{\sigma}_{ab \rightarrow n} (\mu_t, \mu_t)$, one has to integrate over the available phase space, the degrees of freedom in the final state. Let’s suppose also
Chapter 3. QCD and collider physics

that the observable we are interested in is differential in some of these degrees of freedom $F_O$:

$$\frac{d\sigma_{P_1P_2\to n}}{dF_O} = \sum_{ab} \int_0^1 dx_a dx_b \int d\Phi_n f_{a/P_1}(x_a, \mu_f) f_{b/P_2}(x_b, \mu_f) \frac{d\tilde{\sigma}_{ab\to n}(\Phi_n; \mu_r, \alpha_s(\mu_r))}{d\Phi_n} O(\Phi_n),$$

(3.27)

where $\Phi_n$ stands for the $n$-body phase space and $O(\Phi_n)$ denotes the function projecting out the observable $F_O$. For instance, if we were interested in the $\phi_1$ distribution of the first outgoing parton then $O(\Phi_n)$ would be simply a $\delta$-function that removes the phase space integration in that specific variable.

The coupling $\alpha_s$ is evaluated at the renormalization scale $\mu_r$. If $\mu_r \gg m_P$ the process lives in the perturbative regime of QCD as discussed earlier and this implies that $d\tilde{\sigma}$ can be expanded in powers of $\alpha_s$:

$$d\tilde{\sigma}_{ab\to n} = \alpha_s^0 (d\tilde{\sigma}^{(0)} + \alpha_s d\tilde{\sigma}^{(1)} + \alpha_s^2 d\tilde{\sigma}^{(2)} + \ldots),$$

(3.28)

where $\sigma^{(0)}$ is the lowest order cross section corresponding to the process of interest. Depending on the order to which the expansion in 3.28 is performed we refer to a calculation at "leading order" (LO), "next-to-leading order" (NLO), "next-to-next-to-leading order" (NNLO) and so on.

Two main issues are present when calculating a cross section of the type of Eq. 3.27. One of these issues is that one has to calculate the squared of a sum of Feynman partial amplitudes, which can become a difficult task in the case of a high number of partial amplitudes. Nowadays there are fast methods to do so like the so-called helicity amplitudes method (see Ref. [71] for an overview), and implementations of it like HELAS (Ref. [72]). The basic idea is to perform the helicity decomposition of the amplitude and associate a complex number to each resulting partial amplitude. These numbers can then be added up. Also an efficient reuse of already calculated sub-graphs helps to increase the speed of the calculation.

The second issue is that one has to deal with the phase space integrals. For this, Monte Carlo techniques are used. The basic idea is the following: a sampling over the phase space is performed and the cross section differential in the phase space is simply obtained as a multidimensional histogram where each event is weighted according to the corresponding partonic Feynman amplitude squared and a flux factor. Furthermore, one can determine any related observable since applying the function $O(\Phi_n)$ translates to projecting out from the multidimensional histogram.

There are technical issues in performing the phase space integrals. We could take for instance the expansion in Eq. 3.28 up to the second term. Its integral over the phase space would be:

$$\sigma_{ab\to n}^{NLO} = \int (d\tilde{\sigma}^{(0)} + \alpha_s d\tilde{\sigma}^{(1)}) = \int d\Phi_n [\mathcal{B}(\Phi_n) + \mathcal{V}(\Phi_n)] + \int d\Phi_{n+1} \mathcal{R}(\Phi_{n+1}),$$

(3.29)

where $\mathcal{B}(\Phi_n)$ stands for the Born amplitude squared and $\mathcal{V}(\Phi_n)$ constitutes the virtual correction. Both live in the $n$-body phase space. The term $\mathcal{R}(\Phi_{n+1})$ represents the real correction, therefore living in a $(n+1)$-body phase space. The virtual correction $\mathcal{V}(\Phi_n)$ has a divergent behaviour when integrating over the loop momenta. There are two types of divergencies when integrating over the loop momenta. On one hand we have the ultraviolet divergencies,
3.2. Monte Carlo methods and collider physics

which are dealt with by renormalizing the coupling $\alpha_s$, and on the other hand there are infrared divergencies in the case of the loop momenta approaching zero. The latter kind of divergencies can only be cancelled against the also divergent piece $\mathcal{R}(\Phi_{n+1})$ after integrating over the momentum of the extra parton in $\mathcal{R}(\Phi_{n+1})$. This means that, although the sum is finite, the two integrals in Eq. 3.29 are separately divergent leading to a non-straightforward MC implementation of the solution. There are methods to circumvent this problem. One of those is the "slicing method" in which the singular phase space region is separated from the rest by slicing the phase space. The cross section can then be calculated in the regular region, and the divergent parts will cancel in the isolated region with a finite remaining that depends on the cut. Another method is the so-called "subtraction method". The basic idea is to add and subtract a counter term which exactly reproduces the singularities in the cross section but which is analytically integrable. This method is implemented for instance in the NLOJET++ [73] program for calculating jet production cross sections at NLO.

3.2.5 Confinement and hadronization

In this section, the concept of "confinement" and the related hadronization phase of QCD is discussed. Specific models implemented for hadronization in modern MC generators will be commented at the end of the section.

At the end of Sec. 2.4.1 the two different regimes in which the QCD lagrangian density lives were mentioned. One should remember that this separation is arbitrary and it is only linked to our ability to perform perturbative calculations. In other words, we know how to systematically obtain solutions as an expansion in a convergent series. The parameter that is used in the perturbative expansion is the strong coupling $\alpha_s$. However $\alpha_s$ is not constantly small but rather depends on an energy scale (Eq. 3.1). In Eq. 3.2 one can notice that when $\mu \to \Lambda_{\text{QCD}}$ the coupling becomes large consequently making an expansion in terms of it unreliable.

Hadrons live at scales close to $\Lambda_{\text{QCD}}$. One can take the proton as an example. It is a singlet under $SU(3)_c$ transformations despite the fact that the available set of $SU(3)_c$ irreducible representations for three quarks are $1 \oplus 8 \oplus \bar{8} \oplus 10$. We do not know of any hadron that belongs to a tenth color multiplet for instance, in fact we do not know of any hadron which does not live in the singlet representation.

The ideas conveyed in the last two paragraphs can be used to introduce the concept of "confinement", which stands for the believe that free particles, charged under $SU(3)_c$ do not exist in nature. The confinement of partons inside hadrons occurs at hadronic scales, in the non-perturbative regime of QCD. As a consequence, when implementing ISR and FSR following the evolution equations described in previous sections, a cut-off scale of order $\sim 1$ GeV is set, such that the non-perturbative region is avoided. How the mapping between the partons and the experimentally observed colorless hadrons is done relies on models. One of the available models is the so-called "string model" (Ref. [74]) that is implemented in PYTHIA, in which a color string connecting two opposite color charges is assumed. The potential energy stored in the string is considered to grow linearly for large values of the separation $r$ between the charges ($1/r \sim 1$ GeV), a behaviour which is supported by quenched lattice QCD calculations (Ref. [75]). When the color charges move apart the energy stored in the string grows with the separation until a real color-anticolor pair is created by quantum tunnelling out of the string potential energy. This results in the break of the string into
smaller pieces. The process is sketched in Fig 3.3.

![Figure 3.3: Picture of the motion of a $q\bar{q}$ system (taken from Ref. [76]). The string is represented by the red links between the color charges.](image)

In the Lund model, the probability density that governs the fragmentation of a string of transverse mass $m_\perp$ has the following form:

$$f(z) \propto \frac{(1 - z)^a}{z} \exp\left(-\frac{b m_\perp^2}{z}\right),$$

(3.30)

where $z$ represents the fraction of the momentum of the original string taken by the newly created piece. This form of the fragmentation function suppresses small $z$ values for large $m_\perp$ through the parameter $b$, which implies a harder hadron energy spectrum as $b$ grows. The expression Eq. 3.30 also suppresses the production of heavy quarks compared to light quarks.

An additional hadronization model is the so-called "cluster model" (Refs. [77, 78]), which is based on the idea of color singlet clusters of partons that decay into hadrons. This model is used in the HERWIG event generator. Figure 3.4 illustrates how hadrons are mapped to the partonic content in an $e^+e^-$ collision using the cluster model.

![Figure 3.4: Sketch of partonic clusters decaying into hadrons within the cluster model (taken from Ref. [76]).](image)

In the cluster model, the gluons in the final state undergo forced branchings to $q\bar{q}$ pairs, and the resulting quarks and antiquarks are grouped into color singlet clusters. The clusters then undergo isotropic decays into pairs of hadrons, according to the corresponding density of states (phase space times spin degeneracy $(2s_1 + 1)(2s_2 + 1)$ where $s_i$ is the spin of hadron
3.2. Monte Carlo methods and collider physics

Once the hadronization is simulated by the MC event generator, the decay of the resulting unstable hadrons follows, according to the corresponding decay width and available decay channels.

3.2.6 Multiparton interactions

As discussed in previous sections the proton is a composite object, meaning that we should acknowledge the presence of spectator partons to the hard process. One could even think of further interactions from these spectator partons. Following this idea we could ask ourselves how important these contributions are, and even a more basic question, are these contributions present at all?

One can start from the formula for a single parton-parton interaction:

\[
\sigma_{\text{int}} \sim \int_{p_{\perp_{\text{min}}}} dp_{\perp}^2 \frac{d\sigma}{dp_{\perp}^2} \sim \frac{1}{p_{\perp_{\text{min}}}},
\]

where we have taken into account that the t-channel gluon exchange dominates the 2 → 2 QCD cross section. At \( p_{\perp} \sim 4 \) GeV scale, still in the perturbative regime, the partonic cross section interaction (sum over all possible \( a, b, c, d \) in Eq. 3.3) exceeds the non-diffractive proton-proton cross section \( \sigma_{\text{nd}} \) (see Ref. [39] for an overview). A simulation from Ref. [79], using the PYTHIA8 MC generator and the NNPDF2.3 QCD+QED LO PDF set with \( \alpha_s(M_Z) = 0.130 \) [80] (\( M_Z \) is the value at the Z boson mass peak) is shown in Fig. 3.5.

![Figure 3.5: Integrated interaction cross section \( \sigma_{\text{int}}(p_{\text{min}}) \) for \( SppS \) (Super Proton-Antiproton Synchrotron) at 630 GeV, Tevatron at 1.96 GeV and LHC at 13 TeV (taken from Ref. [79]). The corresponding total cross section \( \sigma_{\text{tot}} \) is represented by the horizontal dashed lines, with the non-diffractive part \( \sigma_{\text{nd}} \) of the order of 60% of \( \sigma_{\text{tot}} \). The results were obtained using the PYTHIA8 event generator, and the NNPDF2.3 QCD+QED LO PDF set with \( \alpha_s(M_Z) = 0.130 \) [80].](image)
This apparent contradiction is solved by introducing the multi-parton interaction (MPI) interpretation: the partonic cross section convoluted with the PDFs is no longer equivalent to the corresponding non-diffractive proton-proton cross section ($\sigma_{nd}$), instead, more than one partonic collisions are allowed per proton-proton collision event. The resulting average number of independent partonic interactions per proton-proton collision would result in:

$$\bar{n} = \frac{\sigma_{int}(P_{Tmin})}{\sigma_{nd}}$$

We can assume that only two interactions occur during a pp collision, known as double parton scattering (DPS). The following discussion is based mainly on the work in Ref. [81]. For DPS, two partons inside one proton interact with two partons of the other proton. We could write the DPS cross section as:

$$d\sigma^{DPS}_{ab} = \sum_{abcd} \int d^2y F_{ac}(x_1, x_2, y) F_{bd}(x'_1, x'_2, y) \hat{\sigma}_{ab}(x_1, x'_1) \hat{\sigma}_{cd}(x_2, x'_2),$$

where functions $F_{ac}(x_1, x_2, b)$ is so-called double parton distribution function (DPDF) and represents the probability of finding two partons with flavors $a$ and $c$, and corresponding longitudinal momentum fractions $x_1$ and $x_2$ (similarly for $F_{bd}(x'_1, x'_2, y)$). The partons from the same proton are separated by a transverse distance $|y|$. The cross section for each partonic interaction is represented by $\hat{\sigma}$. If the correlation between the two partons in the proton is neglected one can express the DPDFs in terms of single parton PDFs:

$$F_{ac}(x_1, x_2, y) = \int d^2b f_a(x_1, y) f_c(x_2, y + b),$$

where a possible dependence of the PDF on the transverse position of the partons has been allowed. If we then neglect the correlation between the longitudinal and the transverse directions then $f_a(x, y) = f_a(x) F(y)$, where $f_a(x)$ is the standard single parton PDF introduced in Sec. 3.1, whereas $F(y)$ represents the transverse distribution of matter in the proton (see for instance the discussion on the overlapping function in the MPI models discussed in the following subsections). The renormalization scale dependence of the partonic cross section as well as the factorization scale dependence of the PDFs are implicitly assumed. The resulting simplified DPS cross section can be then written as:

$$d\sigma^{DPS}_{ab} = \frac{1}{C} \int d^2y \left[ \int d^2b F(y) F(y + b) \right]^2 \times \sum_{ab} f_a(x_1) f_b(x'_1) \hat{\sigma}_{ab}(x_1, x'_1) \times \sum_{cd} f_c(x_2) f_d(x'_2) \hat{\sigma}_{cd}(x_2, x'_2),$$

where $C$ is a symmetry factor which is equal to 2 if the final state of the two partonic interactions are identical, and 1 otherwise. The factors in the second and third lines of Eq. 3.35 correspond to the hadronic cross sections for each process $\sigma_1$ and $\sigma_2$, differential in the corresponding longitudinal momentum fractions of the proton momentum. By a simple
3.2. Monte Carlo methods and collider physics

dimensional analysis one can notice that the integral factor on the right-hand side of Eq. 3.35 has the dimensions of \([\text{cross section}]^{-1}\). One can then rewrite Eq. 3.35 as:

\[
\frac{d\sigma^{\text{DPS}}}{dx_1dx'_1dx_2dx'_2} = \frac{1}{C} \frac{1}{\sigma_{\text{eff}}} \frac{d\sigma_1}{dx_1} \frac{d\sigma_2}{dx_2},
\]

(3.36)

where \(1/\sigma_{\text{eff}} = \int d^2y \left[ \int d^2 b F(y) F(y+b) \right]^2\).

The DPS has been measured for instance by the CMS and ATLAS collaborations in W+2jets events (Refs. [82, 83]). The measurements of \(\sigma_{\text{eff}}\) have resulted in \(\sigma_{\text{eff}} \sim 15 \text{ mb}\). The fact that this value is smaller than the predictions from the MPI models from PYTHIA8 and HERWIG++ (30-40 mb) indicates that the models might be oversimplified. The MPI models in PYTHIA and HERWIG will be discussed in the following subsections.

As can be observed in Fig. 3.5, MPI contributions are important mainly at low \(p_T\) scales (~few GeVs). In addition, MPI becomes important at high center-of-mass energies, where smaller fractions of the proton momentum are accessed. In this kinematic limit the partonic content of the proton rises, resulting in the increase of the probability of multiple interactions to occur.

**MPI model in Pythia**

The present subsection is dedicated to the discussion of the MPI modelling in PYTHIA, and it is based mainly on Refs. [76, 79, 84, 85].

The PYTHIA MPI model relies on an ordering in the hardness \((p_\perp)\) of the interactions, in a similar way to the PS formulation. The interactions are assumed to be independent of each other, with the probability for the first interaction to occur given by:

\[
\frac{dP}{dp_{\perp}} = \frac{1}{\sigma_{\text{nd}}} \frac{d\sigma}{dp_{\perp}} \exp \left( - \int_{p_{\perp}}^{\sqrt{s}/2} \frac{1}{\sigma_{\text{nd}}} \frac{d\sigma'}{dp'_{\perp}} \right),
\]

(3.37)

where \(\sqrt{s}\) is the center-of-mass of the partonic collision and a Sudakov factor has been included to ensure that no interaction at a scale harder than \(p_{\perp}\) occurs. Further interactions can take place at lower scales by iterating on Eq. 3.37. The resulting expression for the probability of an interaction to occur at a scale \(p_{\perp}\) can be then written as:

\[
\frac{dP}{dp_{\perp}} = \frac{1}{\sigma_{\text{nd}}} \frac{d\sigma}{dp_{\perp}} \exp \left( - \int_{p_{\perp}}^{p_{\perp-1}} \frac{1}{\sigma_{\text{nd}}} \frac{d\sigma'}{dp'_{\perp}} \right),
\]

(3.38)

where \(i - 1\) indicates the previous interaction. The PDFs have to be rescaled after each iteration according to the parton flavor and proton momentum fraction already used in the previous iteration. In order to avoid the divergence of the partonic cross section \(d\sigma/dp_{\perp}^2\), the PYTHIA MPI model uses the idea of color screening: individual color charges are not any longer resolved when \(1/p_{\perp} > d\), where \(d\) is the typical distance between the individual color charges in the proton (~1 GeV\(^{-1}\)). This results in a reduced effective coupling, which in PYTHIA translates into the following modified expression for the partonic cross section:

\[
\frac{d\sigma}{dp_{\perp}^2} \propto \frac{\alpha_s^2(p_{\perp}^2 + p_{\perp,0}^2)}{(p_{\perp}^2 + p_{\perp,0}^2)^2},
\]

(3.39)

where \(p_{\perp,0} \sim 1/d\) is the non-perturbative screening parameter. In PYTHIA, the following form
for dependence of $p_{\perp 0}$ on the proton-proton center-of-mass energy $E_{\mathrm{CM}}$ is assumed:

$$p_{\perp 0}(E_{\mathrm{CM}}) = p_{\perp 0}^{\text{ref}} \left( \frac{E_{\mathrm{CM}}}{E_{\text{ref}}^{\text{CM}}} \right)^{\text{pow}},$$  

(3.40)

where $E_{\text{ref}}^{\text{CM}}$ is a reference energy value, whereas the parameters $p_{\perp 0}^{\text{ref}}$ and $\text{pow}$ have to be tuned to data.

A dependence on the impact parameter $b$ of the proton-proton collision was introduced in PYTHIA. This implementation allows to better reproduce the "pedestal effect": events with a hard scale have usually more underlying-event (UE) activity, therefore more MPI. This can be understood as the collision being more central. In the PYTHIA event generator the matter density in the proton $\rho$ is assumed spherically symmetric and therefore the function characterizing the overlap $O$ between two colliding protons has the form $O \sim \int d^3x \ dt \rho(x, y, z)\rho(x, y, z - \sqrt{b^2 + t^2})$. The overlapping function can be interpreted in the rest frame of one of the protons as follows: a spherically symmetric proton at rest with density $\rho(x, y, z)$ collides with a proton moving in the $z$ direction whose boosted density is given by $\rho(x, y, z - \sqrt{b^2 + t^2})$. Two different overlapping functions are used in the PYTHIA simulations discussed in this thesis. One of the overlapping functions choice is given by the exponential:

$$O(b) = \exp \left( -b^a \right),$$  

(3.41)

When $a = 2$ the overlapping function has a Gaussian profile, whereas $a \to \infty$ would result in no impact parameter dependence. For instance the tune CUETP8M1 makes use of Eq. 3.41 as overlapping function as will be further discussed in Sec. 8.5.

The other overlapping function employed in this thesis relies on a double Gaussian distribution for the hadronic matter:

$$\rho(r) = \frac{1 - \beta}{a_1^2} \exp \left( -\frac{r^2}{a_1^2} \right) + \frac{\beta}{a_2^2} \exp \left( -\frac{r^2}{a_2^2} \right),$$  

(3.42)

where a narrower distribution of width $a_2$ which contains a fraction $\beta$ of the hadronic matter is embedded in a wider distribution of width $a_1$ containing a fraction $1 - \beta$ of the hadronic matter. The idea behind this model is that due to the valence quark evolution the hadronic matter might tend to gather around the so-called "hot spots" with the valence quarks as centers. For instance, the tunes CP1 and CP2 make use of this parametrization (as will be further discussed in Sec. 8.5).

One of the features of the PYTHIA8 event generator is that the MPI simulation is combined with the ISR and FSR in an interleaved evolution. Such evolution follows the expression:

$$\frac{dP}{dp_{\perp}} = \left( \frac{dP_{\text{MPI}}}{dp_{\perp}} + \sum \frac{dP_{\text{ISR}}}{dp_{\perp}} + \sum \frac{dP_{\text{FSR}}}{dp_{\perp}} \right) \times \exp \left( -\int_{p_{\perp}^{i-1}}^{p_{\perp}} dp'_{\perp} \left( \frac{dP_{\text{MPI}}}{dp_{\perp}} + \sum \frac{dP_{\text{ISR}}}{dp_{\perp}} + \sum \frac{dP_{\text{FSR}}}{dp_{\perp}} \right) \right),$$  

(3.43)

where the sums run over all the ISR and FSR corresponding to the specific MPI sub-collision.
The MC generation of partons using Eq. 3.43 is done using the so-called "the winner takes it all" idea. It works in the following way: a $p_{\perp}$ scale is generated by each individual term in Eq. 3.43, and the process given the highest scale is selected in the generation step.

The evolution of the MPI systems using Eq. 3.43 has a cut-off scale ($\sim 1$ GeV) that avoids the non-perturbative region. Since $1$ GeV $\approx (0.2 \text{ fm})^{-1}$ and since the MPI centers lie within the proton radius, the possibility exists that the resulting partons in a system with high MPI multiplicity overlap in physical space. This gives rise to a collective hadronisation phenomenon correlating different MPI subsystems (see Ref. [85] for an overview). In the PYTHIA8 event generator this effect is modelled by allowing the different MPI subsystems to reconnect. An MPI subsystem which is characterized by a scale $p_{\perp}$ can be reconnected with another subsystem at higher scale with a probability:

$$P = \frac{p_{\perp \text{Rec}}}{p_{\perp \text{Rec}} + p_{\perp}},$$

(3.44)

where $p_{\perp \text{Rec}} = p_{\perp 0}R$, and $p_{\perp 0}$ is the MPI dampening factor introduced in Eq. 3.39. The parameter $R$ is called reconnection range.

**MPI model in HERWIG**

The MPI solution to the problem illustrated in Fig. 3.5 is also implemented in the HERWIG event generator. The present subsection is based mainly on Ref. [86].

The MPI idea in HERWIG starts by also assuming independent partonic interactions, where the average number of interactions $\bar{n}(b; p_{T\text{min}})$ satisfies Eq. 3.32, and depends on the overlapping between the protons through the impact parameter $b$. The statistically independent scatterings follow a Poisson distribution, from which the number of interactions can be sampled once $\bar{n}(b; p_{T\text{min}})$ is determined.

The average number of interactions $\bar{n}(b; p_{T\text{min}})$ is proportional to the corresponding interaction cross section $\sigma_{\text{int}}(p_{T\text{min}})$. If one assumes that the transverse structure of the proton factorizes with the longitudinal part which enters the partonic interaction, the resulting proportionality holds:

$$\bar{n}(b; p_{T\text{min}}) = A(b; \mu)\sigma_{\text{int}}(p_{T\text{min}}),$$

(3.45)

where the proportionality factor $A(b; \mu)$ encodes the transverse structure of the overlapping protons and therefore is called overlapping function. The dependence of the number of interactions on the effective size of the proton is encoded in the parameter $\mu$, which characterizes the inverse of the proton radius. For instance, one could imagine that at a fixed $b \sim 1$ fm, two protons with twice the transverse size should overlap more.

Similarly to the PYTHIA generator, a color reconnection model is implemented in HERWIG. For this purpose, a measure of closeness between color-anticolor charge pairs is defined as follows:

$$\lambda = \sum_{\text{pairs } ij} m_{ij}^2,$$

(3.46)

where $ij$ are color-anticolor pairs that may form neutral clusters in the hadronization model, and $m_{ij}$ is the invariant mass of the pair. The implementation of the color reconnection
model in HERWIG relies on minimizing $\lambda$, which ensures that the resulting color-anticolor charge pairs will lie closely in phase space.

### 3.2.7 Jets

A QCD hard interaction can lead to a very crowded final state. One of the reasons is the evolution of the partons involved in the interaction, in both initial and final states. Additionally one has the confinement phase at hadronic scales which can lead to a rearrangement of partons to the final hadrons. Despite this, in a QCD event with high momentum transfer (high $p_T$ event), what one actually observes are well collimated groups of hadrons in the detector. These bunches are called jets and are a reflection of the hard part of the interaction. A theoretical prediction for a jet observable should be compared to the data only after all the ingredients of the interaction are modelled. As a result the definition of a jet should be the same in both the data and the theoretical prediction which are being compared.

One constructs a jet following a so-called jet algorithm in which the particles are grouped in a jet. The jet algorithm should be applicable at both parton and hadron level meaning that it should be insensitive to the details of the hadronization modelling. And it should be insensitive also to the addition of a soft parton or to the addition of a collinear parton, meaning that emissions which are not resolved should not change the number and kinematics of the jets.

In the present section we focus on sequential jet algorithms, specifically the anti-$k_T$ algorithm. The basic principle is the clustering of pairs of particles in turn, until a resulting object with the four momentum equals to the sum of the four momenta of the clustered particles is left (jet). A "distance" between a particle (object) and the beam line $d_{iB}$, and a "distance" between two particles (objects) $d_{ij}$ are introduced, and in the specific case of the anti-$k_T$ algorithm this distances are defined as:

\[
    d_{iB} = (p_{ti})^{-2}
\]

\[
    d_{ij} = \min\{(p_{ti})^{-2}, (p_{tj})^{-2}\} \frac{R_{ij}^2}{R^2},
\]

where $p_{ti}$ is the transverse momentum of the object $i$, analogously for $j$. The quantity $R_{ij}$ is the distance between the objects $i$ and $j$, in the plane spanned by the rapidity $y$ and the azimuthal angle $\phi$. It is defined as:

\[
    R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2},
\]

The decision of whether a certain particle belongs to a jet or not is somehow arbitrary and can be expressed by a parameter $R$. Let us imagine that we add a particle with small $p_{t1}$ and close in $R$ to a clustered object with high $p_{t0}$. In such a case $d_{01}$ will be small, because of the inverse of $p_{t0}^2$ and because of the distance between the objects. On the other hand $d_{1B}$ will be large meaning that the particle will be clustered and no extra jet is created. This example should give the feeling that the algorithm is infrared safe. The size of the jet in $y - \phi$ space is then driven by the parameter $R$. For instance when $R_{ij} \gtrsim R$ and the two partons have a $p_T$ of the same order they will start being observed as separate objects. The jets clustered using the anti-$k_T$ algorithm are characterized by a circular area in the $y - \phi$ plane.
Using Eqs. 3.47 and 3.48 the procedure is that two objects are clustered if \( d_{ij} < d_{iB} \). A discussion on the performance of the anti-\( k_T \) can be found in Ref. [87]. From the definition one can notice that the algorithm is based on quantities which are invariant under boosts along the beam line.

### 3.3 Matching and/or merging

In the following, a discussion based on Refs. [76, 88, 89] on the concepts of matching and merging is presented. The ideas presented in this section are of big importance for the interpretation of the results of the thesis.

In the previous sections we have discussed the notion of fixed-order calculation and the concept of PS. In this section the relation and interplay between the two will be addressed in collinear factorization.

One of the differences between fixed-order calculation and PS approaches is how exclusive the result can be. For example, for fixed-order 2 \( \rightarrow \) 2 QCD events only one scale is resolved, which can be taken as the transverse momentum of the jets. As a result this calculation can only be used for observables insensitive to other scales (inclusive observables). On the other hand, a PS generates a fully exclusive event, meaning that it provides a non-trivial prediction for observables which are sensitive to multiple scales, for instance a 2 \( \rightarrow \) 2 + PS calculation would give non-trivial predictions for the transverse momentum imbalance of two, three or four jets. Notice that for the latter observables, which involve at least two, three and four scales respectively the fixed-order 2 \( \rightarrow \) 2 calculation would give a trivial result. We call "matching" the process of combining a fixed-order calculation to a PS.

However the PS is an approximation, valid in the soft and collinear limit. Therefore we expect that the PS is not accurate when generating hard scales, for example a hard well separated third parton from a 2 \( \rightarrow \) 2 process. Instead, a fixed-order 2 \( \rightarrow \) 3 calculation is expected to describe better this topology. We call "merging" the process of combining multiple fixed-order multiplicities.

#### 3.3.1 NLO matching

The concepts of matching and merging are very related to each other and one of the goals of this section is to make this apparent. One can start from Eq. 3.29 for the case of an NLO QCD 2 \( \rightarrow \) 2 calculation:

\[
\frac{d\sigma^{NLO}_{2\rightarrow 2}}{d\Phi_B} = \mathcal{E}(\Phi_B) + \mathcal{V}(\Phi_B) + \int d\Phi_r \mathcal{R}(\Phi_B, \Phi_r) \equiv \mathcal{E}(\Phi_B),
\]  

(3.50)

where in this case \( \Phi_B \) represents the two-particle phase space, and \( \Phi_r \) is the real emission phase space spanned by the extra parton. The Born+one parton cross section is represented by \( \mathcal{R}(\Phi_B, \Phi_r) \), while \( \mathcal{V}(\Phi_B) \) corresponds to the virtual correction which lives in the Born phase space. The formula Eq. 3.50 is the two-parton cross section at NLO differential in the Born variables \( \mathcal{E}(\Phi_B) \), since we are integrating over the radiation phase space. For simplicity, in the following the dependence of the terms on the Born phase space will be omitted, and also \( \Phi_r \equiv p_T^2 \) of the radiated parton will be assumed. We can then write the
cross section for the emission of a parton according to Eq. 3.50 as:

\[
\frac{d\sigma^R}{dp_t^2} = \mathcal{R}(p_t^2)
\]

(3.51)

The expression Eq. 3.51 is divergent when the radiation becomes soft or collinear to any of the incoming or outgoing partons. In other words, when we try to have the exclusive cross section in the radiation variables we obtain unphysical results in regions of phase space in which a small scale is resolved.

On the other hand, as commented previously the PS formulation is able to resolve small scales by the subsequent emission of partons (resolvable and non-resolvable). The PS approximation for the radiation of one resolvable parton from the Born configuration is:

\[
\frac{d\sigma^{PS}}{dp_t^2} = \mathcal{R}^{PS}(p_t^2) \exp \left[ - \int dp_t^2 \frac{\mathcal{R}^{PS}(p_t^2)}{\mathcal{B}} \right],
\]

(3.52)

where the Sudakov form factor enforces no further resolvable emission above \(p_t^2\). As can be observed, Eq. 3.52 is analogous to Eqs. 3.21 and 3.22. One should notice that the LO cross section (\(\mathcal{B}\)) is recovered after integrating over the radiation phase space. This should be the case from the probabilistic interpretation of the Sudakov factor, in other words, the PS is by construction, conserving unitarity.

The matching/merging of a PS to a fixed-order ME relies on the relation between Eq. 3.51 and Eq. 3.52. It is interesting to notice that both \(\mathcal{R}\) and \(\mathcal{R}^{PS}\) are divergent in the soft and collinear limit, however in the case of Eq. 3.52 the Sudakov factor tames the divergency as it tends to zero more rapidly in this limit.

**MC@NLO method**

By expanding the Sudakov factor in Eq. 3.52 we obtain the following:

\[
\frac{d\sigma^{PS}}{dp_t^2} \approx \mathcal{R}^{PS}(p_t^2)(1 + O(\alpha_s))
\]

(3.53)

By summing and subtracting the integral of Eq. 3.53 in Eq. 3.50 the resulting expression is:

\[
\mathcal{B} = \mathcal{B} + \left[ \mathcal{V} + \int dp_t^2 \mathcal{R}^{PS}(p_t^2) \right] + \int dp_t^2 \left[ \mathcal{R}(p_t^2) - \mathcal{R}^{PS}(p_t^2) \right]
\]

(3.54)

The terms between brackets in Eq. 3.54 are now finite, since \(\mathcal{R}\) and \(\mathcal{R}^{PS}\) have the same soft and collinear limit. This allows the use of the term \(\mathcal{B} + \left[ \mathcal{V} + \int dp_t^2 \mathcal{R}^{PS}(p_t^2) \right]\), which depends only on the Born variables, to generate Born configurations via MC sampling. Similarly the finite term \(\left[ \mathcal{R}(p_t^2) - \mathcal{R}^{PS}(p_t^2) \right]\) can be used to generate Born+one parton configurations. Subsequent emissions can be then generated by the same PS that was used for obtaining the counter-terms, with the starting scale being the one set by the Born emission.

One feature of the method is that it generates both positive and negatively weighted events, since the terms containing the counter terms are not positive definite. Another feature of the method is that the counter terms have to be generated according to the specific PS that will be added afterwards. In the results section we will discuss the differences observed
when Pythia8 and Herwig++ counter terms are employed and the effect on the observables studied in the present thesis.

This is the basics of the MC@NLO method for matching fixed-order calculations to PS. For further details one can refer to Ref. [90]. It is interesting to notice that one could also call it merging since two different multiplicities are generated at NLO.

**POWHEG method**

The POWHEG method (Ref. [91]) is another method to match/merge fixed-order NLO calculations and PS. It is based on the reweighting of the real emission whose cross section was shown in Eq. 3.51. The expression in Eq. 3.51 diverges for collinear and soft configurations. However, we know that the PS approximation in Eq. 3.52 is finite due to the Sudakov factor. The POWHEG solution corresponds to define a cross section for the real emission which mimics the PS form in Eq. 3.52:

\[
\frac{d\sigma_{PH}}{dp_T^2} = \frac{B}{B} R(p_T^2) \exp \left[ - \int_{p_T^2} dp_T'^2 \frac{R(p_T'^2)}{B} \right],
\]

where the superscript PH stands for POWHEG emission, which corresponds to the real emission weighted by the Sudakov factor contained in Eq. 3.55. The expression in Eq. 3.55 is a mixture of Eq. 3.51 and Eq. 3.52. It contains a Sudakov factor, however this factor carries the real matrix elements emission kernel \( R \). The ratio between the Born matrix elements \( \frac{B}{B} \) is included to ensure that after integrating Eq. 3.55 over the radiation phase space one obtains the NLO cross section in the Born variables \( \mathcal{B} \). In addition, the inclusion of the Sudakov factor formally does not affect, at the given order and large \( p_T \), the accuracy of the cross section:

\[
R(p_T^2) \exp \left[ - \int_{p_T^2} dp_T'^2 \frac{R(p_T'^2)}{B} \right] \approx R(p_T^2)(1 + \mathcal{O}(\alpha_s))
\]

(3.56)

The expression in Eq. 3.55 can then be used to generate Born as well as Born+one parton configurations using MC methods. To see this, one can integrate over the radiation phase space, and split the integration interval using an arbitrary cut \( p_T^{2\text{min}} \):

\[
\mathcal{B} = \mathcal{B} \int_{0}^{p_T^{2\text{min}}} dp_T^2 \frac{R(p_T^2)}{B} \exp \left[ - \int_{p_T^2} dp_T'^2 \frac{R(p_T'^2)}{B} \right] + \mathcal{B} \int_{p_T^{2\text{min}}}^{p_T^2} dp_T^2 \frac{R(p_T^2)}{B} \exp \left[ - \int_{p_T^2} dp_T'^2 \frac{R(p_T'^2)}{B} \right],
\]

(3.57)

After performing the integration of the first term on the right-hand side of Eq. 3.57 we obtain:

\[
\mathcal{B} = \mathcal{B} \exp \left[ - \int_{p_T^{2\text{min}}}^{p_T^2} dp_T^2 \frac{R(p_T^2)}{B} \right] + \mathcal{B} \int_{p_T^{2\text{min}}}^{p_T^2} dp_T^2 \frac{R(p_T^2)}{B} \exp \left[ - \int_{p_T^2} dp_T'^2 \frac{R(p_T'^2)}{B} \right]
\]

(3.58)

The resulting form of Eq. 3.58 has interesting features and a beautiful interpretation from an MC point of view: 1) the first term on the right-hand side depends only on the Born variables and can be used to generate Born events below \( p_T^{2\text{min}} \), weighted by the exponential factor; 2) above \( p_T^{2\text{min}} \) the integrand of the second term can be sampled to generate Born+one
Chapter 3. QCD and collider physics

configurations weighted by the corresponding Sudakov factor; 3) both terms are positive definite, meaning that only positively weighted events are generated; 4) the expression is equal to \( \mathcal{B} \), meaning that the integrated cross section as a function of the Born variables is still accurate at NLO.

A PS can then be easily added using the scale of the POWHEG emission as the starting scale of the shower. Another feature of the method is that it is independent of the PS that would be used afterwards. As in the MC@NLO method, the matching procedure is basically a merging of two multiplicities. However in the POWHEG method there is a clear cut \( p_{t\min}^2 \) defining which multiplicity is generated, whereas in the MC@NLO, although there is a separation between multiplicities, it is not fixed in an event-per-event basis. These features will be discussed in the results section.

In the POWHEGBOX [92] implementation of the POWHEG method the arbitrary cut \( p_{t\min}^2 \) is called \( \text{ptsqmin} \). This parameter and the POWHEG method in general will be also further discussed in the results section.

### 3.3.2 Multi-jet merging

The multi-jet merging concept is very much related to the previous discussion on NLO matching. In the present sub-section only LO multi-jet-merging will be discussed. For further details and information on the multi-jet merging at higher orders refer for instance to Ref. [93].

One can start by remembering that a Born+one parton configuration diverges when the extra parton becomes soft or collinear to any other parton. Therefore one would use the Born level matched to a PS close to the singular regions and the fixed Born+one parton away from it. Let us assume that we can separate the divergent region with a cut \( q_{\text{cut}}^2 \). Then one can rely again on Eq. 3.51 and Eq. 3.52 to generate events.

If we evaluate Eq. 3.51 and Eq. 3.52 at \( q_{\text{cut}}^2 \) we see that the results are not equal, meaning that if we take one of the expressions to generate events below the cut and the other expression to generate events above the cut we end up with an unphysical discontinuity at the cut value. The merging procedures takes care of eliminating this kind of discontinuities when a PS and a fixed-order calculation are combined. The cut parameter \( q_{\text{cut}}^2 \) is the merging scale and it acts similar to \( p_{t\min}^2 \) in the POWHEG method.

One way one can reduce the difference between Eq. 3.51 and Eq. 3.52 at \( q_{\text{cut}}^2 \) is to apply the PS Sudakov to Eq. 3.51 such that:

\[
\frac{d\sigma^{R\ast}}{dp_t^2} = \mathcal{R}(p_t^2) \exp \left[ - \int_{p_t^2} \frac{dR^{PS}(p_t^2)}{\mathcal{B}} \right] \quad (3.59)
\]

The Sudakov included in Eq. 3.59 acts mainly at small \( p_t^2 \), whereas at large \( p_t^2 \) it tends to one. Another important correction that is done in the merging algorithms is to evaluate the emission vertex \( \alpha_s \) at the scale of the emission. Interesting is to notice that the integral over Eq. 3.59 is no longer the LO cross section as it should. Nevertheless, the difference lies in the next order (see discussion for Higgs production in Ref. [89]). The effects of the remaining discontinuity at the merging scale are reduced by a proper choice of this scale.

The expression in Eq. 3.59 can then be used to generate events above \( q_{\text{cut}}^2 \), whereas Eq. 3.52 can be used for generating events below \( q_{\text{cut}}^2 \). The merging procedure is illustrated in Fig. 3.6 for the simpler case of vector boson production.
3.3. Matching and/or merging

Figure 3.6: Illustration of the merging of several event multiplicities for vector boson production (taken from Ref. [94]). The red curly lines represent real emissions whereas the blue curly lines represent PS emissions.

In Fig. 3.6 $q_{\text{cut}}$ represents the merging scale whereas $k_T$ characterizes the relative transverse momentum (not only the $p_T$ of the emission but also the angular separation with respect to other emissions). Emissions from the ME are considered above $q_{\text{cut}}$ whereas the PS populates the phase space below $q_{\text{cut}}$. If the merging of the multiplicities is not properly performed the predictions could suffer from either double counting of emissions, or from gaps in the phase space which might not be filled.

There are various ways of implementing the merging algorithm in a MC generator. In the following the MLM and CKKW merging procedures will be briefly discussed.

MLM merging

The MLM [95] merging procedure implements the combination of Eq. 3.52 and Eq. 3.59 by using the Shower as the Sudakov factor. The idea of the implementation can be summarized as follows: 1) the events are showered starting from the highest scale, 2) the final-state partons are clustered into jets, 3) the partonic jets are compared to the corresponding ME level, 4) only the events that are matched to the corresponding ME level, with no extra jets above the merging scale are accepted. The details of the procedure can be found in the original paper (Ref. [95]).

CKKW merging

The CKKW [96] merging procedure implementation is based on the use of the Sudakov factor in Eq. 3.59 above the merging scale, and the PS (Eq. 3.52) below the merging scale. Above the merging scale, the events generated by the real matrix elements $\mathcal{R}$ are reweighted by the corresponding analytical version of the Sudakov factor in Eq. 3.59. The details of the
Chapter 3. QCD and collider physics

A procedure can be found in Ref. [96]. In the results section the CKKW-L variant proposed in Ref. [97] and implemented in PYTHIA8 (see Ref. [98] for an overview) is employed. The CKKW-L scheme uses the Sudakov factor from the PS by vetoing emissions above the merging scale.

3.3.3 Matrix element correction to parton shower

Additionally, efforts are being made in the opposite direction: correcting the hard limit of the PS using fixed-order calculations. Such corrections are implemented in the VINCIA [99, 100] shower, and also in the PYTHIA8 event generator.

As an example, in order to correct the first PS emission one would start generating the emission using Eq. 3.52. Subsequently, the event is reweighted with the factor $R/R^{PS}$. The idea is analogous to the one used in the POWHEG method and multi-jet merging, with the difference lying on the fact that for the latter cases one starts from the matrix elements and then reweights to the shower, whereas in the present case one starts from the shower and reweights to the matrix elements. This procedure can be iteratively applied if higher matrix elements multiplicities are available (Refs. [99, 100]).

3.4 MC event generators

MC event generators are computer programs used to simulate events in high-energy collisions between particles. An event, as addressed by a MC generator is generally divided in different steps like the ones listed below.

- Matrix element calculation
- Parton shower
- Underlying-event (MPI)
- Hadronization
- Hadron decays

The implementation in MC event generators of the points stated above relies on the validity of the factorization ansatz discussed in Sec. 3.1. There is much freedom in practical implementations of the different points in the list. MC event generators can differ for instance in the use of collinear or $k_t$ factorization, the use of on- or off-shell ME, the PS approach, the matching of the PS to higher-order ME, the MPI and hadronization modelling. And there can be differences even in the details, like the PS scheme (dipole picture or not) and the recoil strategy (conservation of energy and momentum in the PS).

Therefore many MC generators are currently available, and each of them focuses on either one or a few of the bullets listed above. In this way, the comparison to the experimental data, and the tests of the models and schemes in the different generators become very important. In the following, the various event generators used in this thesis and the differences between them are summarized below:
3.4. MC event generators

- **PYTHIA**

  The Pythia8 [101] general purpose event generator relies on collinear factorization. It focuses on the shower treatment (currently $p_t$ ordered shower based on a dipole-like recoil approach), the MPI simulation is interleaved with the PS, and the hadronization simulation is based on the Lund string model. It generates ME for basic LO processes therefore leaving the task of higher-order accuracy of the ME calculation to other generators.

- **HERWIG**

  Herwig++ [102] is a general purpose event generator whose current version is Herwig7 [102, 103]. It relies on collinear factorization. Herwig7 has not only the angular-ordered PS implemented but also a dipole shower option is available. The hadronization simulation relies on the cluster model. Several processes at LO can be generated in Herwig++, whereas Herwig7 provides ME not only at LO but also NLO accurate, with the option of using either the MC@NLO or the Powheg method to match the ME to the PS.

- **MADGRAPH**

  Specifically MadGraph5_aMC@NLO [104] is a ME generator which prioritize computations of cross sections, generation of hard events and their matching with PS generators. The PS, MPI and hadronization has to be simulated by other MC generators like Pythia8 and Herwig++. Multi-partonic final states can be generated at NLO accuracy, relying on the MC@NLO method for the matching to PS.

- **POWHEG**

  The POSitive Weight Hard Emission Generator (Powheg) [92,105,106] is a ME generator at NLO accuracy. It employs the Powheg method to match the ME to PS. The PS, MPI and hadronization have to be simulated by other MC generators like Pythia8 and Herwig++.

- **CASCADE**

  Cascade [57, 58] is a MC generator which is able to simulate events in both collinear and $k_t$-factorization. It employs TMDs, and the generation of ISR is done according to the TMD which is used. Cascade also provides the option of generating ISR using NLO splitting functions (Ref. [107]). Initial-state, on- and off-shell matrix elements can be used in Cascade in combination with the TMDs.

- **DIRE**

  Dire [67,68] (DIpole RESummation) is a PS generator which uses the dipole picture with ordering variables based on transverse momentum of the emissions in the soft limit. The generator can be used as a plugin to Pythia. It includes NLO corrections to the splitting kernels.

- **VINCIA**

  Vincia [99,100] (VIrtual Numerical Collider with Interleaved Antennae) is a PS generator which can be used as a plugin to PYTHIA. It is based on the dipole-antenna picture. The generator includes iterative LO ME corrections to the PS.
3.4.1 Tuning of the underlying-event

Tuning is the name given to the procedure of restricting the parameters of the phenomenological models in a general purpose event MC event generator. The parameters which are commonly tuned are the ones related to the underlying-event (UE) activity, hadronization, color reconnection, and the treatment of the beam remnants. Another source of non-perturbative uncertainty lies in the PS. The on-shell, collinear parton approximation, together with the energy-momentum conservation requirement result in kinematics shifts in the longitudinal momentum distributions (Ref. [59]). One should notice that an important feature of the $k_t$ factorization approach discussed in Sec. 3.1.2 and App. C is that it can reduce the effects of this source of systematic uncertainty.

The tunes are generally obtained by fitting the set of parameters to measurements which are sensitive to the UE activity and hadronization. Two of the observables which are used in order to obtain the different tunes are the charged particles multiplicity, as well as the scalar sum of the charged particles $p_T$ as a function of the leading charged particle transverse momentum.

One has to be careful to only tune the "soft" physics and not the "hard" physics which we generally consider as our signal. For instance, the tuning of hadronization should not affect the prediction for the $p_T$ of a hard jet, or the prediction for the Higgs mass in the $b\bar{b}$ decay channel.

Tunes are obtained by scanning the parameter space until the set that minimizes the $\chi^2$ difference with the data is found (Refs. [108]). The general purpose event generators used in the present thesis, which model the UE activity are PYTHIA8 and HERWIG. The predictions in the present work that make use of the PYTHIA8 event generator are simulated using the CMS tune CUETP8M1 [109], whereas the CMS tune CUETHppS1 [109] is employed for the simulations which make use of the HERWIG++ event generator. For the case of HERWIG7 the tune H7-UE-MMHT [103] was utilized.

The source of uncertainty lie in the parametrization freedom in the models. For instance, among those are the parameters governing the overlapping function in the MPI models (Eqs. 3.41, 3.42, 3.45), and the color reconnection range parameter $R$ which influences the string fragmentation process during the hadronization step (Eqs. 3.44).

The latest CMS tunes CP1 and CP2 [110] (where I was involved in their determination), as well as Monash [111], 4C [112], and the ATLAS tune A14 [113] are used to estimate the differences of using different model parametrizations in the PYTHIA event generator (Sec. 8.5).
CHAPTER 4

PHENOMENOLOGY OF THE $\Delta \phi_{12}$ DISTRIBUTION

Contents

4.1 Fixed-order calculations and parton shower applied to the $\Delta \phi_{12}$ distribution ........................................... 42
   4.1.1 Initial- and final-state shower contributions .................. 44

4.2 Impact of the matching of matrix element to parton shower .... 45
   4.2.1 MC@NLO matching ............................................. 45
   4.2.2 POWHEG matching ............................................. 46
   4.2.3 Relation between POWHEG and MC@NLO. The parameter $\text{hdamp}$ ................................................ 49
   4.2.4 LO multi-jet merging ........................................... 52

4.3 Transverse momentum dependent parton distribution applied to
   $\Delta \phi_{12}$ .......................................................... 54

QCD is able to explain a wide variety of physical phenomena in jet physics. The production of jets is intimately related to the underlying partonic interaction and the subsequent parton evolution. In modern MC generators, the production of jets is simulated with the help of parton showers. The calculations can be improved by using higher-order ME for wide angle hard radiation. As discussed in the previous sections, there are various approaches to not only model the PS but also to match it to higher ME multiplicities. The understanding of this relative freedom and the differences between the different approaches is crucial for a good interpretation of the results in a subsequent comparison to data.

A very sensitive observable to extra radiation in jets physics is the azimuthal separation between the leading jets in the event, $\Delta \phi_{12}$. This observable is sensitive not only to the hard radiation present in the event (small $\Delta \phi_{12}$), but also to the soft radiation accompanying the
hard system (large $\Delta\phi_{12}$). Therefore $\Delta\phi_{12}$ provides a test not only for the higher fixed-order calculations needed to describe small azimuthal separations, but also for the resummation of soft radiation contributions needed when the jets are nearly back-to-back.

In this section we will discuss the phenomenology of the $\Delta\phi_{12}$ distribution as predicted in modern MC generators.

4.1 Fixed-order calculations and parton shower applied to the $\Delta\phi_{12}$ distribution

The inclusive production of two high-$p_T$ jets is described at LO by QCD $2 \rightarrow 2$ partonic ME. However, $\Delta\phi_{12}$ between these jets corresponds to an exclusive calculation for which a fixed-order LO $2 \rightarrow 2$ ME results in the trivial $\Delta\phi_{12} = 180^\circ$ (from momentum conservation in the transverse plane).

On the other hand, a fixed-order $2 \rightarrow 2$ NLO calculation contains the real LO $2 \rightarrow 3$ contribution, which provides a non-trivial description of the $\Delta\phi_{12}$ distribution tail down to the kinematical limit $\Delta\phi_{12} = 120^\circ$. However, as discussed in Secs. 3.2.4 and 3.3 not only the real contribution diverges when the radiation becomes soft ($\Delta\phi_{12} \rightarrow 180^\circ$), but also the virtual correction, which only contributes at $\Delta\phi_{12} = 180^\circ$, is negative and divergent.

In Fig. 4.1, the features mentioned above for a fixed-order QCD $2 \rightarrow 2$ calculation for the $\Delta\phi_{12}$ distribution at both LO and NLO are shown. The leading jet $p_T (p_T^{\text{max}})$ is required to have $600 < p_T^{\text{max}} < 700$ GeV. The distributions are normalized by the corresponding $\sigma$ which is the 2-jet cross section integrated in the full $\Delta\phi_{12}$ range, in the corresponding $p_T^{\text{max}}$ range.

![Figure 4.1: Inclusive a) 2-jet and b) 3-jet distributions as a function of $\Delta\phi_{12}$ as obtained from a fixed-order 2 $\rightarrow$ 2 calculation at LO and NLO, as well as 2 $\rightarrow$ 2 NLO + PS.](image)

When the observable is sensitive to soft radiation, a fixed-order calculation becomes unreliable, or even unphysical like in Fig. 4.1a where the distribution becomes negative at $\Delta\phi_{12} = 180^\circ$. For $\Delta\phi_{12} \approx 180^\circ$ the large logarithmic contributions from soft parton emissions spoil the perturbative expansion. A prediction from $2 \rightarrow 2$ NLO + PS (Powheg-2j +
4.1. Fixed-order calculations and parton shower applied to the $\Delta\phi_{12}$ distribution

PYTHIA8, which includes the resummation of soft radiation contributions, is also shown in Fig. 4.1.

Additionally, Fig. 4.1b depicts the $\Delta\phi_{12}$ distribution, with the requirement that at least a third jet with $p_T > 30$ GeV is present in the event. Oppositely to Fig. 4.1a, the $p_T$ threshold for the third jet avoids the soft, divergent region at NLO $\Delta\phi_{12} = 180^\circ$, therefore only the LO $2 \to 3$ matrix element contributes. However, it is interesting to observe that the differences between the fixed-order calculation and the prediction from the PS are large close to $\Delta\phi_{12} = 180^\circ$. Jets with $p_T \sim 30$ GeV are soft compared to the hard scale ($p_T^{\text{max}} \sim 600$ GeV).

A $\Delta\phi_{12}$ value away from $180^\circ$ is caused by the recoil of the leading system against extra radiation. As a result, the $\Delta\phi_{12}$ separation between the leading jets generally decreases as the hardness of the extra radiation increases. This is illustrated in Fig. 4.2 for the $\Delta\phi_{12}$ distribution in inclusive 2-jet events, as well as in inclusive 3-jet events where several $p_T$ thresholds for the third jet $p_T (p_{T3})$ are applied. Two different $p_T^{\text{max}}$ ranges are shown Fig. 4.2 as examples. The distributions are normalized by $\sigma$, which is the 2-jet cross section integrated in the full $\Delta\phi_{12}$ range, in the corresponding $p_T^{\text{max}}$ range. The predictions were obtained using the PYTHIA8 MC generator.

If we compare the inclusive 2-jet distribution with the inclusive 3-jet distributions in Fig. 4.2 one can notice that when the $p_T$ threshold for the third jet increases, the azimuthal separation of the leading system generally decreases. More important is to notice that, for example, requiring the presence of at least a third jet with $p_T > 30$ GeV translates into a rather large $\Delta\phi_{12}$ separation (value at which the inclusive 2- and 3-jet distributions start to differ) of around $173^\circ$ for $p_T^{\text{max}} \sim 600$ GeV, and $\Delta\phi_{12} \sim 176^\circ$ for $p_T^{\text{max}} \sim 1200$ GeV. This is understandable since the same recoil from a third jet would provoke a larger boost to the leading system for lower $p_T^{\text{max}}$ scales.

A measurement of $\Delta\phi_{12}$ would be an important test not only for the resummation of soft
radiation but also for the matching schemes between higher-order ME and PS. In addition, at high $p_T^{\text{max}}$ the requirement of extra, much softer measurable jets would be an important complementary probe of soft QCD radiation effects.

### 4.1.1 Initial- and final-state shower contributions

As discussed in Sec. 3.2, during the evolution partons can emit radiation in the form of PS. The separation between the leading jets $\Delta \phi_{12}$ is caused by the recoil produced by the radiation. One can investigate the impact that ISR and FSR separately have on the $\Delta \phi_{12}$ distribution. Figure 4.3 shows the $\Delta \phi_{12}$ distribution when only FSR is generated, when only ISR is generated, and when both ISR and FSR are considered.

\[
\begin{align*}
\text{2} \rightarrow 2 & + \text{ISR} + \text{FSR} \\
\text{2} \rightarrow 2 & + \text{ISR} \\
\text{2} \rightarrow 2 & + \text{FSR}
\end{align*}
\]

Figure 4.3: Predictions from PYTHIA8 MC generator for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$ are shown. Three cases are considered: 1) both ISR and FSR are generated, 2) only ISR is generated, and 3) only FSR is generated.

One can observe that the differences between the distribution where only ISR is considered, and the case where both ISR and FSR are generated are small compared to the case when only FSR is considered. The recoil caused by ISR plays the main role in describing the overall shape of the $\Delta \phi_{12}$ distribution, while FSR acts like a correction mainly close to $\Delta \phi_{12} = 180^\circ$. This observation is an additional reason to study different approaches for the generation of ISR. One of these approaches is discussed in App. C, and it is based on the generation of ISR according to the TMD evolution using the parton branching method. The impact of adding a transverse momentum according to the TMD evolution will be discussed in Sec. 4.3, as well as the contributions from the PS when ISR is generated according to the TMD.
4.2 Impact of the matching of matrix element to parton shower

In Sec. 3.3.1 we introduced different ways to improve a QCD calculation by the combination of PS and ME. In the present section the impact on the $\Delta \phi_{12}$ distribution of the matching and merging approaches for combining PS and ME, will be discussed. The distributions in this section are shown in the range $120^\circ < \Delta \phi_{12} < 180^\circ$ ($2.09 < \Delta \phi_{12} < 3.14$ rad) with a bin size of $1^\circ$ ($\approx 0.0175$ rad) in order to visualize the features of the theoretical models not only near the back-to-back configuration but also at smaller $\Delta \phi_{12}$ where the contribution from extra, hard radiation becomes important.

4.2.1 MC@NLO matching

An exclusive $2 \rightarrow 2$ NLO calculation in QCD diverges at $\Delta \phi_{12} = 180^\circ$ for the $\Delta \phi_{12}$ distribution as observed in Fig. 4.1a. In Sec. 3.3.1 the MC@NLO method for matching NLO matrix elements to PS has been discussed. The MC@NLO expression in Eq. 3.54 preserves the NLO accuracy of the cross section integrated over the radiation phase space, however it modifies the calculation of Born and Born+one parton configurations through the PS counter terms (Eq. 3.54). As a result, the generation of the Born and Born+one parton configurations depend on the specific PS used for determining the counter terms.

The unphysical, fixed-order $2 \rightarrow 2$ MC@NLO prediction (calculated using MadGraph5_aMC@NLO [104]) for the inclusive 2-jet cross section as a function of $\Delta \phi_{12}$ is shown in Fig. 4.4 in logarithmic scale (left) as well as in linear scale (right), in order to facilitate the interpretation of the results. The prediction from fixed-order $2 \rightarrow 2$ NLO calculation is also shown in Fig. 4.4.

![Figure 4.4](image_url)

Figure 4.4: Fixed-order $2 \rightarrow 2$ NLO predictions for the inclusive 2-jet observables. The MC@NLO method has been employed with counter terms from Pythia8 as well as from Herwig++. The distributions are shown in logarithmic scale (left) as well as in linear scale (right).

It is interesting to notice the impact of the counter terms from Eq. 3.54 on the MC@NLO fixed-order predictions compared to the pure NLO distribution: 1) the $2 \rightarrow 3$ configurations
can now carry negative weights as can be observed from the negative part of the distribution for $\Delta\phi_{12} \lesssim 180^\circ$, and this is due to the term $R^{PS}(p_T^2)$ that was subtracted in Eq. 3.54; 2) the $2 \rightarrow 2$ configurations can now be positively weighted as opposed to the pure NLO case, and this is because the positive term $\int dp_T^2 R^{PS}(p_T^2)$ was added to the Born term in Eq. 3.54; 3) the tail of the MC@NLO distributions agree with the real contribution from the NLO calculation; 4) the behaviour of the soft limit $\Delta\phi_{12} \lesssim 180^\circ$ is very different depending on the PS counter terms, which is an additional motivation to study this phase space region experimentally.

As mentioned earlier the distributions differ considerably depending on the PS that is used to determine the counter terms, either Pythia8 or Herwig++ in the present case. The splitting kernels of the shower are constrained to reproduce the DGLAP splitting kernels in the collinear limit. However, in the soft limit the splitting kernels also receive contributions from sub-leading color terms, leading to the coherence feature which was mentioned in Sec. 3.2.2. The differences observed in Fig. 4.4 between Pythia8 or Herwig++ in the region $\Delta\phi_{12} \lesssim 180^\circ$ (soft limit) is then due to the different splitting kinematics, and different splitting kernel functional form away from the collinear limit. As mentioned in Sec. 3.4, the Pythia8 event generator uses a dipole recoil picture for the PS (ordering in $p_T$) whereas the Herwig++ event generator employs an angular ordering for the emissions.

As discussed in Sec. 3.3.1, after the counter terms are added, the PS that corresponds to the used counter terms can be matched to the ME using the Born emission scale as the starting scale of the shower. As a result, two event samples will be generated for the case of $2 \rightarrow 2$ NLO (MC@NLO): the $2 \rightarrow 2$ and $2 \rightarrow 3$ multiplicity samples. These samples are shown in Fig. 4.5 for the cases where subtraction terms from Pythia8 (Fig. 4.5a) and Herwig++ (Fig. 4.5b) are employed. In order to better visualize the features of the calculations, the distributions in Fig. 4.5 are shown in logarithmic scale (left) as well as in linear scale (right).

As one can observe in Fig. 4.5, there is not a clear separation between the $2 \rightarrow 2$ and $2 \rightarrow 3$ sub-samples contributing to the $2 \rightarrow 2$ MC@NLO + PS, instead, both the multiplicities span the full $\Delta\phi_{12}$ range. This is not the case for other matching/merging methods as will be observed in the upcoming sections. One of the features observed in Fig. 4.5 is that when the Herwig++ subtraction terms are employed the $2 \rightarrow 3$ contributions are of the order of less than 1% at large $\Delta\phi_{12}$, whereas for the case of Pythia8 subtraction terms the contributions from the $2 \rightarrow 3$ multiplicity are of the order of 10%. This important feature means that when Herwig++ counter terms are used, the contribution from the $2 \rightarrow 3$ sample partially cancels itself out, and as a consequence the distribution at high $\Delta\phi_{12}$ is generated by the showered $2 \rightarrow 2$ multiplicity.

For both subtraction terms choices the events from the $2 \rightarrow 2$ sub-sample are mainly positively weighted whereas the $2 \rightarrow 3$ events are negatively weighted at $\Delta\phi_{12} \sim 180^\circ$, which is in correspondence with the features observed at the ME level in Fig. 4.4.

### 4.2.2 POWHEG matching

In Sec. 3.3.1 the POWHEG method for matching NLO ME to PS has been discussed. Using the first term on the right-hand side of Eq. 3.58 the $2 \rightarrow 2$ partonic configurations (Born) are calculated, and the events are weighted by the accompanying Sudakov factor. The $2 \rightarrow 3$ partonic configurations (Born+one parton) are calculated from the second term on the right-hand side of Eq. 3.58, and it is the accompanying Sudakov factor which prevents
4.2. Impact of the matching of matrix element to parton shower

Figure 4.5: Predictions from a) $2 \rightarrow 2$ MC@NLO + Pythia8 and b) $2 \rightarrow 2$ MC@NLO + Herwig++ are shown in logarithmic (left) and linear scales (right), for one $p_T^{\text{max}}$ region. The corresponding contributions from $2 \rightarrow 2$ (dotted line) and $2 \rightarrow 3$ (dashed line) ME configurations are shown as dotted and dashed lines respectively.
Chapter 4. Phenomenology of the $\Delta\phi_{12}$ distribution

the divergence when two of the partons become collinear or one becomes soft. The Born+one parton configurations are generated above the scale $\text{ptsqmin}$ (minimum $p_T$ squared cut for the real emission introduced in Eq. 3.57), whereas the Born configurations are generated below $\text{ptsqmin}$.

In Fig. 4.6 the fixed-order prediction given by the POWHEG method (Eq. 3.58) for the inclusive 2-jet cross section is shown as a function of $\Delta\phi_{12}$, for one $p_T^{\text{max}}$ range. Different values of $\text{ptsqmin}$ are depicted.

Figure 4.6: Fixed-order POWHEG-2j predictions are shown for the inclusive 2-jet observable as a function of $\Delta\phi_{12}$, for one $p_T^{\text{max}}$ range. Various $\text{ptsqmin}$ values are compared.

It is very interesting to notice how the features of Eq. 3.58 are reflected in Fig. 4.6. The separation between multiplicities given by the scale $\text{ptsqmin}$ translates in the step seen in the curves. One should remember that because of the momentum conservation the $2 \to 2$ Born configurations populate $\Delta\phi_{12} = 180^\circ$. Using Eq. 3.58 one realizes that if $\text{ptsqmin}$ tends to infinity (or more precisely to the upper scale of the process) then only Born configurations ($2 \to 2$) would be generated (first term on the right-hand side of Eq. 3.58). This agrees with the observation in Fig. 4.6, that the last bin population grows as $\text{ptsqmin}$ increases, and already for $\text{ptsqmin} = (600 \text{ GeV})^2$ the largest fraction of the events is located at $\Delta\phi_{12} = 180^\circ$.

The normalization factor $\sigma$ (2-jet cross section at NLO integrated over the full $\Delta\phi_{12}$ range in the corresponding $p_T^{\text{max}}$ range) is the same for all the distributions in Fig. 4.6.

The default value of $\text{ptsqmin}$ for the POWHEG predictions presented in Sec. 9.3 is $\text{ptsqmin} = (0.8 \text{ GeV})^2$. This value is very small and this implies that no step is present in Fig. 4.6 (the step lies within the last bin), and after showering the events the main contributions to the phase space will be given by the $2 \to 3$ configurations, except for the last bin. As a result the bulk of the events are generated by the second term on the right-hand side of Eq. 3.58.

The two multiplicity samples generated according to Eq. 3.58 and matched/merged using the PS generator PYTHIA8 are shown in Fig. 4.7 for the default $\text{ptsqmin} = (0.8 \text{ GeV})^2$ as
4.2. Impact of the matching of matrix element to parton shower

well as $\text{ptsqmin} = (145 \text{ GeV})^2$, for the inclusive 2-jet $\Delta \phi_{12}$ distribution.

Figure 4.7: The Powheg-2j predictions using a) $\text{ptsqmin} = (0.8 \text{ GeV})^2$, and b) $\text{ptsqmin} = (145 \text{ GeV})^2$ are shown for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$, for one $p_T^{\text{max}}$ range. The $2 \to 2$ and $2 \to 3$ multiplicity samples contributing to the distribution are also depicted.

One can observe that $\text{ptsqmin}$ acts like a merging scale which, for the case in which $\text{ptsqmin} = (145 \text{ GeV})^2$ (Fig. 4.7b) it separates the $2 \to 2$ and $2 \to 3$ multiplicities at $\Delta \phi_{12} \approx 170^\circ$. This is the reason why the value $\text{ptsqmin} = (145 \text{ GeV})^2$ was chosen as an example, to investigate the impact of filling up the region $170^\circ < \Delta \phi_{12} < 180^\circ$ with the Pythia emissions instead of the Powheg real emission.

In addition, one can observe in Fig. 4.7a that, since $\text{ptsqmin} = (0.8 \text{ GeV})^2$ is very small compared to $p_T^{\text{max}}$ the phase space is mainly filled by $2 \to 3$ configurations, and the $2 \to 2$ events only contribute to the last bin where they represent $\approx 0.5\%$.

4.2.3 Relation between POWHEG and MC@NLO. The parameter $\text{hdamp}$.

The Powheg formula in Eq. 3.58 is NLO accurate after integration over the radiation phase space. The $2 \to 3$ configurations generated using the second term on the right-hand side of Eq. 3.58 are weighted by a Sudakov factor which is not present in the starting NLO formula in Eq. 3.29. This Sudakov factor can in principle affect the generation of events even if there is a hard radiation, and as a result there can be differences in the tail of the $\Delta \phi_{12}$ distribution compared to the pure NLO prediction. One can observe this feature in Fig. 4.8 where the fixed-order Powheg-2j is compared to the $2 \to 2$ NLO prediction as well as to the prediction from fixed-order $2 \to 2$ MC@NLO using Pythia8 subtraction terms. The normalization factor $\sigma$ ($2$-jet cross section at NLO integrated over the full $\Delta \phi_{12}$ range in the corresponding $p_T^{\text{max}}$ range) is the same for all the distributions in Fig. 4.8.

As observed in Fig. 4.8 the tail of the distribution from the fixed-order Powheg-2j does not follow the real $2 \to 3$ ME from the NLO calculation. On the other hand we see that the
Chapter 4. Phenomenology of the $\Delta \phi_{12}$ distribution

\[ \Delta \phi_{12} \text{ inclusive 2-jet, } 600 < p_T^{\text{max}} < 700 \text{ GeV} \]

Figure 4.8: Comparison of the fixed-order predictions from Powheg-2j, 2 → 2 NLO, and 2 → 2 MC@NLO (using Pythia8 subtraction terms) for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$. The bottom plot shows the ratio of the distributions to the fixed-order 2 → 2 NLO prediction. The solid band represents the statistical uncertainty.

As discussed earlier, the reason for the difference in the tail of the $\Delta \phi_{12}$ distribution observed in Fig. 4.8 between the fixed-order Powheg-2j and pure 2 → 2 NLO, lies in the Sudakov weight which is applied to the real 2 → 3 ME in the case of Powheg-2j (Eq. 3.58). In the following, a parameter available in Powheg will be discussed, which controls the importance of the contribution from the Sudakov factor in Eq. 3.55, for different $p_T$ values of the real emission.

We can start from Eq. 3.55 and introduce an additional reference scale $h$ such that $p_T \gg h$ implies $d\sigma^{\text{PH}}/dp_T^2 \approx R(p_T^2)$. In order to achieve this behaviour of the real contribution, one can define a function $D(p_T; h)$ such that: $D(p_T; h) \approx 0$ when $p_T \gg h$, and $D(p_T; h) \approx 1$ when $p_T \ll h$. Using the function $D$ one can then rewrite Eq. 3.55 as:

\[
\frac{d\sigma^{\text{PH}}}{dp_T^2} = B \frac{R(p_T^2)}{R(p_T^2)} \exp \left[ - \int_{p_T^2}^{\infty} dp_T^2 \frac{D R(p_T^2)}{B} \right] + (1 - D) R(p_T^2) \quad \text{(4.1)}
\]

If $D \approx 1$ the starting expression Eq. 3.58 is recovered whereas if $D \approx 0$ then Eq. 4.1 approaches $R$ and no Sudakov weight is applied to the event. In the Powhegbox the
4.2. Impact of the matching of matrix element to parton shower

The parameter \( h \) is named \( \text{hdamp} \) in the POWHEGBOX library. The default value of \( \text{hdamp} \) is infinity, meaning that \( D = 1 \) and one recovers Eq. 3.55 for the generation of the real emission.

In Fig. 4.9, the fixed-order POWHEG-2j prediction using default value \( \text{hdamp} = \infty \) is compared to the prediction using \( \text{hdamp} = 145 \) GeV. The latter value of \( \text{hdamp} \) is a generic small value (compared to the scale of the process \( p_{T}^{\text{max}} \)). We observed in Fig. 4.6 that a real emission with \( p_{t} = 145 \) GeV causes a \( \Delta \phi_{12} \) of around 2.9 rad for \( 600 < p_{T}^{\text{max}} < 700 \) GeV. As a result one would expect that below \( \Delta \phi_{12} \approx 2.9 \) rad the fixed-order \( 2 \to 2 \) NLO would start to agree with the fixed-order from POWHEG-2j (using \( \text{hdamp} = 145 \) GeV). This comparison is shown in Fig. 4.9 (the normalization factor \( \sigma \) is the same for all the distributions).

The predictions in Fig. 4.9 nicely show the effect from \( \text{hdamp} \) and how the choice of a small value (compared to the scale of the process) recovers the LO \( 2 \to 3 \) description of the

---

\[
D(p_{t}; h) = \frac{h^2}{h^2 + p_{t}^2},
\]

(4.2)
tail of the $\Delta \phi_{12}$ distribution.

### 4.2.4 LO multi-jet merging

In this section, the impact of LO multi-jet merging on the $\Delta \phi_{12}$ distribution is investigated. This section is based on the ideas presented in Sec. 3.3. The MLM merging method is used for the discussion on the general features of the merging procedure including the impact of the merging scale choice. For the studies in this section, LO $2 \rightarrow 2$ and $2 \rightarrow 3$ QCD samples generated using MadGraph are employed. The events are showered with the Pythia8 event generator. The $2 \rightarrow 4$ sample was not included (as opposed to the predictions in Sec.9.3) because it does not change the general picture and its generation is very inefficient due to the high number of partial amplitudes which are computed.

As discussed in Sec. 3.3, the merging scale $q_{\text{cut}}$ splits the phase space in regions which will be then filled by the different multiplicities. In Fig. 4.10 the resulting distributions from $2 \rightarrow 2$ and $2 \rightarrow 3$ multiplicities at the ME level are shown for the inclusive 2-jet distribution.

For the predictions, merging scales $q_{\text{cut}} = 20$ GeV and $q_{\text{cut}} = 140$ GeV are studied. The choice $q_{\text{cut}} = 20$ GeV is the default value used in the CMS collaboration for producing jet samples. The choice $q_{\text{cut}} = 140$ GeV is the result of a detailed study presented in App. E on the proper merging scale choice depending on the phase space region which is probed by the observable. In App. E, the impact of a bad merging scale choice is investigated, using the inclusive jets $p_T$ measurement as an example.

![Figure 4.10: Inclusive 2-jet distribution at the ME level for $q_{\text{cut}} = 20$ GeV and $q_{\text{cut}} = 140$ GeV.](image)

Similar to $p_{\text{tsqmin}}$ for Powheg, the merging scale $q_{\text{cut}}$ divides the phase space. The $2 \rightarrow 2$ configurations are generated below $q_{\text{cut}}$ whereas the $2 \rightarrow 3$ configurations are generated above $q_{\text{cut}}$. At the ME level the $2 \rightarrow 2$ only contribute at $\Delta \phi_{12} = 180^\circ$. We observed for
4.2. Impact of the matching of matrix element to parton shower

instance that real emissions from the $2 \rightarrow 3$ event sample, with a $p_T$ close to the threshold $q_{\text{cut}} = 140 \text{ GeV}$ cause $\Delta\phi_{12}$ separations of around 2.9 rad.

Figure 4.11: Predictions from QCD merged samples from MadGraph, obtained using a) $q_{\text{cut}} = 20 \text{ GeV}$ and b) $q_{\text{cut}} = 140 \text{ GeV}$ are shown for the inclusive 2-jet observable as a function of $\Delta\phi_{12}$, for one $p_T^{\text{max}}$ range. The $2 \rightarrow 2$ and $2 \rightarrow 3$ multiplicity samples contributing to the distributions are also depicted.

In Fig. 4.11 the result from the merging of the $2 \rightarrow 2$ and $2 \rightarrow 3$ multiplicities are shown for the inclusive 2-jet observable. The PYTHIA8 generator was used for showering the events, and the MLM scheme was employed for the merging of the multiplicities. The predictions using $q_{\text{cut}} = 20 \text{ GeV}$ are shown in Fig. 4.11a and the predictions using $q_{\text{cut}} = 140 \text{ GeV}$ are depicted in Fig. 4.11b. One can clearly observe the following: when the merging scale is small most of the events are $2 \rightarrow 3$ configurations, even for the case of the last bin. The resulting distribution shown in Fig. 4.11a using $q_{\text{cut}} = 20 \text{ GeV}$ differs from the one depicted in Fig. 4.11b using $q_{\text{cut}} = 140 \text{ GeV}$, as will be shown later in Secs. 9.4.4 and 9.2.4 in the comparison to data. The reason for this is that $q_{\text{cut}} = 20 \text{ GeV}$ is not a correct choice for the merging scale as will be discussed in the following.

As observed earlier in this section and in Sec. 3.3.2, a third jet can be emitted either as PS radiation below the merging scale or as a real emission from the ME above the merging scale. As a result, a distribution like the $p_T$ of the third jet ($p_{T3}$) can be sensitive to the merging scale, and a not so smooth transition at the merging scale value in the $p_{T3}$ distribution would be an indication that the merging scale has not been properly chosen. Another observable which is sensitive to the merging scale is the so-called differential jet rate (DJR), which characterizes the scale at which the transition between the (n)-jet and (n+1)-jet configurations occurs. The DJR distribution is also discussed in App. E for the case of the transition from 2-jet configurations to 3-jet configurations (DJR23), since we merge $2 \rightarrow 2$ and $2 \rightarrow 3$ event multiplicities. In the studies reported in App. E I show that a merging scale $q_{\text{cut}} = 20 \text{ GeV}$ is not a proper choice at high $p_T^{\text{max}}$ scales. I applied the studies successfully to two additional, important jet observables: the inclusive jets $p_T$ cross section, and the minimum azimuthal
Chapter 4. Phenomenology of the $\Delta \phi_{12}$ distribution

separation between any two of the four highest $p_T$ jets in inclusive 4-jet events (reported in App. E and App. F respectively).

4.3 Transverse momentum dependent parton distribution applied to $\Delta \phi_{12}$

One of the common features of the theory models discussed in the previous sections is that the ME are convoluted with collinear PDFs. In this section, the impact on the $\Delta \phi_{12}$ distribution of using TMDs instead of collinear PDFs is investigated.

In Sec. 3.1.2 the evolution Eq. 3.5 for a transverse momentum dependent parton distribution was introduced. This equation can be iteratively solved using the Parton Branching method (PB) introduced discussed in App. C. One of the main features of this method is that the solution is fully exclusive, meaning that at each parton splitting the kinematical properties of the partons are known. As a result, the transverse momentum of the generated partons is also known and the TMD can be constructed using the cumulative transverse momentum of the branchings.

The TMD used in the present section was fitted to deep inelastic scattering (DIS) data at various beam energies from HERA 1+2 [114]. The TMD evolution scale obeys an angular ordering condition whereas the renormalization scale at each branching was chosen as the $p_T$ of the emission (under the angular ordering prescription for the evolution scale). The ISR can be generated according to the TMD evolution. More details on the TMD determination, fitting and application can be found in App. C.

Figure 4.12 shows predictions from Powheg-2j at the ME level as well as Powheg-2j + PS. In addition, a prediction from Powheg-2j at the ME level, with the addition of a transverse momentum according to the TMD evolution is shown.

As one can observe in Fig. 4.12, Powheg-2j at the ME level falls steeply at 120° (2.09 rad), which is the kinematical limit for $2 \rightarrow 3$ LO configurations. On the other hand, when a transverse momentum is added according to the TMD, smaller values of $\Delta \phi_{12}$ can be reached. As a consequence, the phase space which is left for the subsequent shower gets reduced. This can be observed from the difference of these distributions with the Powheg-2j ME + TMD + PS prediction.

Additionally, Fig. 4.13 shows the prediction from Powheg-3j at the ME level. Powheg-3j is a $2 \rightarrow 3$ NLO generator therefore it can fill the region $\Delta \phi_{12} < 120^\circ$ with the LO $2 \rightarrow 4$ parton multiplicity. The predictions from Powheg-2j ME and Powheg-2j ME + TMD are also depicted in Fig. 4.13.

As one can observe in Fig. 4.13, the fixed-order $2 \rightarrow 4$ LO predictions from Powheg-3j and the $2 \rightarrow 3$ LO + TMD from Powheg-2j with a transverse momentum according to the TMD evolution, give similar descriptions of the $\Delta \phi_{12}$ distribution tail, despite the fact that the latter uses ME with one final-state parton less than the former.
4.3. Transverse momentum dependent parton distribution applied to $\Delta \phi_{12}$

Figure 4.12: Predictions from POWHEG-2j at the ME level, POWHEG-2j ME including a transverse momentum according to the TMD evolution, and POWHEG-2j + TMD + PS are shown for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$. 

Figure 4.13: Predictions from POWHEG-2j, POWHEG-3j, and POWHEG-2j including a transverse momentum according to the TMD evolution are shown for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$. 

55
Chapter 4. Phenomenology of the $\Delta\phi_{12}$ distribution
At the Large Hadron Collider (LHC) [115], the Compact Muon Solenoid (CMS) [116] stands as one of the four big experiments located along the accelerator ring. Its three other siblings are named ATLAS (A Toroidal LHC ApparatuS) [117], ALICE (A Large Ion Collider Experiment) [118], LHCb (Large Hadron Collider beauty) [119] and they will only be briefly mentioned. In this section the LHC will be introduced. Moreover, the CMS and its different sub-detector systems will be discussed. In addition, the trigger system is introduced in Sec. 5.2.5. The last section is dedicated to a brief description of the CMS detector simulation.

5.1 The Large Hadron Collider

The Large Hadron Collider, located in the European Organization for Nuclear Research (CERN) in Geneva, is world’s largest circular accelerator. Protons and heavy ions are carried by its 27 km long, underground double-ring structure and, in the case of proton beams, they undergo collisions, reaching center-of-mass energies of up to 13 TeV, as of the moment of
writing this thesis. These high energies have allowed for the exploration of the SM and also searches for new physics scenarios that could answer some of the remaining open questions in particle physics.

The protons which later undergo the acceleration process are produced from standard molecular hydrogen. The electrons from the hydrogen atoms are stripped away using an electric field. The resulting protons are then injected to the acceleration system. The protons (or heavier ions) first undergo a ladder-like pre-acceleration process in the linear accelerator (LINAC 2), the proton synchrotron booster (PSB), the proton synchrotron (PS) and super proton synchrotron (SPS) facilities [120] to then enter the LHC as a last acceleration step before collisions occur:

- LINAC 2, energy of the output protons: 50 MeV
- PSB, energy of the output protons: 1.4 GeV
- PS, energy of the output protons: 25 GeV
- SPS, energy of the output protons: 450 GeV

The fundamental constituents of matter and the interactions between them are studied in the four experiments CMS, ATLAS, ALICE and LHCb. ATLAS is a general purpose detector with the particularity that it makes use of a toroidal magnet to bend the charged particle trajectories. Together with CMS it was designed for precise measurements in the SM sector and searches for BSM physics. ALICE is a detector devoted to investigate the early universe conditions by studying the quark-gluon plasma in heavy ion collisions. The LHCb experiment addresses the question of matter-antimatter asymmetry in the universe by looking at CP violation in the production of B-mesons. The CMS detector will be discussed in the next sections. In the following only pp collisions will be considered.

In 2016 the peak instantaneous luminosity delivered by the LHC was $1.52 \times 10^{34}$ cm$^{-2}$s$^{-1}$. The number of colliding bunches was 2208 with $1.18 \times 10^{11}$ protons per bunch. The bunch spacing was 25 ns. By using 1232 dipole magnets capable of producing magnetic fields of about 8.33 T, and 392 quadrupole magnets, the proton beams are kept in orbit and focused respectively.

The LHC will experience a major upgrade in the following years. The project of construction of its successor, the High-Luminosity LHC (HL-LHC) [121] was approved by the CERN Council. It is expected to deliver a peak luminosity of $5 \times 10^{34}$ cm$^{-2}$s$^{-1}$ and an integrated luminosity of 3000 fb$^{-1}$ spanned over ten years of operation starting after 2025.

A new and exciting era is therefore not far ahead, as HL-LHC will provide a sensible increase in the amount of data that will allow for the study of rare processes and extreme phase space regions. The plan until the year 2037 is depicted in 5.1, where 'LS' stands for long shutdown.

The LHC might undergo another important upgrade in the late 2030s. The project is called High-Energy LHC (HE-LHC) and it relies on magnets that can deliver at least twice the LHC dipoles magnetic field strength [123]. This would allow to reach center-of-mass energies of $\sim 27$ TeV.
5.1. The Large Hadron Collider

![Graph showing integrated and peak luminosities delivered by the LHC, and longer term schedule as reported on the LHC commissioning (see Ref. [122]).]

5.1.1 Luminosity and data taking

The number of events per unit time taking place \((dN/dt)\), corresponding to a certain physics process is proportional to the probability of the process to occur and therefore to the cross section \(\sigma\). This can be expressed as:

\[
\frac{dN}{dt} = \mathcal{L} \cdot \sigma
\]  

(5.1)

The proportionality factor \(\mathcal{L}\) is called luminosity and it contains the information on the experimental set-up. The luminosity, for the case of two identical head-on colliding beams with Gaussian profile is defined as [124]:

\[
\mathcal{L} = \frac{N^2 n_b f}{4\pi \sigma_x \sigma_y}
\]  

(5.2)

where \(N\) and \(n_b\) are the number of protons per bunch and the number of bunches respectively. The parameter \(f\) corresponds to the revolution frequency and \(\sigma_x, \sigma_y\) to the cross-sectional width parameters of the Gaussian beams. Even though the starting point of the luminosity evaluation is the Eq. 5.2, in practice it has to be modified to account for corrections due to the non-zero crossing angle, the collision offset, the non-Gaussian profiles, the non-zero dispersion at the collision point, and the dependence of the transverse beam sizes on the longitudinal position (hourglass effect) (Ref. [124]). These corrections account for the beams not being necessarily Gaussian-shaped, the collisions not occurring head-on, the variation of the beams shape in the longitudinal direction, as well as the possible variation of the beams focusing.
The integrated luminosity can then be written as:

\[ \mathcal{L}_{\text{int}} = \int dt \mathcal{L} \quad (5.3) \]

Two methods are used for the measurement of the luminosity in CMS: the on- and off-line methods [125]. The hadronic forward calorimeter (HF) is used for the online measurement (Refs. [126, 127]). The offline method relies on the pixel detector (Ref. [126]).

The measurement of the luminosity implies an inherent systematic uncertainty on the normalization that propagates to the measurement of the observable of interest. In order to reduce the impact of the luminosity uncertainty on the observables, distributions normalized to the cross section lead to a cancellation of the effects. In the studies presented in this thesis the observables are normalized to the cross section in their respective phase space region.

The data used in the present thesis corresponds to the LHC run 2, in the year 2016. The Fig. 5.2a shows the integrated luminosity as a function of time delivered by the LHC and recorded by CMS in 2016. These two numbers are not necessarily the same since the data recorded by CMS have to fulﬁl quality criteria. Figure 5.2b shows the peak luminosity as a function of time.

![Figure 5.2: a) Integrated luminosity as a function of days delivered by the LHC and recorded by CMS in 2016, and b) the peak luminosity in the same period. The information was taken from Ref. [128].](image)

The analyses presented in this work use data from pp collisions at 13 TeV, recorded during the 2016 data taking period and corresponding to an integrated luminosity of 35.9 fb\(^{-1}\).

### 5.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) experiment is located about 100 m underground at Point 5, CERN. The content of this section is mainly based on Refs. [116, 129]. The detector is designed for studying the physics of the SM and beyond by accurately identifying and measuring muons, electrons, photons, and hadrons over a large energy range. For this purpose, its main features include: good muon identification, momentum resolution resolution and dimuon mass resolution; good charged-particle momentum resolution and reconstruction.
efficiency in the inner tracker; good electromagnetic energy, diphoton and dilepton mass resolution; good missing transverse energy and dijet-mass resolution. The CMS detector is arranged in an "onion-like" configuration, symmetric around the beam pipe, and each layer specialized in obtaining different information on the produced particles.

A distinctive feature of the CMS detector is its superconducting 3.8 T magnet designed to bend the trajectory of high-energetic charged particles. The magnet has a solenoidal shape, and it is located around the beam pipe. "Solenoid" came to be part of the name of CMS due to the importance of the magnet. Three detector systems are embedded inside the magnet: the tracking system, the electromagnetic and hadronic calorimeters. This explains the name 'Compact'. Outside the magnet, the dedicated muon stations interleaved with the iron yokes are placed, justifying the 'M' of CMS. The detector has a weight of 12500 t, a diameter of 14.6 m, and a length of 21.6 m. A sketch of CMS and its components is depicted in Fig. 5.3. These components and their main features will be addressed in the following sections.

Figure 5.3: Schematic view of the CMS detector showing the different components (taken from Ref. [116]).

**CMS coordinate system definition**

The CMS coordinate system is defined as right-handed, with the $xy$ plane transverse to the beam pipe. The $x$-axis points towards the center of the LHC ring and the $y$-axis points vertically upwards. The right-handedness convention fixes the $z$-axis. The polar coordinate system is defined accordingly, with the $\phi$ angle measured from the $x$-axis and the radius $r$ being the radius in the $xy$ plane. The angle $\theta$ identifies the polar angle in the $rz$ plane.
Chapter 5. Experimental setup

Important CMS related observables

In hadron-hadron collisions the center of mass of the partonic collision suffers from boosts depending on the longitudinal momentum fractions of the partons undergoing the interaction. Therefore the use of quantities which are invariant under Lorentz boosts along the $z$-axis become important. As a consequence the pseudorapidity $\eta$, defined in Eq. 5.4, is commonly used instead of $\theta$. Differences of this quantity, are Lorentz invariant under boosts along the $z$-axis for massless particles.

$$\eta = -\ln \tan \frac{\theta}{2}$$ (5.4)

As a convention, the region $|\eta| < 2.5$ is called central region whereas the region $|\eta| > 2.5$ is called forward region. Other invariant quantities can be constructed using the symmetry of the collisions at the LHC. For instance the transverse momentum of the object $p_T$, defined in the $xy$ plane, as well as the transverse mass (or transverse energy) defined as $E^2_T \equiv m^2_T = m^2 + p^2_T$, where $m$ is the invariant mass of the object. The variable $R_{ij}$, very important for jet physics and defined in section 3.2.7 shares also this property.

Another related quantity, important for the purposes of the CMS experiment is the so-called missing $E_T$ ($\slashed{E}_T$). This observable allows to detect the presence of weak interacting particles that might escape from detection giving rise to a missing net transverse momentum $\slashed{p}_T = | - \sum_i \vec{p}_{T,i} |$ ($i$ runs over the detected particles). Confusingly enough $\slashed{E}_T$ is defined as $\slashed{E}_T \equiv \slashed{p}_T$, so no strict association with energy should be made by the reader since the mass of the constituents is not taken into account in the definition. The accurate measurement of the $\slashed{E}_T$ is only achievable at CMS due to its very hermetic configuration, as we will discuss in the next sections.

5.2.1 Inner tracking system

The purpose of the tracker is to measure the momentum of the charged particles by tracking its path through the magnetic field, as well as to precisely reconstruct secondary vertices. The CMS tracker system is made up of an innermost pixel detector and an outermost silicon strip tracker.

During the 2016 data taking period the pixel detector was composed of 1440 pixel modules distributed over three barrel layers and complemented with two endcap disks. The size of a pixel module is $100 \times 150 \mu m^2$. The pixel barrels are placed between 4.4 cm and 10.2 cm from the beam line. It has a resolution of $10 \mu m$ in $r\phi$, and $20 - 40 \mu m$ in $z$, as well as a high efficiency.

The silicon strip tracker spans the space around the beam line between 20 cm and 116 cm. It has four Tracker Inner Barrel layers (TIB) located below 55 cm in radius and three Tracker Inner Disks (TID) at each end. They are made of 320 $\mu m$ as well as 500 $\mu m$ thick silicon strip sensors. The Tracker Outer Barrel (TOB), composed by six layers of 500 $\mu m$ thick strip sensors, surrounds the TIB and TID, and extends in radius to 116 cm. The TOB spans the range $|z| < 118$ cm. The Tracker EndCaps (TEC), composed of nine disks of 320 $\mu m$ thick strip sensors each, are located beyond this range (124 cm < $|z|$ < 282 cm). The resulting tracker system coverage is $|\eta| < 2.5$. 

62
5.2. The Compact Muon Solenoid

5.2.2 Electromagnetic calorimeter and preshower detector

The Electromagnetic Calorimeter (ECAL) is a fine granularity, fast detector with a good energy resolution, mainly designed to detect the Higgs boson decaying into two photons. Its dense and homogeneous structure collects and measures the energy deposited by the particles entering the ECAL, initiators of electromagnetic showers. It is composed by the EB (barrel in the central region with 61200 lead tungstate crystals), flanked by two endcaps (EE) made of 7324 crystals. The crystals provide a homogeneous scintillating medium for optimal energy resolution. The scintillation decay time of the high-density crystals is similar to the LHC bunch crossing time density (25 ns). The ECAL is located within magnet and hadron calorimeter. The energy resolution for electrons and photons above 100 GeV is around 0.5%.

The EB is around 3m length with 36 "supermodules" organized in two halves, each spanning $|\eta| < 1.48$. It makes use of avalanche photodiodes (APDs) to detect the scintillation light. The EEs cover the region $1.48 < |\eta| < 3.0$. They make use of Vacuum Photo-triodes (VPTs) which are more resistant to radiation damage than the diodes.

Preshower detector

During the collisions in the CMS detector short-lived $\pi^0$ mesons are also produced. They subsequently decay into two highly collimated photons, difficult to distinguish as individual entities by the ECAL, especially in the high $|\eta|$ region. In order to increase the resolving power of the detector, a preshower detector (ES), based on lead absorber and silicon strips sensors, is placed in front of the EEs ($1.65 < |\eta| < 2.6$). With its finer granularity (2 mm wide strips compared to the 3 cm wide ECAL crystals), it is able to resolve the decay products of the pion as separate photons.

5.2.3 Hadronic calorimeter system

As its name suggests, the Hadronic Calorimeter (HCAL) is designed to measure the energy of hadrons entering its volume. It makes use of alternating layers of absorber and scintillator materials in order to find the position and energy of the particle passing through. The HCAL is a key player in the measurement of hadron jets and $E_T$ from neutrinos or possibly new physics, weakly interacting particles.

The HCAL is located between the ECAL and the magnet coil ($177 < r < 295$ cm). Since the amount of stopping material that fits inside the magnet is constrained, an outer hadron calorimeter (HO) is placed in the central region outside the magnet balancing the decrease of material towards the center of the detector. An inner barrel (HB), endcaps (HE) and a forward calorimeter (HF) are also part of the HCAL system.

The HB, whose baseline active material is 3.7 mm thick Kuraray SCSN81 plastic scintillator, covers the pseudorapidity region $|\eta| < 1.3$. It has a resulting segmentation in the $\eta\phi$ plane corresponding to $(\Delta \eta, \Delta \phi) = (0.087, 0.087)$. With the same segmentation, the HO covers the region $|\eta| < 1.26$. By increasing the effective thickness of the calorimeter, the HO helps improving the energy resolution. The HEs span the pseudorapidity region $1.3 < |\eta| < 3.0$. It is placed at the ends of the magnetic solenoid, which is the reason for its active material to be chosen non-magnetic.

The HF is located 11 m away from the interaction point. It assures a hermetic configuration of the HCAL by extending its coverage to $|\eta| < 5.2$. It is a Cherenkov based detector.
composed by grooved steel absorber plates with quartz fibres inserted as active medium. It is sensitive to both the electromagnetic and hadronic components of the particles shower.

5.2.4 Muon detector system

The outermost detector system corresponds to three dedicated muon gaseous detector chambers, located in the range $4 < r < 7$ m. Additional disk-shaped caps extend the coverage of the muon detector up to $|\eta| < 2.4$. The detector is divided in several layers interleaved with iron yokes for the return of the magnetic flux that also act as a shield against hadrons, improving the muon identification.

Four layers of up to twelve drift tube (DT) chambers cover the region $|\eta| < 1.2$, where the magnetic residual flux and the muon rate are low. On the other hand, the region $1.2 < |\eta| < 2.4$ is exposed to high magnetic flux as well as high muon rate. In that region, cathode strip chambers (CSC) that provide a fast response as well as a good time and spatial resolution are employed. Additionally, resistive plate chambers (RPC) complement the DTs and CSCs in the region $|\eta| < 1.6$. The RPCs have a good time resolution with a better than 25 ns response meaning that they can accurately identify muons coming from the same bunch crossing. The RPC are therefore also used in the trigger system. In the muon momentum reconstruction the information from the muon chambers is complemented with the tracker information.

5.2.5 Trigger system

The high collision rate ($\sim 100$ MHz) in the CMS detector produces an enormous amount of data, most of which is soft QCD as can be expected from the fact that the $t$-channel amplitude becomes large for small transverse momentum (Sec. 3.2.6). A small fraction of the events corresponds to potentially interesting events ($\sim 100$ Hz), which for the main purposes of the CMS experiment occur at high momentum transfer. When events at small transverse momentum are the subject of investigations, for example to improve the MPI models and therefore the underlying-event description, then only a fraction of the same type of events is recorded. The events with a minimum of requirements (activity in the forward detector for instance) are called minimum-bias events. However, the concept of minimum-bias event does not correspond to the underlying-event activity. An easy way to notice this is that the underlying-event is related to the activity around a high momentum transfer partonic collision, and this likely implies a small impact parameter between the parent protons; therefore the underlying-event activity implies a bias by definition.

A selection system, able to quickly decide whether an event is interesting to be further analysed, is needed. Such a task relies on the trigger system [130]. It is divided in two levels: the level 1 trigger (L1), and the high level trigger (HLT) [131]. The L1 trigger is applied at the hardware level and it uses information from the muon chambers and the calorimeter system. It reduces the amount of events to be recorded to $\sim 100$ kHz. Thereafter, the HLT level trigger is applied to the output from the L1, in an offline base. It depends on the specific trigger path, meaning the reconstructed physics objects that are used. In Sec. 6.2, the specific HLT triggers used for the purpose of the present thesis will be discussed. The HLT reduces the rate of events to be analysed to $\sim 100$ Hz.
5.2.6 Detector simulation

In general, the events observed in a physics experiment have the footprint of the specific experimental set-up used for the measurement, and even of the specific experimental conditions. In the case of CMS the recorded events are shaped, for example, by the efficiency of the different detector systems, the granularity and resolution, the pileup conditions and possibly even the weather on the specific date. These conditions fold the events and make a direct comparison to a theory unreliable unless the theoretical predictions account for the same contributions. The GEometry ANd Tracking (Geant4) [132] is the platform commonly used, among many other fields, in high energy physics specifically for the simulation of the CMS detector geometry (active and passive volumes), passage of particles through matter (interactions with media, decay and energy deposition), electronics effects, and the detector response.

The basic principle is the following: an event record from a theoretical model, including the information of the particles in the event are fed into Geant4. The particles are propagated according to their interaction length in the specific medium, with the interaction being sampled according to the corresponding cross section for the particle to interact with the specific material. This implies for instance that the simulation time depends on the specific particles that are being propagated.

The events are finally reconstructed using the same reconstruction algorithms, and trigger information as in the data. Both the event information before the simulation of the detector effects is included (so-called GEN level) and the information after the event is propagated into the simulation of the detector (DET level information) are very important, as they allow to know the functional form of the mapping of the event between the GEN and DET levels. This mapping and its relevance will be the subject of following sections.
CHAPTER

6

EVENT RECONSTRUCTION AND SELECTION

Contents

6.1 Data and simulated samples ........................................ 68
6.2 Physics objects and vertex reconstruction ...................... 69
  6.2.1 Jet reconstruction ........................................... 70
  6.2.2 Trigger path ................................................. 70
  6.2.3 Jet energy calibration and jet energy resolution .......... 73
6.3 Event selection ...................................................... 76
  6.3.1 Observables .................................................. 76
6.4 Bin size and resolution ............................................ 77
  6.4.1 Coarser binning ............................................. 78
  6.4.2 Finer binning ............................................... 78
  6.4.3 Relation between tracks $\phi$ resolution and $\Delta\phi_{12}$ resolution .... 81
6.5 Purity, stability, acceptance and background .................. 83
  6.5.1 The relation between resolution and purity/stability ...... 84
6.6 Pileup ............................................................... 86
6.7 The measurement at detector level ................................ 90
6.8 Detector effects and unfolding to particle level .............. 93
  6.8.1 Unfolding methods .......................................... 93
  6.8.2 Response matrices and unfolded distributions .......... 94
  6.8.3 Regularization .............................................. 97
  6.8.4 D’Agostini vs Tikhonov unfolding ......................... 97
  6.8.5 Migration matrices obtained with a new method .......... 99
Chapter 6. Event reconstruction and selection

6.8.6 Validation of the unfolding ........................................ 103

The information provided by the different sub-detectors of CMS can be combined and further analysed offline in order to build physics objects that can be subsequently used in the experimental analyses. Jets of hadrons are among these objects. This chapter is devoted to discuss the reconstruction of the events and the associated physics objects in CMS. Emphasis will be put on jets as they are the main subject of this thesis. Subsequently, detector effects on the observables are studied and the unfolding algorithm to correct for them is discussed.

6.1 Data and simulated samples

The investigations presented in this thesis are based on proton-proton collision data collected with the CMS experiment at 13 TeV of center-of-mass energy, corresponding to an integrated luminosity of $35.9 \text{ fb}^{-1}$. The LHC operated at 25 ns bunch spacing. The data samples were collected during the 2016 period of data taking. The data taking period is divided in a number of runs, each of them undergoing a process of certification and calibration. In this thesis the following data sets are employed:

- RUN B: /JetHT/Run2016B-23Sep2016-v3/AOD
- RUN C: /JetHT/Run2016C-23Sep2016-v1/AOD
- RUN D: /JetHT/Run2016D-23Sep2016-v1/AOD
- RUN E: /JetHT/Run2016E-23Sep2016-v1/AOD
- RUN F: /JetHT/Run2016F-23Sep2016-v1/AOD

As mentioned in the previous sections, the Monte Carlo simulation of high energetic collisions, and of the detection system is a vital tool not only for testing the validity of the models and theories but also for the good understanding of the detector effects on the physics objects being measured.

The following official CMS MC samples are also used in this analysis:

- The Pythia8 tune CUETP8M1 $p_T$ slice MC: QCD_Pt_xxttoxx_TuneCUETM1_13TeV_pythia8
- The MadGraph MC: QCD_HTxxttoXX_TuneCUETP8M1_13TeVmadgraphMLMpythia8
- The Herwig++ Tune CUETS1 flat MC: QCD_Pt-15to7000_TuneCUETHS1_Flat_13TeV_herwigpp

The Pythia8 [101] MC samples were generated in ranges of the partonic transverse momentum ($p_T$), in other words the events are effectively weighted such that events with high $p_T$ jets are produced with high statistics. The PDF set NNPDF2.3LO [80, 133] was used and the CUETP8M1 [109] tune as the choice for the UE parametrization.

68
6.2 Physics objects and vertex reconstruction

Similarly, the MadGraph [104, 134] MC samples are generated in ranges of the scalar sum of the hard process partons $p_T (H_T = \sum_i p_T$, where $i$ runs over the partons in the hard process), allowing a rather flat event density up to $p_T \sim 1 \text{ TeV}$. The simulation contains QCD $2 \rightarrow 2$, $2 \rightarrow 3$, and $2 \rightarrow 4$ partonic multiplicities at LO. The divergencies in the last two types of process (two partons becoming collinear or a parton becoming soft) are avoided by setting a minimum distance (analogous to the anti-$k_T$ definition) between any pair of partons ($x_{\text{qcut}} = 10 \text{ GeV}$). The Pythia8 event generator with the tune CUETP8M1 was employed for the PS and UE simulation. Since the overlap between the different event multiplicities can occur when they are made exclusive (by means of a Sudakov/PS as discussed in previous sections), a merging based on the MLM scheme was used with a merging scale $q_{\text{cut}} = 20 \text{ GeV}$.

The Herwig++ [102] sample was generated using the PDF set CTEQ6L1 [135] and the tune CUETHppS1 [109] for the UE. The events are unweighted, meaning that the number of events generated as a function of the $p_T$ exchange in the hard process falls rapidly with the $p_T$.

The MC samples from Pythia8, Herwig++, and MadGraph MC generators contain the information on the generated particles before undergoing the detector response simulation (GEN level), and after being subjected to the detector simulation (DET level).

Predictions based on NLO perturbative QCD are obtained with the Powhegbox library [105, 136, 137] and the Herwig7 [103] event generator. Among the NLO event generators the Powheg in dijet mode [138], referred to as Powheg-2j is investigated. Powheg-2j provides an NLO $2 \rightarrow 2$ calculation. Additionally, the Powheg generator in three-jet mode [139], referred to as Powheg-3j and providing a NLO $2 \rightarrow 3$ calculation was utilized. The Powheg predictions were obtained using the default $p_{\text{tsqmin}}$ cut (introduced in Sec. 3.3.1) of $(0.8 \text{ GeV})^2$. The impact and the meaning of this parameter will be further discussed in the results section.

The Pythia8 (tune CUETPSM1) and Herwig++ (tune CUETHppS1) generators are used to dress the Powheg output with PS, hadronization, and MPI. Predictions from the Herwig7 event generator are based on the MMHT2014 PDF set [140] and the default tune H7-UE-MMHT [103] for the UE simulation.

The details on the MC generators mentioned above can be found in Sec. 3.4, as well as in Apps. B, A. In addition, the Cascade event generator, introduced in Sec. 3.4 and App. C, is employed together with TMD sets (see App. C for details on the TMD sets determination). Finally the PS generators Vincia and Dire, introduced in Sec. 3.4, is employed to test the impact on the observables when higher-order corrections to the PS are considered.

6.2 Physics objects and vertex reconstruction

In a highly energetic collision many particles are produced. The unstable particles will undergo subsequent decays and only photons, electrons, muons and hadrons will remain to be detected in the dedicated sub-detectors. In practice what we "see" in the detector are energy deposits in the calorimeters and hits in the tracking system together with signals in the muon chambers. It is the task of the particle-flow (PF) algorithm (Ref. [141]) to combine this information and associate a group of hits in the tracker to the energy deposits in the ECAL and HCAL, and reconstruct their four-momenta.

The first step in the reconstruction is to associate hits in the tracker and muon chambers
Chapter 6. Event reconstruction and selection

with tracks. Depending on the resulting combined track fit muons are identified. Remaining tracks that are associated with energy depositions in the ECAL are tagged as electrons. The remaining tracks that have a corresponding HCAL cluster are considered as charged hadrons and their momenta are reconstructed from both the tracker information and the energy deposition in the HCAL. Photons and neutral hadrons are associated with energy depositions without associated tracks in the ECAL and HCAL respectively. One should mention that quality cuts are implicit in the procedure, for instance a track should have a minimum number of hits and a threshold for the energy deposition in the calorimeters is also present. The successfully reconstructed objects (PF candidates) with their respective four-momenta are then used as input for the reconstruction of jets, τ leptons and $E_T$. A τ lepton decaying in the hadronic decay channel ($\sim 60\%$) for instance, results in a few, collimated hadrons depositing their energy in the HCAL. The $E_T$ corresponds to the transverse imbalance of the event, which is not zero for the case that weakly interacting particles are produced and pass undetected, neutrinos for instance.

With the track information the collision vertex is determined. This is especially important in the luminosity conditions of the LHC that led to an average number of collisions per bunch crossing (pileup) of 27 in 2016. This pileup (PU) value was determined using the LHC standard value of 80 mb for the minimum bias cross section (see Ref. [128] for further details). The 'pileup' concept and its implications for the work presented in this thesis will be discussed in the next sections. The vertex position is determined by means of a vertex fit algorithm which makes use of tracks, preselected according to their impact parameter significance and their $p_T$. The vertex candidates are then ranked by the scalar sum $\sum_i p^2_T$, where $i$ runs over the associated jets, clustered with the anti-$k_T$ algorithm described in Sec. 3.2.7. The vertex with the largest $\sum_i p^2_T$ is denoted as the primary vertex. More details are given in Ref. [142].

6.2.1 Jet reconstruction

As described in Sec. 3.2.7, jets are defined by a jet algorithm. Jets are also defined at detector level, and as mentioned in Sec. 3.2.7 the same definition has to be applied to the data and the corresponding simulation. In the case of CMS the reconstructed PF objects enter the jet algorithm. The jet four-momentum is then computed from the four-momenta of the PF objects composing the jet.

Together with the PF reconstructed jets, in CMS there are other jet reconstruction definitions which make use of specific information from the detector. The reconstruction of the so-called calo-jets only uses information from the calorimeter systems. Similarly, the track-jets are reconstructed using only information from the tracker. Since the PF jets are reconstructed using the information from both the tracker and the calorimeters, it provides the best performance (Ref. [143]). Only PF jets are used for the purpose of this thesis.

6.2.2 Trigger path

Jet triggers are very important, especially for QCD physics. Single, double and triple jet triggers can be defined (Ref. [144]). The definition of a single trigger is given by a transverse energy threshold, the angular region ($\eta, \phi$), and by the prescale factor which reduces the trigger output rate. The multi-jet triggers require a minimum separation in $R$ between the jets. Single jet triggers are suited for QCD analyses whereas double triggers are normally
used in energy calibration studies (Ref. [131]).

Eight HLT single PFJet triggers, labelled as 'HLT_PFJetX' were studied for the purpose of this thesis, with X taking the following values: 80, 140, 200, 260, 320, 400, 450 and 500. The first six triggers are prescaled whereas the two highest HLT_PFJet450 and HLT_PFJet500 are not. The effective luminosities for each of the triggers are shown in Tab. 6.1, along with the corresponding L1 and HLT thresholds. For instance, when the $E_T$ deposited in the calorimeter exceeds 140 GeV and the clustered jet $E_T > 200$ GeV, the event is selected by the HLT_PFJet200 trigger.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HLT_PFJet80</td>
<td>60</td>
<td>80</td>
<td>0.003 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet140</td>
<td>80</td>
<td>140</td>
<td>0.024 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet200</td>
<td>140</td>
<td>200</td>
<td>0.104 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet260</td>
<td>200</td>
<td>260</td>
<td>0.592 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet320</td>
<td>260</td>
<td>320</td>
<td>1.765 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet400</td>
<td>320</td>
<td>400</td>
<td>5.171 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet450</td>
<td>400</td>
<td>450</td>
<td>35.877 fb$^{-1}$</td>
</tr>
<tr>
<td>HLT_PFJet500</td>
<td>450</td>
<td>500</td>
<td>35.877 fb$^{-1}$</td>
</tr>
</tbody>
</table>

Table 6.1: The HLT PFJet triggers, the corresponding L1 and HLT thresholds as well as the effective luminosity.

The investigations in this thesis are done in intervals of the highest $p_T$ jet in the event ($p_T^{\text{max}}$). Therefore it is convenient to choose for each $p_T^{\text{max}}$ range the trigger which gives the highest number of events. In practice there is a compromise between this requirement and the fact that the trigger should be efficient in the corresponding region.

The efficiency of each trigger is evaluated by means of the trigger emulation method. This method consists in using two consecutive triggers such that the efficiency of the higher is measured relative to that of the lower one. The lower trigger thresholds (L1 and HLT) are emulated on the events that have passed a higher, more efficient trigger. The advantage of emulating the lower trigger thresholds is that the resulting statistics is higher than when the lower trigger itself is used.

The efficiency is parametrized by the $p_T^{\text{max}}$ of the event. The resulting efficiency curves are shown in Fig. 6.1. In order to estimate when a trigger becomes fully efficient with respect to the previous trigger, an error function shifted to be between 0 and 1 (Eq. 6.1) is employed:

$$\epsilon(p_T; a, b) = 0.5 + 0.5 \times \text{erf}(a \times p_T - b)$$  \hspace{1cm} (6.1)

When the function in Eq. 6.1 reaches the value 0.999, the trigger is considered to be fully efficient and the corresponding $p_T^{\text{max}}$ value is taken as the turn-on point of the trigger. In addition, the trigger efficiency depends upon $\eta$ as is depicted in fig. 6.1. The highest values between the two $\eta$ ranges were considered as the final turn-on points.
Figure 6.1: HLT PFJet triggers efficiency curves for a) $|\eta| < 1.0$ and b) $1.0 \leq |\eta| < 2.5$. 
The HLT triggers turn-on points are shown in Tab. 6.2. Table 6.2 also shows the triggers that are used depending on the $p_T^{\text{max}}$ range, as well as the corresponding turn-on points. Only one trigger is used per $p_T^{\text{max}}$ range.

<table>
<thead>
<tr>
<th>HLT Path</th>
<th>Turn-on points (GeV)</th>
<th>$p_T^{\text{max}}$ regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLT_PFJet140</td>
<td>183</td>
<td>200-300 GeV</td>
</tr>
<tr>
<td>HLT_PFJet200</td>
<td>253</td>
<td>300-400 GeV</td>
</tr>
<tr>
<td>HLT_PFJet320</td>
<td>391</td>
<td>400-500 GeV</td>
</tr>
<tr>
<td>HLT_PFJet400</td>
<td>485</td>
<td>500-600 GeV</td>
</tr>
<tr>
<td>HLT_PFJet450</td>
<td>548</td>
<td>&gt;600 GeV</td>
</tr>
</tbody>
</table>

Table 6.2: The HLT trigger turn-on points and the corresponding $p_T^{\text{max}}$ regions.

6.2.3 Jet energy calibration and jet energy resolution

A precise knowledge of the jet energy scale (JEC) factors as well as the jet energy resolution (JER) is decisive for the correct interpretation of any jet observable. Both quantities can cause a shift of the measured $p_T$ of the jets. For instance, the precise knowledge of the JES and JER is important for a good determination of the $E_T$ which in turn is crucial for the searches of new physics, often predicted as the $E_T$ from the weakly interacting particles of the underlying theory. It is also crucial for understanding the migration effects between bins in the $p_T$ spectrum. The latter is especially important for the work presented in this thesis, since the observables under study are measured in ranges of the highest jet $p_T$ in the event.

An essential quantity for the calibration of jets is the jet response $\mathcal{R}$. This quantity is a measure of the effects that the characteristics of the detector system have on the properties of the jets that are measured. It is defined as:

$$\mathcal{R} = \frac{p_T^{\text{GEN}} - p_T^{\text{DET}}}{p_T^{\text{GEN}}},$$

(6.2)

where $p_T^{\text{GEN}}$ stands for the $p_T$ of the 'jet generated by nature' or equivalently the $p_T$ of the jet that would be measured by an ideal, perfect detector. On the other hand $p_T^{\text{DET}}$ would be the jet $p_T$ that is actually measured in the detector. We do not know precisely how nature ultimately works, rather nature is approached using the theories and models that we develop. Therefore $p_T^{\text{GEN}}$ will stand for the $p_T$ of the jet generated in a MC simulation without the simulation of the detector response. As $p_T^{\text{DET}}$ can be both simulated and measured in the real experiment, it will stand for both and a specification will be made in case of its usage.

Jet energy scale

The jet energy calibration (correction) maps the jet energy measured in the detector to that at the generator level (measured by an ideal perfect detector). A corrected jet energy
Chapter 6. Event reconstruction and selection

$E_{\text{cor}}$ is obtained by applying a correction factor $C$ to the jet energy given by the calorimeter tower $E_{\text{uncor}}$ such that $E_{\text{cor}} = C \cdot E_{\text{uncor}}$. The correction factors is defined as:

$$C = C_{\text{off}}(p_T^{\text{uncor}}, \eta) \circ C_{\text{rel}}(p_T', \eta) \circ C_{\text{abs}}(p_T'') \circ C_{\text{res}}(\eta) \quad (6.3)$$

The multiplication symbol 'o' indicates that the different corrections are not commuting therefore have to be applied in the given order. First, the offset factor $C_{\text{off}}(p_T^{\text{uncor}}, \eta)$ corrects for electronic noise, as well as pileup effects by subtracting an average energy per unit jet area as will be discussed in Sec. 6.6. The correction allows to remove the average contribution of neutral hadrons from pileup. Subsequently, the relative correction factor $C_{\text{rel}}(p_T', \eta)$ is applied to the resulting jet in order to ensure a uniform jet response as a function of $\eta$ with respect to a control $\eta$ region (central region). The factor is obtained in dijet events, back-to-back in $\phi$ such that one of the jets lies in the $\eta$ control region and while the other one lies in the $\eta$ region being probed. Subsequently the absolute correction $C_{\text{abs}}(p_T'')$ corrects the control region response. This is achieved by measuring a jet together with a back-to-back well measured and calibrated object such as $\gamma$ or $Z$. Finally, $C_{\text{res}}(\eta)$ removes the small residual differences that might remain after applying the previous factors. Details of the procedure can be found in Ref. [145].

The resulting factor has an associated uncertainty which can lead to systematic effects on the results. A dedicated study of these effects is presented in Sec. 6.

The jet energy scale (JES) factors used in this thesis are the ones recommended by the JetMet CMS group: Summer16_23Sep2016(RUN). The determination of the JES factors for the 2016 data resulted in unexpectedly large values for $\eta > 2.5$. The source of this result was later on understood during the 2018 commissioning, when it was observed that the L1 level trigger was being fired earlier than expected because of transparency degradation in ECAL crystals at high $|\eta|$ (see Refs. [146, 147]). The studies presented in this thesis were not affected by this issue because the reconstructed jets are selected with rapidity $|y| < 2.5$, as will be discussed in Sec. 6.3.

Jet energy resolution

The jet response, defined in Eq. 6.2, can be characterized by a Gaussian distribution (Fig. 6.2), although non-Gaussian low response tails are present. The jet $p_T$ resolution, responsible for migration of events between the $p_T^{\text{max}}$ bins, is then related to the width of the jet response. Some features of the non-Gaussian tails will be discussed later in this section.

The Gaussian core of the jet $p_T$ resolution is shown in Fig. 6.3 as extracted from the official PYTHIA8 tune CUETP8M1 and MadGraph CMS MC simulations. The values of the jet $p_T$ resolution, obtained from the two MC event generators are summarized in 6.3. The resolution becomes better as the jet $p_T$ increases. At high $p_T$, the compromise between the better performance of the calorimeter and the decrease of the accuracy of the tracker implies that the resolution becomes nearly flat. The two generators depicted in Fig. 6.3 result in a similar jet $p_T$ resolution. The resolution curves are fitted with a function which increases as $\sim 1/p_T$ as $p_T$ decreases, and becomes flat at high $p_T$. The resulting fit parameters agree between the two MC samples.
6.2. Physics objects and vertex reconstruction

Figure 6.2: The $p_T$ core resolution obtained from PYTHIA8 official CMS MC for two ranges of $p_T$. 

![Figure 6.2](image)

Figure 6.3: a) Fractional $p_T$ Gaussian core resolution (with respect to the $p_T$ of the jet) obtained from PYTHIA8 and b) MadGraph official CMS MC.

![Figure 6.3](image)
6.3 Event selection

Each event is required to have at least one offline-reconstructed vertex (Ref. [148]). The following conditions are imposed:

- at least one primary vertex (PV) candidate;
- the $z$ component of PV satisfies $|z(\text{PV})| < 24$ cm;
- the radius in the $x-y$ plane of the PV satisfies $< 2$ cm;
- the number of degrees of freedom in the vertex fit $> 4$;

where $|z(\text{PV})|$ represents the position of the proton-proton collision along the beam-line and $z = 0$ indicates the center of the CMS detector.

In order to prevent spurious contributions tight identification criteria are applied (Ref. [149]) to the reconstructed jets: each jet should contain at least two particles, one of which is a charged hadron, and the jet energy fraction carried by neutral hadrons and photons should be less than 90%. These criteria have an efficiency greater than 99% for genuine jets.

The observables discussed in this thesis can be separated in two groups: observables measured with a fine binning and observables measured with a coarser binning. In the following, we introduce and discuss the observables measured with a fine binning. The distributions which are measured with a coarser bin size will be mentioned in Sec. 6.3.1 and the details on the corresponding event selection, resolution studies, and systematic effects can be found in Sec. A.

All the jets with a minimum $p_T$ of 30 GeV and $|y| < 5$ are reconstructed. We define the 2- and 3-jet inclusive event samples as follows:

- Inclusive 2-jet:
  1. events with at least 2 jets (labelled as '1' and '2')
  2. $|y_1| < 2.5$ and $|y_2| < 2.5$ are required
  3. $p_{T1} > 200$ GeV and $p_{T2} > 100$ GeV are required

- Inclusive 3-jet:
  1. events that satisfy the inclusive 2-jet definition
  2. the presence of at least a third jet with $|y_3| < 2.5$ and $p_{T3} > 30$ GeV is imposed

The set of cuts $p_T/\text{GeV} = \{200, 300, 400, 500, 600, 700, 800, 1000, 1200\}$ define the ranges in which the phase space is sectioned. The measurement is then performed in the intervals $p_{Ti} < p_{Ti}^{\text{max}} < p_{Ti+1}$, and also in the two additional $p_{T}^{\text{max}} > 1200$ GeV and $p_{T}^{\text{max}} > 2000$ GeV.

6.3.1 Observables

The main observables under study in this thesis are the inclusive 2- and 3-jet cross sections as a function of the azimuthal separation $\Delta \phi_{12}$ between the highest $p_T$ jets. The definitions of inclusive 2- and 3-jet topologies were introduced in the previous section.
6.4 Bin size and resolution

The inclusive 2- and 3-jet cross sections are measured differentially in \( \Delta \phi_{12} \), and in ranges of the highest \( p_T \) among the jets in the event (\( p_T^{\text{max}} \)):

\[
\frac{1}{\sigma} \frac{d\sigma}{d\Delta \phi_{12}}(\Delta \phi_{12}, p_T^{\text{max}}),
\]

(6.4)

where the distributions are normalized to the inclusive 2-jet cross section \( \sigma \) in the corresponding \( p_T^{\text{max}} \) range, integrated in the full \( \Delta \phi_{12} \) range. The observable is invariant under longitudinal boosts. The radiation which is emitted off the hard system produces a recoil that results in \( \Delta \phi_{12} \) taking values different than \( 180^\circ \). The measurements and the investigations discussed in this thesis are focused on event topologies with nearly back-to-back leading jets (\( 170^\circ < \Delta \phi_{12} < 180^\circ \)), for which a bin size of \( 1^\circ \) was chosen. The jets being approximately back-to-back generally implies a small recoil from soft radiation, against the hard system. One also refers to this region as "Sudakov region" or "resummation region".

Observables measured with a coarser bin size

The azimuthal correlation between the leading jets \( \Delta \phi_{12} \) in inclusive 2-jet topologies is also measured using a coarser bin size of \( 5^\circ \), in the wider range \( 90^\circ < \Delta \phi_{12} < 180^\circ \). For these measurements the distributions are normalized to the inclusive 2-jet cross section \( \sigma \) in the corresponding \( p_T^{\text{max}} \) range, integrated in the range \( 90^\circ < \Delta \phi_{12} < 180^\circ \). The results of this interesting measurement, which I have significantly contributed to, are published in Ref. [6] and can also be found in App. A. The corresponding event selection corresponds to the one presented in the previous section for the inclusive 2-jet configuration.

Missing transverse energy

The contributions from events originated by \( t\bar{t} \) and heavy vector boson production to the low \( \Delta \phi_{12} \) region has been investigated in Ref. [5]. The studies in this thesis are focused on the range \( 170^\circ < \Delta \phi_{12} < 180^\circ \), but the observables are normalized to the inclusive 2-jet cross section integrated over the full \( \Delta \phi_{12} \) range. As a consequence, the normalization of the distributions can be affected by \( t\bar{t} \) and heavy vector boson contributions at low \( \Delta \phi_{12} \). Since a large fraction of \( t\bar{t} \) and Z/W+jets events have neutrinos as decay products, these contributions are characterized by a large \( \sum_i E_T \).

The distributions in Fig. 6.4 depict the \( E_T / \sum_i E_T \) (where \( i \) runs over all the PF candidates) of the data compared to the simulation of QCD events using the \textsc{Pythia8} event generator. A disagreement is observed for the low \( \Delta \phi_{12} \) region and high \( E_T / \sum_i E_T \).

In order to reduce the contributions at large \( E_T / \sum_i E_T \), in the event selection the events are required to have \( E_T / \sum_i E_T < 0.26 \).

6.4 Bin size and resolution

Even though the azimuthal correlation in inclusive 2-jet events have been measured previously by several collaborations (Refs. [1–6]), the studies presented in this thesis contain a novel, precise and dedicated investigation of the Sudakov region. In the following, the resolution studies of the measurements, as well as the bin migration and detector effects on the observables are discussed.
Chapter 6. Event reconstruction and selection

6.4.1 Coarser binning

The resolution studies regarding the measurement of azimuthal correlations using a coarser bin size of 5° are analogous to the ones presented in the following section. I have studied the resolution using 5° as the choice for the bin size, the results are reported in App. A. The use of a coarser bin size of 5° allowed us to also focus on different, complementary regions of phase space, where statistics would have been otherwise compromised.

6.4.2 Finer binning

A finer bin size has to be inevitably taken as a compromise with the resolution of the observable which is being measured. In order to perform the $\Delta \phi_{12}$ resolution studies, the Pythia8 tune CUETP8M1 and the MadGraph + Pythia8 tune CUETP8M1 MC event generators are used. The distribution in Fig. 6.5 shows the difference between the generated and reconstructed $\Delta \phi_{12}$ for one $p_T^{\text{max}}$ range as an example, for the inclusive 2- and 3-jet samples. The resolution curves which I have obtained for the full list of $p_T^{\text{max}}$ ranges can be found in App. H. The distributions are fitted with a function consisting of the sum of two Gaussians and an exponential (expression 6.5).

$$f(x; A_1, \sigma_1, m_1, A_2, \sigma_2, m_2, a, b) = f_1(x; A_1, \sigma_1, m_1) + f_2(x; A_2, \sigma_2, m_2) + f_3(x; a, b, m_2), \quad (6.5)$$

where:

$$f_1(x; A_1, \sigma_1, m_1) = A_1 e^{-\frac{(x-m_1)}{\sigma_1}^2}, \quad (6.6)$$

$$f_2(x; A_2, \sigma_2, m_2) = A_2 e^{-\frac{(x-m_2)}{\sigma_2}^2}, \quad (6.7)$$

Figure 6.4: The $E_T/\sum E_T$ in data and simulation for the two $\Delta \phi_{12}$ regions: a) $0^\circ < \Delta \phi_{12} < 180^\circ$ and b) $170^\circ < \Delta \phi_{12} < 180^\circ$. 
The narrower Gaussian describes the core of the resolution, whereas the wider Gaussian and the exponential function describe the tails. Very good $\chi^2/\text{ndf}$ values were achieved, which is impressive if we keep in mind how simple the function’s form is. As $p_T^{\text{max}}$ increases the width of the Gaussians becomes smaller.

The two Gaussians distributions included in the resolution curves were approximated by one Gaussian in order to get a width 'average' estimation. The resulting Gaussian core width is shown as a function of $p_T^{\text{max}}$ for the inclusive 2-jet distributions (Fig. 6.6 a)) as well as for the inclusive 3-jet observables (Fig. 6.6 b). The curves were obtained using the two MC generators Pythia8 and MadGraph. There is little difference observed between the resolution obtained using the two different MC generators.

Figure 6.6 also shows that the $\Delta\phi_{12}$ resolution improves as we go up in $p_T^{\text{max}}$. In addition it shows that the model dependence of the resolution is small, when comparing the simulations from Pythia8 tune CUETP8M1 and MadGraph + Pythia8 tune CUETP8M1 MC event generators.

At high $p_T$ the resolution does not improve significantly, taking similar values for $p_T^{\text{max}} \approx 1000$ GeV and for $p_T^{\text{max}} \approx 2000$ GeV, as can be observed in Fig. 6.6.

The values of the resolution shown in Fig. 6.6 imply that in the current case ($\Delta\phi_{12} \in [170^\circ, 180^\circ]$), the $\Delta\phi_{12}$ resolution is better than when the full range is considered (see App. B). For example, the resolution for the full range takes values of $1^\circ$ ($0.5^\circ$) for $p_T^{\text{max}}$ of the order of 200 GeV (1 TeV), whereas for the case $\Delta\phi_{12} \in [170^\circ, 180^\circ]$ the resolution takes values of $0.8^\circ$ ($0.4^\circ$) for $p_T^{\text{max}}$ around 200 GeV (1000 GeV).

The bin width for the measurements of the normalized 2- and 3-jet cross sections was set to $1^\circ$ which is around twice the value of the resolution core width. It is therefore very important to further check and quantify the stability of the results with the chosen bin size and the given resolution. Furthermore a significant migration between phase space points

\[ f_3(x; a, b, m_2) = e^{a-b|x-m_2|} \]  

(6.8)
Chapter 6. Event reconstruction and selection

![Graph](image)

Figure 6.6: The Gaussian core $\Delta \phi_{12}$ resolution as a function of $p_T^{\text{max}}$ for a) 2-jet and b) 3-jet inclusive events. The predictions obtained using Pythia8 tune CUETP8M1 and MadGraph + Pythia8 tune CUETP8M1 MC generators are compared.

...of the observable is expected due to the size of the resolution compared to the bin width. Dedicated sections to study these issues and the proper recovery from the detector effects are presented later on.

One of the sources contributing to the tails of the resolution curves corresponds to cases in which the $p_T$ ordering of the jets changes due to detector effects, for example, when the second hardest jet in the event at GEN level corresponds to the third hardest jet at DET level. In order to estimate the impact of these contributions the jets at GEN and DET levels are matched not only in $p_T$ (by taking the leading and subleading in both samples) but also a minimum distance $R$ between the matched candidates of 0.2 is required. Figure 6.7 depicts the resolution curve corresponding to one $p_T^{\text{max}}$ region for the case in which the jets are matched in $p_T$ compared to the case when the jets are also matched in $R$.

When the matching is done in both $p_T$ and $R$ the tails are significantly reduced.
6.4. Bin size and resolution

6.4.3 Relation between tracks φ resolution and ∆φ_{12} resolution

The values of the ∆φ_{12} resolution are small, of the order of half a degree. It is therefore interesting and important to check how these values compare to the precision of the measurement of the φ angle of individual tracks.

Figure 6.8 illustrates the φ resolution of single tracks in t\bar{t} events (taken from Ref. [150]). Figure 6.8a depicts the track φ resolution, as a function of the track η for single, isolated pions with transverse momenta of 1, 10, and 100 GeV. On the right, Fig. 6.8b shows the track φ resolution, as a function of the track p_T for charged particles in simulated t\bar{t} events with pileup. The solid (open) symbols correspond to the half-width for 68% (90%) intervals centered on the mode of the distribution in residuals, as described in the Ref. [150]).

Tracks with a low p_T (~ 1 GeV) have a φ resolution of ~ 0.5° (0.0087 rad), which is of the same order of the resolution found for the reconstructed jets. One reason for the good resolution observed at the level of the jet is that high p_T jets are well collimated. In addition, the contribution from high p_T tracks, which generally have a good resolution as observed in Fig. 6.8, becomes more important for high p_T jets.

In order to assess the contribution to the ∆φ_{12} resolution coming from the inaccuracy of the tracks φ 2-jet inclusive events were simulated using the PYTHIA8 MC generator.

The charged particles φ was smeared according to the functional form depicted in Fig. 6.8. Subsequently the resulting ∆φ_{12} resolution is estimated from the difference between the ∆φ_{12} obtained with the original tracks, and the one after the tracks φ is smeared. The result is depicted in Fig. 6.9 together with a fit using the function defined in Eq. 6.5. This can be considered as a partial simulation of the detector response, where only the effect of the tracker on the φ of the tracks is estimated.

The resulting ∆φ_{12} resolution is ~ 0.02°. As expected, the ∆φ_{12} resolution is much
Figure 6.8: Track $\phi$ resolution as a function of a) the track $\eta$ for single, isolated pions with transverse momenta of 1, 10, and 100 GeV and b) the track $p_T$ for charged particles in simulated $t\bar{t}$ events with pileup. The solid (open) symbols correspond to the half-width for 68% (90%) intervals centered on the mode of the distribution in residuals, as described in the Ref. [150].

Figure 6.9: Difference between the $\Delta\phi_{12}$, obtained with tracks simulated using the PYTHIA8 MC generator, and the $\Delta\phi_{12}$ that results after smearing the tracks $\phi$ according to the distributions in Fig. 6.8.
smaller than the one of individual low $p_T$ tracks, when only the tracks $\phi$ smearing is considered. Contributions like the smearing in $p_T$ and $\eta$ as well as the information from the calorimeters account for the remaining difference with the resolution obtained from a full detector simulation (Fig. 6.5). The asymmetry observed between the tails is due to the correlation between the distributions before and after smearing (see discussion on resolution and migration matrix in App. D).

6.5 Purity, stability, acceptance and background

Following the previous discussion, the detector set-up induces a smearing of $\Delta \phi_{12}$ that would have been measured in a perfect, ideal detector. The fact that the resulting resolution for the fine binning observables is of the order of one half of the chosen bin size implies that the migration effects have to be carefully checked. One of the important checks is to quantify the amount of events that are measured in neighbouring $\Delta \phi_{12}$ bins to that in which the event would have had ideally contributed to. Events can migrate in and out of a bin. If the bin size is large (for instance $\sim 10 \times \text{resolution}$) then the bin content would not be dramatically affected by the migration. However, if the resolution is of the order of the bin size (like in the present case) the migration can be large.

Migration can occur inside or outside the defined phase space. In order to study the migration effects inside the phase space, the concepts of "purity" and "stability" characterizing the observables have been studied. The concept of purity denotes the fraction of events that was produced in a bin at GEN level, and is measured in the same bin after the detector simulation. Similarly, stability represents the part of the events that is measured in a bin at DET level and that came from the same bin at GEN level.

On the other hand, the quantities "acceptance" and "background" have been used to study the effects of event migration outside the phase space. "Acceptance" corresponds to the fraction of events selected at GEN level for which a corresponding DET level event is found. Additionally, "background" denotes the fraction of events which are measured at DET level, and do not have a corresponding GEN level.

Figure 6.10 shows the purity, stability, acceptance and background distributions for both the inclusive 2- and 3-jet observables in bins of $\Delta \phi_{12}$ for two $p_T^{\text{max}}$ regions. The same quantities for the remaining $p_T^{\text{max}}$ ranges can be found in Figs. H.3 and H.4 (App. H).

The purity and stability curves are smooth and flat (except for the last bin), suggesting that the resolution and the migration are stable across the $\Delta \phi_{12}$ bins. The last bin is different in the sense that, for instance, events at $\Delta \phi_{12} = 180^\circ$ can only migrate to lower bins resulting in a higher values of purity and stability. The purity and stability increase with $p_T^{\text{max}}$, from $\sim 40\%$ at low $p_T^{\text{max}}$ to $\sim 70\%$ at high $p_T^{\text{max}}$ (App. H). This trend is a consequence of the improvement of the resolution as $p_T^{\text{max}}$ increases (Sec. 6.4). The acceptance is smooth and stable with values between 75%-95%, whereas the background takes values of 5%-25%.

These values suggest the following: 1) there is a considerable migration within the phase space especially for low $p_T^{\text{max}}$ due to the fact that the resolution is comparable to the bin size; 2) the migration inside the phase space is stable and smooth; 3) the migration outside the phase space is small, stable and smooth. With this in mind Sec. 6.8 will be devoted to the recovery of the observables from the migration effects caused by the detector response.
Chapter 6. Event reconstruction and selection

6.5.1 The relation between resolution and purity/stability

Before moving on to the next section, the relation between the concept of purity and the resolution of the observable is investigated. In this section I arrive to a simple formula that characterizes this relation and that can be easily extended to other observables. The relation between the resolution and the concept of stability is analogous to the case of purity discussed here, as can be inferred from the agreement between the purity and stability in Figs. H.3 and H.4 (App. H).

Due to the resolution of the experiment, the events that would be placed in a certain bin by an ideal, perfect detector, might be assigned to a different bin by our real detector system. In Fig. 6.11, an arbitrary resolution curve (blue) is superimposed to a bin whose size is \((a + b)/2\). The x-axis corresponds to an observable \(x\) at GEN level.

The purity would be represented by the part of the resolution distribution that lies within the bin. Following this line of thoughts, the dependence of the purity \(P\) on the resolution function \(f\) could be expressed as:

\[
P = \frac{\int_{-a/2}^{b/2} d\mu \rho(\mu) \int_{-a/2}^{b/2} dx f(x; \mu, \sigma, \ldots)}{\int_{-a/2}^{b/2} d\mu \rho(\mu) \int_{-\infty}^{+\infty} dx f(x; \mu, \sigma, \ldots)},
\]

where \(\mu\) parametrizes the center of the resolution curve, the \(\sigma\) represents the width of the resolution, and \((\ldots)\) stands for additional parameters that could account for possible asymmetries of the resolution distribution, the shape of the tails, or corrections to the width (for instance the parametrization in Eq. 6.5). The density \(\rho(\mu)\) accounts for the fact that the distribution of the observable inside the bin is not necessarily flat. For instance, the \(\Delta \phi_{12}\) distributions for the inclusive 2-jet \(\Delta \phi_{12}\) distributions are steeply falling, as will be observed in Fig. 9.13.

For simplicity let us assume that \(\rho(\mu) = \text{constant}\), which is a better approximation for the inclusive 3-jet \(\Delta \phi_{12}\) observables (Fig. 9.16) than for the inclusive 2-jet distributions
6.5. Purity, stability, acceptance and background

Figure 6.11: Representation of a resolution curve (blue) of an observable \( x \) superimposed to a bin of size \( (a + b)/2 \).

(Fig. 9.13). In addition, since we want an estimation of the dependence of the purity on the resolution parameter \( \sigma \), let us consider that \( f \) is a Gaussian distribution. By setting \( a = b = a' = b' = 1 \) one arrives to a simple expression that can be applied to the bins other than the last bin of the \( \Delta \phi_{12} \) distributions:

\[
P(\sigma) = \text{erf}\left(\frac{1}{\sqrt{2\sigma}}\right) + \sqrt{\frac{2}{\pi}} \sigma \left(\exp\left(-\frac{1}{2\sigma^2}\right) - 1\right)
\]

The expression in Eq. 6.10 cannot be applied to the last bin for the case of \( \Delta \phi_{12} \) since the events originated in that bin can only migrate to lower bins. One can check the validity of Eq. 6.10 by comparing the resulting purity to the values depicted in Fig. H.4 (excluding the last bin) for the case of the inclusive 3-jet observable, for each \( p_T^{\text{max}} \) range. The Gaussian core resolution \( \sigma \) corresponding to each \( p_T^{\text{max}} \) range is extracted from Fig. 6.6b. The resulting ratio of the purity \( (P^{\text{true}}) \) obtained from Fig. H.4 (excluding the last bin) to the prediction from Eq. 6.10, is shown in Fig. 6.12.

As one can observe in Fig. 6.12, the simple expression Eq. 6.10 is in good agreement (differences \( \sim 5\% \) at high \( p_T^{\text{max}} \)) with the purity values depicted in Fig. H.4. The resolution values that correspond to the purity in Fig. H.4 (excluding the last bin), are shown in Fig. 6.6b for each \( p_T^{\text{max}} \) range.

The expression in Eq. 6.10, and the more general Eq. 6.9 can be naturally extended to other observables. This expressions also serves as a validation of resolution studies and can additionally be used to estimate the resolution, differential in the observable itself once the purity is known.

As commented earlier, the expression in Eq. 6.10 cannot be applied to the last bin for the case of \( \Delta \phi_{12} \) since the events originated in that bin can only migrate to lower bins. An analogous expression, applicable to the last bin, can be obtained by imposing that events which 'intend' to migrate to higher bins are counted as part of the last bin. This is achieved by setting \( a = b = a' = 1, b' = +\infty \) in the general expression Eq. 6.9. The resulting purity
Chapter 6. Event reconstruction and selection

Figure 6.12: Ratio of the purity values ($P_{\text{true}}$) obtained from Fig. H.4 (excluding the last bin) to the prediction from Eq. 6.10. The ratio is shown as a function of $p_T^{\text{max}}$.

for the last bin $P_{\text{last}}(\sigma)$ obeys the following simple relation with the purity $P(\sigma)$:

$$\frac{P_{\text{last}}(\sigma) - P(\sigma)}{P(\sigma)} = \frac{1 - P(\sigma)}{2P(\sigma)}$$  \hspace{1cm} (6.11)

One should notice that Eq. 6.11 characterizes the size of the step observed in the purity curves shown in Figs. H.3 and H.4. Let us define the left-hand side of Eq. 6.11 as $\Delta_{\text{true}}^P$ and the right-hand side as $\Delta_P$. In order to check the validity of Eq. 6.11 one can compare $\Delta_{\text{true}}^P$ for each $p_T^{\text{max}}$ region using the purity values in Fig. H.4, with the prediction that results from evaluating $\Delta_P$ with the corresponding $P(\sigma)$ from Eq.6.10. The resulting ratio $\Delta_{\text{true}}^P/\Delta_P$ is shown in Fig. 6.13.

It is quite impressive that the simple expression Eq. 6.11, derived by means of simple approximations like a purely Gaussian shape for the resolution distribution and a constant $\rho(\mu)$ within the bin, is able to accurately predict the size of the step observed in the purity and stability curves, specifically for the case of the inclusive 3-jet distributions in Fig. H.4 at high $p_T^{\text{max}}$. As $p_T^{\text{max}}$ increases, the predictive power of Eqs. 6.11 improves. This is in accordance with the fact that as $p_T^{\text{max}}$ increases, the Gaussian components of the resolution curves become more important (see Figs. H.1 and H.2), hence the Gaussian approximation in the derivation of Eqs. 6.10 and 6.11 becomes more accurate. Additionally, the predictive power of Eqs. 6.10 and 6.11 is higher for the inclusive 3-jet observables depicted in Fig. H.3 than for the 2-jet distributions in Fig. H.3 because the approximation $\rho(\mu) = \text{constant}$ is less accurate for the latter, as will be observed in Figs. 9.13 and 9.16. Nevertheless, the more general Eq. 6.9 remains valid and can be used and extended to other observables.

6.6 Pileup

Due to the high luminosity achieved at the LHC, several interactions per bunch crossing (pileup) can occur. Contributions from events in the same bunch crossing are called in-time
pileup, similarly the contribution from collisions from a different bunch crossing is called out-of-time pileup. The average number of pileup interactions per single bunch crossing in the data is observed to be about 27. It is proportional to the inelastic proton-proton cross section, and also to the instantaneous luminosity per lumi section (time in which the luminosity is considered stable) per bunch crossing. This implies the uncertainty on the luminosity translates into an uncertainty on the pileup.

The observables investigated in this thesis are quite insensitive to the pileup contributions. One reason is the normalization to the 2-jet inclusive cross section which cancels the dependence on the luminosity, and another reason is that the jets are measured in the tracker acceptance region, where the charged hadrons tracks from pileup vertices are removed using the so-called charged-hadron subtraction technique (see Ref. [145]). Also the contribution from neutral hadrons is reduced by applying a jet-area-based correction (see Ref. [151]) as part of the $\eta$ dependent JES factor discussed in Sec. 6.2.3, which subtracts the offset energy of the jets. The pileup removal algorithm has an efficiency of $\sim 99\%$ for jets with $30 < p_T < 50$ GeV and $|y| < 2.5$ (see Ref. [152]). Nevertheless, since we want a very precise measurement of angular variables it is important to assess the impact that the pileup contributions have on the results.

The pileup contribution to an event is simulated by overlapping the interaction with already generated minimum bias events. Since the pileup conditions change with time the generated MC samples have to be adapted to the real conditions. There is an easy way to achieve this by reweighting the distribution of pileup events in the MC as to reproduce the pileup distribution in the data. Since there is a linear dependence between the number of pileup events and the number of reconstructed primary vertices the latter will be used. In Fig. 6.14a one can observe the distributions of primary vertices corresponding to both the simulation before reweighting and the data.

Figure 6.14a shows that the number of primary vertices in the simulation is higher than in

Figure 6.13: Ratio of $\Delta_P^{\text{true}}$ obtained from Fig. H.4, to the prediction that results from evaluating $\Delta_P$ with the corresponding $P(\sigma)$ from Eq.6.10. The ratio is shown as a function of $p_T^{\text{max}}$. 

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure6_13.png}
\caption{Ratio of $\Delta_P^{\text{true}}$ obtained from Fig. H.4, to the prediction that results from evaluating $\Delta_P$ with the corresponding $P(\sigma)$ from Eq.6.10. The ratio is shown as a function of $p_T^{\text{max}}$.}
\end{figure}
Chapter 6. Event reconstruction and selection

Figure 6.14: Distribution of primary vertices in a) the simulation before reweighting and in data, and b) the simulation after reweighting and in data. The average number of primary vertices in data is 15.

The data. Since we want to reproduce the experimental conditions we can sample the number of primary vertices in the simulation and correct the event weight with the factor \( n_{\text{PV}}^{\text{data}} / n_{\text{PV}}^{\text{MC}} \), where \( n_{\text{PV}}^{\text{data}} \) and \( n_{\text{PV}}^{\text{MC}} \) correspond to the number of primary vertices in data and simulation respectively. The procedure can be iteratively applied until the desired agreement with the data is achieved. Figure 6.14b shows the distribution of primary vertices in the data, and in the simulation after the weights are applied. Three iterations were used in order to get the reweighted distribution in Fig. 6.14b.

The same procedure was used in order to estimate the pileup uncertainty, which is subsequently propagated into the observables of interest. The primary vertex distribution follows closely a Poisson distribution with mean value 15. The MC simulation is then reweighted in order to get corresponding distributions with mean values 14 and 16 (6% variation). The resulting primary vertex distributions are shown in Fig. 6.15.

The variations are an overestimation since the uncertainty on the luminosity during the 2016 data taking period is 2.5% (Ref. [153]), and the uncertainty on the proton-proton inelastic cross section is \( \sim 2\% \) (Ref. [154]).

The variations were then propagated into the observables of interest, namely the 2- and 3-jet inclusive cross section as a function of \( \Delta \phi_{12} \). The results are depicted in Fig. 6.16 for two \( p_T^{\text{max}} \) regions. One can observe that a 6% variation of the average number of pileup events translates into less than 0.2% effect for the observables of interest.
6.6. Pileup

Figure 6.15: Reweighted distributions of primary vertices with average number 14, 15 and 16.

Figure 6.16: Ratios of the inclusive a) 2-jet and b) 3-jet cross sections differential in $\Delta\phi_{12}$, obtained with average PV of 14 and 16, to the one obtained by using an average PV of 15. Two $p_T^{\text{max}}$ ranges are shown.
6.7 The measurement at detector level

The inclusive 2-jet distributions differential in $\Delta \phi_{12}$ are shown in Fig. 6.17a for various $p_T^{\text{max}}$ ranges at detector level. The distributions are normalized to the inclusive 2-jet cross section integrated over the full $\Delta \phi_{12}$ range $\sigma^{\text{DET}}$, for the corresponding $p_T^{\text{max}}$ range. The data, represented by markers, are compared with the full simulated predictions from \textsc{Pythia8} tune CUETP8M1 (dotted lines) as well as \textsc{Herwig++} tune CUETHppS1 (solid lines). The statistical uncertainty is represented by error bars (generally small therefore not always visible). The $\Delta \phi_{12}$ distributions are strongly peaked at 180° and become steeper with increasing $p_T^{\text{max}}$. The corresponding ratios of the MC predictions to data are displayed in Fig. 6.17b for ten $p_T^{\text{max}}$ ranges. The predictions used in the ratios are DET level simulations from \textsc{Pythia8} tune CUETP8M1, \textsc{MadGraph} +\textsc{Pythia8} tune CUETP8M1, and \textsc{Herwig++} tune CUETHppS1. The \textsc{MadGraph} +\textsc{Pythia8} event generator performs better than \textsc{Pythia8} and \textsc{Herwig++}.

The inclusive 3-jet distributions are depicted in Fig. 6.18a at detector level for several $p_T^{\text{max}}$ ranges. The full simulated predictions from \textsc{Pythia8} tune CUETP8M1 (dotted lines) as well as \textsc{Herwig++} tune CUETHppS1 (solid lines) are compared to data (markers). The distributions are also normalized to the inclusive 2-jet cross section integrated over the full $\Delta \phi_{12}$ range $\sigma^{\text{DET}}$, for the corresponding $p_T^{\text{max}}$ range. The predictions for the inclusive 3-jet distributions are flatter than those for the inclusive 2-jet distributions. The reason for this behaviour is that exclusive 2-jet topologies contribute mainly at $\Delta \phi_{12} \approx 180^\circ$, making therefore the inclusive 2-jet distributions steeper. The ratios of the predictions from \textsc{Pythia8} tune CUETP8M1, \textsc{MadGraph} +\textsc{Pythia8} tune CUETP8M1, and \textsc{Herwig++} tune CUETHppS1 samples are depicted in Fig. 6.17b. Oppositely to the performance observed for the inclusive 2-jet case, \textsc{Pythia8} and \textsc{Herwig++} perform better than \textsc{MadGraph} +\textsc{Pythia8} for the inclusive 3-jet distributions.
6.7. The measurement at detector level

Figure 6.17: a) The normalised inclusive 2-jet DET level distributions differential in $\Delta \phi_{12}$ overlaid with full simulated predictions from PYTHIA8 MC with tune CUETP8M1 (dotted lines) as well as HERWIG++ tune CUETHppS1 (solid lines). b) Ratios of MC predictions from PYTHIA8 tune CUETP8M1, MadGraph + PYTHIA8 tune CUETP8M1, and HERWIG++ tune CUETHppS1 to data. The error bars indicate the statistical uncertainty.
Figure 6.18: a) The normalised inclusive 3-jet DET level distributions differential in $\Delta\phi_{12}$ overlaid with full simulated predictions from PYTHIA8 MC with tune CUETP8M1 (dotted lines) as well as HERWIG++ tune CUETHppS1 (solid lines). b) Ratios of MC predictions from PYTHIA8 tune CUETP8M1, MadGraph + PYTHIA8 tune CUETP8M1, and HERWIG++ tune CUETHppS1 to data. The error bars indicate the statistical uncertainty.
6.8 Detector effects and unfolding to particle level

The unfolding of the data to particle level is the procedure of correcting the data (DET level) for detector effects, for instance resolution and efficiency. The unfolding of the data is therefore a mapping between the GEN and the DET levels. In other words, when the data is unfolded to particle level one is determining the outcome that an ideal, perfect detector would have given. One should notice that one of the advantages of the unfolding of the data is that it removes the dependency of the measurement on the detector response, meaning that it allows to compare in a common ground, outputs from different experiments.

In order to understand the detector effects on the physics objects, hence the measurement, we simulate the GEN level prediction using MC event generators and subsequently simulate the detector response on the events, resulting in the corresponding DET level prediction. The resulting mapping between the GEN and DET levels can be cast in the following form:

\[ d = \hat{A}g, \] (6.12)

where the DET level measurement \( d \) and the GEN level \( g \) are mapped by the matrix \( \hat{A} \). The expression in Eq. 6.12 can be easily extended to higher dimensional objects. The objects \( d \) and \( g \) can be taken as vectors whose components are the bin contents of the histograms at the respective levels. The matrix \( \hat{A} \) is the normalized response matrix.

Once \( \hat{A} \) is determined from the simulation, it is used in order to find \( g \) given the experimental data \( d \).

6.8.1 Unfolding methods

The different unfolding procedures differ in the way they solve Eq. 6.12. In principle it is a simple linear equation that would have an inversion solution, however in practice the matrix \( \hat{A} \) can be singular. One of the reasons for this behaviour corresponds to the statistics limitation of the simulation and subsequent uncertainty on the entries of the matrix, which can result in a singular output.

If \( \hat{A} \) is a diagonal matrix the solution of Eq. 6.12 will be simply \( g_i = 1/A_{ii}d_i \), where \( i \) indicates the specific bin. This is the so-called correction factor method or bin-by-bin method, and it can be useful when the detector effects provoke negligible migration between bins.

If \( \hat{A} \) is invertible then the solution will be simply \( g = \hat{A}^{-1}d \). Not surprisingly this method is known as the inversion method. Although all the methods try to invert \( \hat{A} \), the inversion method refers to the exact inversion of the matrix.

As discussed in Secs. 6.4 and 6.5, although the migration between bins is stable, the resolution of the observables investigated in this thesis is of the same order as the bin size which is employed. This means that the use of the correction factor method would not be accurate. On the other hand, due to fact that the size of the response matrices is large and that they have low populated regions (as will be observed in Sec. 6.8.2), the inversion method does not give stable results. As a result, the unfolding of the observables under study in this thesis relied on methods that make use of a regularization of the matrix inversion. Specifically the D’Agostini method [155] was used for the unfolding of the observables whereas the Tikhonov regularization method [156] was employed as a consistency check of the unfolding procedure. The two methods will be briefly introduced in the following subsections.
Chapter 6. Event reconstruction and selection

D’Agostini

The D’Agostini unfolding method relies on the Bayes theorem of conditional probabilities and therefore it can be expressed as follows:

\[
P(g_i^{(0)}|d_j) = \frac{P(d_j|g_i^{(0)})g_i^{(0)}}{\sum_k P(d_j|g_k^{(0)})g_k^{(0)}}
\] (6.13)

The conditional probability in Eq. 6.13 that given the DET level \(d_j\) one finds the particle level \(g_i^{(0)}\) is a first approximation of the inverse of \(\hat{A}\) in Eq. 6.12. The procedure can be iteratively used, for instance with the resulting \(P(g_i^{(0)}|d_j)\) one determines a new prior \(g_i^{(1)}\) which is then used to get the next approximation of the inverted response matrix \(P(g_i^{(1)}|d_j)\). The details on the D’Agostini method and its implementation can be found in Refs. [155,157].

Tikhonov regularization

The Tikhonov regularization unfolding (TUnfold) method relies on minimizing the following \(\chi^2\) measure:

\[
\chi^2 = (d - \hat{A}g)^T\hat{V}_{yy}^{-1}(d - \hat{A}g) + \tau^2\|L\|g\|
\] (6.14)

where \(\hat{V}_{yy}\) corresponds to the covariance matrix of the measurement, \(\hat{L}\) is the regularization matrix, and \(\tau\) is corresponds to the so-called regularization parameter. The minimization gives an estimator of the distribution \(g\). The details on the procedure and its implementation can be found in Refs. [156,158].

6.8.2 Response matrices and unfolded distributions

The measured cross sections are corrected for detector smearing effects and unfolded to stable-particle level. We used the D’Agostini method as implemented in the RooUnfold software package (Ref. [157]). The consistency of the unfolding procedure was also checked against the alternative TUnfold package (Refs. [156,158]) that uses a least-square minimisation with Tikhonov regularisation. Both methods gave equivalent results as will be shown in Sec. 6.8.4.

The unfolding procedure parametrizes the measurement effects using a response matrix that maps the true distribution onto the measured one. The construction of the response matrix is derived using two methods:

- Using the official PYTHIA8 tune CUETP8M1 CMS simulated Monte Carlo (MC) samples. Also the HERWIG++ and MADGRAPH responses were used as cross checks and to estimate the model systematic uncertainty in the unfolding procedure (Sec. 7.3).

- A novel method I developed by parametrizing the response matrix according to its projections in the diagonal and anti-diagonal axes (App. D). The method was used to estimate the systematic uncertainty arising from the limited knowledge of the real \(\Delta\phi_{12}\) resolution (Sec. 7.3.1). Even though it is not used for the final unfolded results, its consistency and validity was checked (App. D).
The response matrix in the unfolding procedure is usually derived using simulated Monte Carlo (MC) samples (e.g. Pythia8 tune CUETP8M1). The CMS detector response of the simulation is based on Geant4 [132]. One of the disadvantages of using only the fully simulated MC response matrices is that controlling the systematic effects from the limited knowledge of the $\Delta \phi_{12}$ resolution in data is at least very difficult to achieve.

Figure 6.19: The response matrices derived the official Pythia8 (tune CUETP8M1) CMS MC for the inclusive a) 2-jet and b) 3-jet distributions as a function of $\Delta \phi_{12}$, for one $p_T^{\text{max}}$ range.

The response matrices, derived using the official CMS simulated MC samples from Pythia8 tune CUETP8M1, are depicted in Fig. 6.19 for the inclusive 2- and 3-jet observables. The response matrices for the remaining $p_T^{\text{max}}$ ranges can be found in Figs. I.1 and I.2 (App. I). The matrices are diagonal for both the differential 2-jet (a), and 3-jet (b) cross sections (this property is used in the derivation of the method in the App. D). The ratios of the unfolded data to the data at DET level, for the differential a) 2-jet and b) 3-jet distributions as a function of $\Delta \phi_{12}$ are shown in Fig. 6.20. The unfolding corrections are of the order of a few percent.

As a complement of the discussion in Sec. 6.5 on migration, Fig. 6.21 shows the correlation matrices for the 2- and 3-jet differential observables. Only the lower half of the matrices is depicted since the matrices are symmetric. The correlation coefficients corresponding to the remaining $p_T^{\text{max}}$ ranges can be found in the App. I. The matrices show how the different bins get correlated due to migration of events to and from neighbouring bins. The off-diagonal correlation coefficients drop rapidly and are not strongly oscillating, which is consistent with the stability studies discussed in Sec. 6.5.

Migrations between $p_T^{\text{max}}$ bins can also occur. From the fractional jet $p_T$ resolution curve shown in Fig. 6.3 one can extract the absolute $p_T$ resolution of the leading jet, for instance, when $p_T^{\text{max}}$ is $\sim 200$ GeV the corresponding absolute resolution is $\sim 15$ GeV. As $p_T^{\text{max}}$ increases the corresponding resolution becomes smaller. In comparison to the $p_T^{\text{max}}$ bin sizes used in this thesis (equal or larger than 100 GeV), the $p_T^{\text{max}}$ resolution is small. In addition, the distributions are normalized to the inclusive 2-jet cross section in the corresponding $p_T^{\text{max}}$ bin, which reduces the effects of $p_T^{\text{max}}$ bin migration. Therefore the unfolding of the data is only considered in bins of $\Delta \phi_{12}$.
Chapter 6. Event reconstruction and selection

Figure 6.20: Ratio of the unfolded data to the data at DET level for the inclusive a) 2-jet and b) 3-jet distributions as a function of $\Delta \phi_{12}$. The error bars indicate the statistical uncertainty. The response matrices are derived from the official Pythia8 MC (tune CUETP8M1).
6.8. Detector effects and unfolding to particle level

6.8.3 Regularization

The regularization (number of iterations) of the unfolding procedure was chosen by comparing the difference in $\chi^2$ between data and MC at DET level to that between data and MC at GEN level. As an example, Fig. 6.22 shows the resulting $\chi^2$ curves for four of the $p_T^{\text{max}}$ ranges. The minimum of the curves indicates the iteration number for which the information given by the difference between data and MC at GEN level is equal to the information given by the same difference at DET level. The minimum value does not necessarily correspond to the number of iterations which is finally chosen. For the determination of the final values also the fractional statistical errors are considered.

The conservation of information indicates that the statistical error of the unfolded data should be generally larger than the statistical errors of the DET level data. The number of iterations for the unfolding is considered as the closest value to the minimum of the corresponding $\chi^2$ curve (see the examples in Fig. 6.22), and satisfies, in addition, that the statistical errors after unfolding is larger than the statistical errors before unfolding. The statistical fractional errors of the data before and after unfolding are shown in Fig. 6.23 for the inclusive 2- and 3-jet differential cross sections in one $p_T^{\text{max}}$ range. The fractional errors corresponding to the remaining $p_T^{\text{max}}$ ranges can be found in Figs. I.3 and I.4. The number of iterations which were used for the final results are also indicated in Figs. I.3 and I.4.

6.8.4 D’Agostini vs Tikhonov unfolding

The unfolding procedure performance and consistency is also cross checked with the alternative TUNFOLD package (Ref. [156, 158]) that uses a least square minimisation with Tikhonov regularisation.

The regularization parameter in Eq. 6.14 was determined using the L-curve scan method which is implemented in the TUNFOLD package. The L-curve is defined by the pair of values $(\log \chi^2_A, \log \chi^2_L)$, while $\chi^2_A$ corresponds to the first term on the right-hand side of Eq. 6.14 whereas $\chi^2_L$ corresponds to the regularization term (second term on the right-hand side of the Eq. 6.14). The unfolding was performed for twenty values of $\tau$, and the optimal $\tau$ was chosen.
Chapter 6. Event reconstruction and selection

Figure 6.22: Comparison of the difference in $\chi^2$ between data and simulation at DET level to that between data and simulation at GEN level. The distributions correspond to the inclusive 2-jet observables as a function of $\Delta \phi_{12}$, for four $p_T^{\text{max}}$ regions.

Figure 6.23: The fractional statistical errors of the unfolded and the measured inclusive a) 2-jet and b) 3-jet distributions as a function of $\Delta \phi_{12}$. 
as the one for which the curvature in the L-curve is the highest. The resulting optimal $\tau$ value was found in the range $[4 \times 10^{-4}, 1 \times 10^{-3}]$ for the observables considered in this thesis.

Figure 6.24 shows the ratios of the unfolded data using the Tikhonov method to the unfolded data using the D’Agostini method for the inclusive 2-jet distributions as a function of $\Delta \phi_{12}$. Four $p_T^{\text{max}}$ ranges are depicted as examples. As observed in Fig. 6.24, the results from the D’Agostini and Tikhonov unfolding methods are compatible within the statistical uncertainty.

The final results of this thesis are obtained using the D’Agostini method.

6.8.5 Migration matrices obtained with a new method

By applying the novel method described in App. D I was able to obtain alternative response matrices that can be used to unfold the data. The method uses the symmetry properties of the response matrices in order to reproduce the migration and correlation between bins far away from the diagonal. It uses toy MC events that can be generated with unlimited statistics. An advantage of the method is that the fluctuations which usually affect the stability of the unfolding are drastically reduced, and also that the characteristics of the new response matrices are controlled by a few parameters. The latter feature allows the estimation of the systematic uncertainty arising from the limited knowledge of the $\Delta \phi_{12}$ resolution.
Chapter 6. Event reconstruction and selection

in the data (Sec. 7.3.1).

As mentioned in App. D the new method has important advantages with respect to the procedure which is currently used in CMS called toy MC method or forward smearing. The toy MC method is based on the use of the resolution curves obtained from the CMS fully simulated MC samples to smear the GEN level distribution, and obtain the corresponding toy response matrix. The main disadvantage of this method is the non-trivial form of the resolution curves, generally asymmetric and varying differentially in the observable, which makes the fitting difficult specially for the tail of the distributions.

Figure. 6.25a shows the original response matrix obtained from the fully simulated Pythia8 samples for the differential 2-jet distributions, for one \( p_T^{\text{max}} \) range. Figure 6.25b depicts the response matrix derived by means of the new method introduced in App. D. As can be observed in Fig. 6.25c, the resulting ratio between the two matrices is around 1, even far away from the diagonal. Similar ratios corresponding to the remaining \( p_T^{\text{max}} \) ranges are shown in App. D. Using the alternative, parametrized response matrices, the corresponding DET level distributions are derived by folding the original GEN level with the new response matrices. The method was used to estimate the systematic uncertainty arising from the limited knowledge of the \( \Delta \phi_{12} \) resolution in the data (Sec. 7.3.1).

A consistency check was done in order to assess the validity of the response matrices determined by the new method. Figure 6.26 shows the ratio of the unfolded data using the alternative response matrices to the unfolded data using the original Pythia8 response matrices. The ratio is consistent with 1 within the unfolding uncertainty (estimated in Sec. 7.3). Figure 6.26 is not only a test of the validity of the method derived in the App. D, but also of the unfolding procedure itself. The unfolded data using the response from MadGraph full simulation is also shown in Fig. 6.26, and it also agrees, within the unfolding systematic uncertainty, with the unfolded data using the Pythia8 and the alternative response matrices. The determination of the unfolding systematic uncertainty (solid band in Fig. 6.29) will be discussed in Sec. 7.3.
Figure 6.25: a) Response matrix from the CMS fully simulated Pythia8 MC sample, b) response matrix obtained with the new method, and c) the ratio of the new matrix to the original one. The matrices correspond to the inclusive 2-jet distributions as a function of $\Delta \phi_{12}$ for $500 < p_T^\text{max} < 600$ GeV. Similar ratios corresponding to the remaining $p_T^\text{max}$ ranges are shown in App. D.
Figure 6.26: Ratio of the unfolded data using the MadGraph response and the response derived using the method described in the App. D, to the unfolded data using the Pythia8 generator response. The solid band represents the unfolding systematic uncertainty and the error bars represent the statistical uncertainty.
6.8. Detector effects and unfolding to particle level

6.8.6 Validation of the unfolding

Four tests to the unfolding procedure are performed in order to demonstrate the consistency of the unfolding method and the validity of the software. In this section only the unfolding tests corresponding to the inclusive 2-jet distributions as a function of $\Delta \phi_{12}$ are presented. The plots corresponding to the unfolding tests of the inclusive 3-jet distributions as a function of $\Delta \phi_{12}$ can be found in App. I.

Internal closure test

The detector (reconstructed) level of the MC simulation is unfolded using the same MC sample. The result is then compared to the stable particle level of the simulation. The results, using the official MC Pythia8 (tune CUETP8M1) are shown in Figs. 6.27 for the case of the inclusive 2-jet observables, and in J.1 for the case of the inclusive 3-jet distributions. The resulting ratios are consistent with 1.

![Diagram](image-url)

Figure 6.27: Internal closure test for the unfolding of the inclusive 2-jet distributions as a function of $\Delta \phi_{12}$ using the official Pythia8 (tune CUETP8M1) MC samples. The detector level predictions from the MC simulation are unfolded using the same MC sample and compared to the corresponding stable particle level predictions.
Backfolding test

Using the response matrices from the MC simulation, the unfolded distributions are folded back and then compared to the distributions at reconstructed level. The results, using the official MC Pythia8 (tune CUETP8M1) are shown in Fig. 6.28 for the inclusive 2-jet observables, and in Fig. J.2 for the inclusive 3-jet distributions. The resulting ratios are consistent with a flat behaviour at 1, within the statistical uncertainties.

Figure 6.28: Backfolding test for the inclusive 2-jet distributions differential in $\Delta \phi_{12}$. The unfolded distributions are folded back and then compared to the distributions at reconstructed level. The official Pythia8 (tune CUETP8M1) MC samples are used.
### 6.8. Detector effects and unfolding to particle level

**Bottom line test**

As a check of consistency of the procedure, the level of agreement between the data and the model is investigated at both stable particle level (GEN) and reconstructed level (DET). In the case in which the unfolding procedure is able to correct the measurement without making use of any regularisation, the level of agreement at both levels should be the same. In our case, we use a regularized unfolding procedure (D’Agostini method), so the difference between the model and the data at DET level should remain small and generally larger than the same difference at GEN level. The results, using the official MC Pythia8 tune CUETPSM1 are shown in Tabs. 6.3 and J.1.

<table>
<thead>
<tr>
<th>$p_T^{\text{max}}$ [GeV]</th>
<th>$\chi^2/\text{ndf}$ unfolded</th>
<th>$\chi^2/\text{ndf}$ folded back</th>
</tr>
</thead>
<tbody>
<tr>
<td>200-300</td>
<td>25.429</td>
<td>25.675</td>
</tr>
<tr>
<td>300-400</td>
<td>30.405</td>
<td>30.774</td>
</tr>
<tr>
<td>400-500</td>
<td>33.145</td>
<td>39.127</td>
</tr>
<tr>
<td>500-600</td>
<td>29.921</td>
<td>38.908</td>
</tr>
<tr>
<td>600-700</td>
<td>156.727</td>
<td>202.672</td>
</tr>
<tr>
<td>700-800</td>
<td>67.979</td>
<td>94.788</td>
</tr>
<tr>
<td>800-1000</td>
<td>42.868</td>
<td>62.018</td>
</tr>
<tr>
<td>1000-1200</td>
<td>12.076</td>
<td>16.548</td>
</tr>
<tr>
<td>&gt;1200</td>
<td>6.275</td>
<td>8.531</td>
</tr>
</tbody>
</table>

Table 6.3: Bottom line test for the inclusive 2-jet observables as a function of $\Delta \phi_{12}$. The value of $\chi^2/\text{ndf}$ of the unfolded data compared to the Pythia8 GEN level (middle column), and $\chi^2/\text{ndf}$ of the DET level data compared to the Pythia8 DET level (right column). The official MC Pythia 8 (tune CUETPSM1) MC simulation is used.
Chapter 6. Event reconstruction and selection

External closure test

The detector level (reconstructed) of a MC simulation A is unfolded using another, independent MC simulation B. The result is then compared to the stable particle level of the MC simulation A. The results, using the official MC PYTHIA8 tune CUETP8M1 to unfold the distributions from MADGRAPH +PYTHIA8 tune CUETP8M1 are shown in Fig. 6.29 for the inclusive 2-jet observables, and in Fig. J.3 for the inclusive 3-jet distributions. The observed agreement (within the unfolding systematic uncertainties) between the unfolded distributions further supports the consistency and validity of the unfolding procedure. The determination of the unfolding systematic uncertainty (solid band in Fig. 6.29) will be discussed in Sec. 7.3.

Figure 6.29: External closure test corresponding to the inclusive 2-jet distributions as a function of $\Delta\phi_{12}$. The DET level predictions from the official MC PYTHIA8 (tune CUETP8M1) MC simulation are unfolded using another, independent MC simulation (official MADGRAPH +PYTHIA8 tune CUETP8M1). The result is then compared to the stable particle level (GEN) from the MC PYTHIA8 (tune CUETP8M1) simulation. The solid band represents the unfolding systematic uncertainty.
An experimental measurement is likely to be meaningless if the accuracy at which the result is obtained is not specified. One of the sources determining the accuracy of an experiment is the statistical uncertainty, which depends on the amount of events that fall into the phase space region being probed and that we are able to detect. It is driven mainly by the corresponding cross section, the luminosity, the efficiency of the detector system and even on how the phase space is defined. The other main origin of uncertainty is the systematic source. It is not directly dependent on the amount of events that are measured, instead it is determined by the inaccuracy of the detector system and also it accounts for biases that we might arbitrarily introduce to the measurement, like for instance, when a specific theoretical model is used to treat the data. In the following sections, the main sources of systematic uncertainty affecting the observables presented in this thesis are discussed, followed by a discussion on the statistical uncertainty on the measurement.

7.1 Jet energy scale

In the last chapter the importance of the jet calibration procedure was discussed. The overall correction applied to the jets suffers from the intrinsic uncertainty on the values of
Chapter 7. Experimental uncertainties

the different factors. This implies that a jet generated with certain $p_T$ can be measured with a shifted $p_T$ governed by the uncertainty in the correction factor that is applied.

The JES uncertainty has been studied in Ref. [145] at 8 TeV, and in Ref. [159] at 13 TeV of center-of-mass energy. For jets with $p_T$ larger than 100 GeV and $|\eta| < 2.5$ the uncertainty on the JES factors is $\sim 1\%$. On the other hand, jets with $p_T \sim 30$ GeV are affected by an uncertainty on the calibration that ranges from 4\% at $|\eta| \sim 2.5$ down to 3\% at $|\eta| \sim 0$. This implies that the impact of the energy calibration uncertainty on the 2-jet observables at high $p_T$ is expected to be smaller than on the inclusive 3-jet observables for which a jet with $p_T > 30$ GeV is required.

The JES uncertainty is propagated to the observables under consideration in this thesis by shifting the $p_T$ of the measured jets up and down using the corresponding JES uncertainty. The jets are then reordered in $p_T$ and the uncertainty on the observable is estimated as the absolute difference of the result given by the central value compared to the corresponding variations.

The resulting JES uncertainty is depicted in Fig.7.1 for one $p_T^{\text{max}}$ range as an example. As a general feature, the impact of the JES uncertainty is smaller for the inclusive 2-jet observables compared to the inclusive 3-jet distributions. One of the reasons for this behaviour is that the distributions are normalized to the inclusive 2-jet integrated cross section, and that a low $p_T$ jet is required for the case of the inclusive 3-jet distributions.

![Figure 7.1: JES uncertainty for the inclusive a) 2-jet and b) 3-jet cross section differential in $\Delta \phi_{12}$, for one $p_T^{\text{max}}$ range.](image)

The JES uncertainty for the remaining $p_T^{\text{max}}$ ranges is depicted in Fig.7.6 for the inclusive 2-jet observables. Towards high $\Delta \phi_{12}$ it reaches its maximum of $\sim 0.4\%$ at low $p_T^{\text{max}}$ (Fig.7.6a), up to values of $\sim 1\%$ for the high $p_T^{\text{max}}$ ranges (Fig.7.6b).

Unlike the 2-jet case, for the inclusive 3-jet observables the impact of the JES factors uncertainty is large, even being the leading source of experimental uncertainty for $200 < p_T^{\text{max}} < 400$ GeV (Fig.7.7a). It reaches values of $\sim 3\%$ at low $p_T^{\text{max}}$ (Fig.7.7a) down to $\sim 1\%$ for higher $p_T^{\text{max}}$ ranges (Fig.7.7b).

As commented in Sec. 6.2.3, the JES factors determined in 2016 were unexpectedly large for $\eta > 2.5$. The problem was related to the transparency degradation in ECAL crystals at
high $|\eta|$ (see Refs. [146, 147]). The studies presented in this thesis were not affected by this issue because the reconstructed jets are selected with rapidity $|y| < 2.5$ (see Sec. 6.3).

### 7.2 Jet energy resolution

In the previous chapter the influence of the detector response on the measured $p_T$ was discussed: a jet generated with a certain $p_T$ is then reconstructed with a different transverse momentum due to the detector response. The jet energy resolution (JER) is measured in data by using the jet response in di-jet and $Z$+jet events (Ref. [145]), similar to the discussion in Sec. 6.2.3 on the determination of the jet energy scale factors.

Since the JER that is simulated depends on the theoretical model and on the simulation of the detector response, it does not necessarily correspond to the JER that is measured in data. Therefore the CMS JetMET group also performs dedicated studies to evaluate the correction factors that should be applied to the simulated JER in order to reproduce the JER present in the data (Ref. [160]). This correction factors possess an intrinsic uncertainty that can be a significant source of systematic effects for the observables under study. The JER scaling factors as a function of $|\eta|$ are shown in Tab. 7.1 together with the corresponding uncertainty.

| $|\eta|$ region | scaling factor ± uncertainty |
|----------------|----------------------------|
| 0.000 - 0.522  | 1.1595 ± 0.0645            |
| 0.522 - 0.783  | 1.1948 ± 0.0652            |
| 0.783 - 1.131  | 1.1464 ± 0.0632            |
| 1.131 - 1.305  | 1.1609 ± 0.1025            |
| 1.305 - 1.740  | 1.1278 ± 0.0986            |
| 1.740 - 1.930  | 1.1000 ± 0.1079            |
| 1.930 - 2.043  | 1.1426 ± 0.1214            |
| 2.043 - 2.322  | 1.1512 ± 0.1140            |
| 2.322 - 2.500  | 1.2963 ± 0.2371            |

Table 7.1: JER scale factors recommended by the JetMET CMS group and their uncertainty as a function of $|\eta|$ (Ref. [160]).

The JER obtained from the simulated samples was shown in Figs. 6.2 and 6.3. No big difference between the JER obtained using the Pythia8 and the MadGraph MC generators was observed (Fig. 6.3). In order to propagate the uncertainty on the scaling factors to the observables under study the JER extracted from Pythia8 was scaled according to the factors depicted in Tab. 7.1. The JER effect on the observables is then estimated by smearing the $p_T$ of the generated jets with a Gaussian distribution with a width corresponding to the nominal value of the JER given by the factors in Tab. 7.1, and the nominal value shifted up and down by the uncertainty on the JER depicted also in Tab. 7.1. The resulting uncertainty is then
estimated as the absolute difference of the result given by the nominal value compared to the corresponding variations.

As an example, the resulting JER uncertainty is depicted in Fig. 7.2 for one $p_T^{\text{max}}$ range. The impact of the JER uncertainty is in general smaller for the case of the inclusive 2-jet distributions compared to the inclusive 3-jet observables.

![Figure 7.2: JER uncertainty for the inclusive a) 2-jet and b) 3-jet cross section differential in $\Delta\phi_{12}$, for one $p_T^{\text{max}}$ range.](image)

The effect of JER uncertainty on the inclusive 2- and 3-jet distributions for the rest of the $p_T^{\text{max}}$ ranges is shown in Figs. 7.6 and Figs. 7.7 respectively. In comparison to the contributions to the total experimental uncertainty from the other sources, the uncertainty from the JER is negligible for all the $p_T^{\text{max}}$ ranges.

### 7.3 Unfolding

Two different sources of systematic uncertainties arise from the unfolding procedure described in the previous chapter: the systematic uncertainty due to the limited knowledge of the real $\Delta\phi_{12}$ resolution, and the systematic uncertainty due to the modelling of the detector response.

#### 7.3.1 $\Delta\phi_{12}$ resolution

The uncertainty arising from the limited knowledge of the $\Delta\phi_{12}$ resolution in data is estimated using the method developed in the App. D. The detector response is reproduced according to a parametrization of its diagonal and anti-diagonal projections. The migration is caused by the elements off the anti-diagonal axis. One should remember that the $\Delta\phi_{12}$ resolution curves are nothing else than projections of the response matrix onto its $x$-axis. Therefore the diagonal projections of the response matrix are also related to the $\Delta\phi_{12}$ resolution ($\sigma_{\text{diagonal}} \approx \sqrt{2}\sigma$, where $\sigma$ represents the $\Delta\phi_{12}$ resolution). The JetMET group recommends $\pm 10\%$ as a conservative estimate for the evaluation of the systematic errors due to the jet resolution effects [149,161].
Variations of the original response matrices were generated by using the parametrization of the detector response described in the App. D. The Gaussian core of the diagonal projection was varied in ±10%. Subsequently, the matrix variations were used to fold the GEN level, which resulted in new DET level distributions which therefore carry the migration governed by the variations. The data were then unfolded using the variations of the detector response (the matrix variations together with the corresponding DET level distributions). Finally, the $\Delta\phi_{12}$ resolution uncertainty is given by the difference of the resulting unfolded distributions compared to the unfolded data using the original response.

The procedure described in App. D and in the previous paragraph is a new, general methodology to estimate the uncertainty due to migration. It can be also extended to other observables. Moreover, it is independent from the unfolding procedure. The main advantage of this method lies in the fact that the correlations given by the detector response are consistently propagated.

The resulting uncertainty from the limited knowledge of the $\Delta\phi_{12}$ resolution in the data is depicted in Fig. 7.3 for one $p_T^{\text{max}}$ range as an example. In general the impact of the $\Delta\phi_{12}$ resolution uncertainty increases with the increase of $\Delta\phi_{12}$. This can be understood in terms of the migration of events caused by the $\Delta\phi_{12}$ resolution, whose effects are higher at high $\Delta\phi_{12}$ where there is less compensation of events going in and out of the bins compared to the case of lower $\Delta\phi_{12}$ bins.

The uncertainty arising from the limited knowledge of the $\Delta\phi_{12}$ resolution for the remaining $p_T^{\text{max}}$ ranges is depicted in Fig. 7.6 for the inclusive 2-jet observables. It constitutes one of the dominant source of uncertainty for the low $p_T^{\text{max}}$ ranges and high $\Delta\phi_{12}$ with values of up to 1% (Fig. 7.6a), whereas for high $p_T^{\text{max}}$ (Fig. 7.6b) it does not contribute significantly to the total experimental uncertainty (less than 0.3%). This observation agrees with the fact that the $\Delta\phi_{12}$ resolution becomes better towards high $p_T^{\text{max}}$ (Sec. 6.4). The decrease of the uncertainty towards low $\Delta\phi_{12}$ is consistent with the shape of the distribution, which becomes less steep towards low $\Delta\phi_{12}$ (as will be observed in Figs. 9.13 and 9.13).

Analogously, Fig. 7.7 shows the uncertainty from the limited knowledge of the $\Delta\phi_{12}$ resolution for the rest of the $p_T^{\text{max}}$ ranges for the inclusive 3-jet observables. In this case, the
\[\Delta \phi_{12}\] resolution does not constitute a main source of systematic uncertainty.

### 7.3.2 Model dependence

The unfolding procedure brings another source of systematic uncertainty: the model dependence of the procedure. Even if the Gaussian core resolution is similar between different models (Fig. 6.6), small differences in the migration can still arise from sources like the exact form of the resolution tails, the steepness of the response matrix and the exact shape of the generated distributions. The model dependence is taken care of by evaluating the difference between the data unfolded with the PYTHIA8 MC generator response compared to the unfolded data using the HERWIG++ and MADGRAPH MC generators. The HERWIG++ generator gives the largest difference since, unlike the MADGRAPH generator, it also models the PS differently in comparison to PYTHIA8. Therefore the absolute difference between the data unfolded with the PYTHIA8 and HERWIG++ event generators is considered as the systematic uncertainty due to the model dependence.

As an example, the resulting model dependence uncertainty is depicted in Fig.7.4 for one \(p_T^{\text{max}}\) range. One of the features which is observed in general is the flat behaviour of the model dependence uncertainty.

![Figure 7.4: Model dependence uncertainty for the inclusive a) 2-jet and b) 3-jet cross section differential in \(\Delta \phi_{12}\), for one \(p_T^{\text{max}}\) range.](image)

The resulting uncertainty for the remaining \(p_T^{\text{max}}\) ranges is depicted in Fig. 7.6 for the inclusive 2-jet observables. The model dependence constitutes the main source of systematic uncertainty except for the first \(p_T^{\text{max}}\) range and the \(\Delta \phi_{12}\) resolution uncertainty dominates towards large \(\Delta \phi_{12}\). The model dependence systematic uncertainty generally increases with \(p_T^{\text{max}}\), from \(\sim 1\%\) at low \(p_T^{\text{max}}\) (Fig. 7.6a) up to \(\sim 2.5\%\) at high \(p_T^{\text{max}}\) (Fig. 7.6b).

Similarly, Fig. 7.7 depicts the model for the inclusive 3-jet distributions. Unlike the 2-jet case, the model dependence is only dominant at high \(p_T^{\text{max}}\) values (Fig. 7.7b). Its contribution ranges from \(\sim 1.5\%\) at low \(p_T^{\text{max}}\) up to 3% for the high \(p_T^{\text{max}}\) ranges.
7.4 Statistical uncertainty

The statistical uncertainty of the data is modified by the unfolding procedure as can be observed in the fractional errors comparisons in Figs. 6.23, I.3 and I.4. The statistical uncertainty resulting after unfolding includes contributions from correlated bins, therefore it accounts for event migration effects.

![Graphs showing statistical uncertainty for inclusive 2-jet and 3-jet cross section](image)

Figure 7.5: Statistical uncertainty for the inclusive a) 2-jet and b) 3-jet cross section differential in $\Delta \phi_{12}$, for one $p_T^{\text{max}}$ range.

The statistical uncertainty corresponding to the 2-jet observables for the remaining $p_T^{\text{max}}$ ranges is depicted in Figs. 7.6. For low $p_T^{\text{max}}$ ranges and low $\Delta \phi_{12}$ values it constitutes the second largest experimental uncertainty after the model dependence. At low $p_T^{\text{max}}$, the statistical uncertainty ranges from ~1.6% at low $\Delta \phi_{12}$ down to ~0.4% at high $\Delta \phi_{12}$, where the statistics is larger. For $p_T^{\text{max}} > 700$ GeV the statistical uncertainty also constitutes the second largest experimental uncertainty after the model dependence, reaching values of ~2% at low $\Delta \phi_{12}$ down to 0.5% at high $\Delta \phi_{12}$ for the case of $p_T^{\text{max}} \sim 1$ TeV.

In the case of the 3-jet observables depicted in Figs. 7.7 for the remaining $p_T^{\text{max}}$ ranges, the statistical uncertainty only contributes significantly at $p_T^{\text{max}} \sim 1$ TeV, where it constitutes the second largest experimental uncertainty after the model dependence, reaching values of up to 2% at low $\Delta \phi_{12}$.

7.5 Experimental uncertainties overview

The experimental uncertainties that result from the JES, JER, $\Delta \phi_{12}$ resolution, model dependence, and the statistics limitation are compared in Figs. 7.6 and 7.7. The distributions correspond to the inclusive 2-jet (Fig. 7.6) and the inclusive 3-jet (Fig. 7.6) observables respectively. The sources of experimental uncertainty are assumed uncorrelated, and the total experimental uncertainty is determined by summing over the individual sources in quadrature. The total experimental uncertainty is also shown in Figs. 7.6 and 7.7.
Chapter 7. Experimental uncertainties

Figure 7.6: Experimental uncertainty sources for the inclusive 2-jet cross section differential in $\Delta \phi_{12}$, for a) $200 < p_T^{\text{max}} < 600$ GeV, and b) $p_T^{\text{max}} > 600$ GeV.
Figure 7.7: Experimental uncertainty sources for the inclusive 3-jet cross section differential in $\Delta \phi_{12}$, for a) $200 < p_{T}^{\text{max}} < 600$ GeV, and b) $p_{T}^{\text{max}} > 600$ GeV.
In Chapter 3 we have observed that the cross section in a perturbative QCD calculation depends on various parameters. Some of these parameters, like the renormalization scale $\mu_r$ and factorization scale $\mu_f$, are non-physical reminders that the perturbation series has been truncated. The strong coupling $\alpha_s$ itself also suffers from an uncertainty which propagates into the calculation: the precision at which it was measured. Additional uncertainties of a QCD cross section calculation arise from the parametrization of the PDF, both from the determination of the parametrization by fitting the experimental data, and from the chosen parametrization (the number of parameters for instance). Moreover, in the PS, the subsequent splittings of the partons introduce additional vertices, and the scale at which the corresponding $\alpha_s$ is evaluated has to be chosen. This can be regarded as an additional source of uncertainty from the resummation done by the PS. Uncertainties from non-perturbative origins are also present. For instance, in MC generators the non-perturbative part of the simulation is modelled and parametrized, and the uncertainty on the determination of the parameters can also propagate into the final result.

Various sources of theoretical uncertainty of perturbative origin are presented in this chapter, followed by a discussion on the impact of the non-perturbative uncertainty sources on the results.

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Scale dependence</td>
<td>118</td>
</tr>
<tr>
<td>8.2 Parton distributions</td>
<td>118</td>
</tr>
<tr>
<td>8.3 Strong coupling</td>
<td>120</td>
</tr>
<tr>
<td>8.4 Parton showering</td>
<td>120</td>
</tr>
<tr>
<td>8.5 Non-perturbative sources</td>
<td>123</td>
</tr>
<tr>
<td>8.6 Comparison between theoretical and experimental uncertainties</td>
<td>126</td>
</tr>
</tbody>
</table>
8.1 Scale dependence

The truncation of the perturbative series in Eq. 3.28 constitutes a source of uncertainty due to the missing terms in the fixed-order calculation. Two arbitrary scales are introduced as a remainder: the renormalization scale $\mu_r$ and the factorization scale $\mu_f$. These scales are normally set to the scale of the hard process. The dependence on the unphysical scales is reduced when including higher terms (at least for inclusive observables). Once set to the scale of the process, $\mu_r$ parametrizes the physics well above this scale (absorption of ultraviolet divergences) and $\mu_f$ encodes the physics down to the proton scale (absorption of collinear singularities in the initial state). The residual dependence of the observables on the arbitrary scales is used as an estimator of the uncertainty from the missing higher-order contributions.

The uncertainties were evaluated by using the NLO Powheg-2j generator. One should remember that the lowest non-trivial order for the $\Delta \phi_{12}$ distribution is given by the real contribution ($2 \rightarrow 3$ LO) of the $2 \rightarrow 2$ NLO calculation, except for $\Delta \phi_{12} = 180^\circ$ where the virtual corrections are included for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$ (see Sec. 4.1). The default $\mu_r = \mu_f = p_T$ of the underlying Born configuration was varied between $p_T/2$ and $2p_T$ in the following seven combinations: $(\mu_r/p_T, \mu_f/p_T) = \{(1/2, 1/2), (1/2, 1), (1, 1/2), (1, 1), (1, 2), (2, 1), (2, 2)\}$. The factors 1/2 and 2 are arbitrary and follow the prescription of choosing the lower and upper variations of the same order of that of the central value. The impact of the $\mu_r$ and $\mu_f$ variations on the 2- and 3-jet inclusive cross distributions is depicted in Fig. 8.1. The envelope (blue band) of the scale variations (red curves) represents the scale uncertainty on the observables. The uncertainty lies below 0.5% for both the 2- and 3-jet observables, even when the statistical uncertainty of the predictions (represented as error bars) are considered.

8.2 Parton distributions

The PDF, which enters the cross section formula (Eq. 3.27), has non-perturbative initial conditions encoded in its parametrization, that have to be, at the present, determined from data. Consequently the PDF is affected by the uncertainty on the experimental data from which it was determined, as well as from the parametric form of the fit to the data (Ref. [162]). Additionally, the evolution of the PDF to higher scales lies in the perturbative regime as was discussed in Sec. 3.1.3. This implies that the perturbative evolution of the PDF is affected by the uncertainty on the strong coupling. These sources of uncertainty propagate to the observables of interest. In this section the propagation of the parametrization and experimental PDF uncertainty to observables will be discussed. The uncertainty from the strong coupling value, which also enters the hard process will be discussed in the next section. It should also be mentioned that the PDF enters three places in the calculation of the cross section: in the convolution with the ME (Eq. 3.27), in the Sudakov governing the ISR (Eq. 3.19), and also the MPI modelling (Eq. 3.33). As a result, different PDFs at different accuracies may be used throughout the same calculation, implying that the concept of PDF uncertainty impact on the observable is at least ambiguous. In this section, only the propagation of the uncertainties of the PDF that is convoluted with the ME is being investigated. The other two sources are covered by the PS and non-perturbative uncertainty sources which are also discussed in this chapter.

The Neural Network based NNPDF3.0 NLO PDF set (Ref. [163]) was used to evaluate
Figure 8.1: Ratios of the a) 2-jet, and b) 3-jet distributions obtained by evaluating $(p_T^\text{max}/p_T, \mu_T/p_T) = (1/2, 1/2), (1/2, 1), (1/1/2), (1, 1/2), (1, 1),(1,2),(2,1),(2,2)$, to the nominal ones obtained by using the central choice $(p_T^\text{max}/p_T, \mu_T/p_T) = (1, 1)$. The envelope (solid blue band) represents the scale variation uncertainty whereas the error bars represent the statistical uncertainty of the predictions.
Chapter 8. Theoretical uncertainties

the impact of the uncertainty to the observables of interest. The NNPDF collaboration provides a group of equally probable PDF replicas, obtained in a way that they reproduce the uncertainty on the observables used to fit the PDF. The replicas can be therefore used to estimate the propagation of this uncertainty to the observables discussed in this thesis. The cross section is calculated for each of the 100 replicas of the NNPDF3.0 NLO PDF set and the uncertainty is evaluated as the resulting envelope. The PDF is used together with the NLO POWHEG-2j + PYTHIA8 event generator.

Figures 8.2a and 8.2b show the resulting PDF uncertainty for the 2- and 3-jet inclusive observables respectively, as the envelope of the NNPDF replicas (red curves). For the case of the inclusive 2-jet distributions in Fig. 8.2a, the uncertainty increases with $p_{T}^{\text{max}}$ from less that 1% at low $p_{T}^{\text{max}}$ up to 2% for high $p_{T}^{\text{max}}$ ranges, except for the last $\Delta \phi_{12}$ bin for which it reaches up to 5% for $p_{T}^{\text{max}} > 1200$ GeV. Similarly, for the case of the inclusive 3-jet observables in Fig. 8.2b, the resulting uncertainty increases with $p_{T}^{\text{max}}$ from less that 1% at low $p_{T}^{\text{max}}$ up to 2% for high $p_{T}^{\text{max}}$ ranges.

8.3 Strong coupling

As mentioned in the previous section, the strong coupling $\alpha_s$ used in the evolution of the PDF carries an uncertainty corresponding to the accuracy at which it was measured. Conventionally, $\alpha_s$ is measured at the scale of the $Z$ boson mass ($\alpha_s(M_Z)$). The propagation of the $\alpha_s$ uncertainty can be estimated at 68% confidence level by varying $\alpha_s(M_Z)$ by 0.001 up and down as recommended in Ref. [164]. The impact of the $\alpha_s$ variation on the inclusive 2- and 3-jet distributions is shown in Fig. 8.3, where the envelope (blue band) of the up and down variations (red curves) represents the resulting uncertainty on the observables. The uncertainty lies below 0.1% for both the inclusive 2- and 3-jet $\Delta \phi_{12}$ distributions.

8.4 Parton showering

As discussed in previous chapters the PS provides an all-order resummation of large logarithms. There are various sources of uncertainties that stem from a PS algorithm (Ref. [69]): sub-leading perturbation terms, the recoil scheme, the parameters entering the shower algorithm (for example the cut-off scale at which the PS is truncated). In the present section we focus on the first source of PS uncertainty mentioned earlier: the uncertainty stemming from the sub-leading logarithms which are not exponentiated in the Sudakov form factor. In the following, "PS uncertainty" stands only for the perturbative uncertainty source.

The study was performed using the LO MC generator PYTHIA8. The PS uncertainty was estimated by varying the default renormalization scale choice for the $\alpha_s$ evaluation at the branching vertices $\mu_r = p_T$, between $\mu_r/2$ and 2$\mu_r$ for both the initial and final state radiation. The following nine combinations were considered: $(\gamma_1^{\text{ISR}}/p_T, \gamma_1^{\text{FSR}}/p_T) = \{(1/2, 1/2), (1/2, 1), (1/2, 2), (1, 1/2), (1, 1), (1, 2), (2, 1/2), (2, 1), (2, 2)\}$. Figures 8.4a and 8.4b show the resulting PS uncertainty for the 2-jet and 3-jet inclusive observables as the envelope (blue band) of the different scale combinations (red curves).

For the inclusive 2-jet distributions the uncertainty from PS is dominant compared to the other sources discussed in this section, reaching 5% at high $\Delta \phi_{12}$ values. At low $\Delta \phi_{12}$ the PS uncertainty increases with $p_{T}^{\text{max}}$, from 1% at low $p_{T}^{\text{max}}$ to 3% for high $p_{T}^{\text{max}}$ ranges.
Figure 8.2: Ratios of the a) 2-jet, and b) 3-jet distributions obtained from the variations of NNPDF3.0 NLO PDF set to the nominal choice. The envelope of the variations (red curves) represents the uncertainty from the PDF.
Figure 8.3: Ratios of the a) 2-jet, and b) 3-jet distributions obtained using $\alpha_s(M_Z) = 0.117, 0.118, 0.119$, to the nominal ones obtained by using the central value $\alpha_s(M_Z) = 0.118$. The envelope (solid blue band) represents the uncertainty from the $\alpha_s(M_Z)$ choice.
8.5. Non-perturbative sources

In the case of the inclusive 3-jet distributions, the uncertainty from PS is significantly reduced at high $\Delta \phi_{12}$ compared to the 2-jet case, with values of $\sim 1\%$ for all the $p_T^{\text{max}}$ ranges. This difference is due to the fact that by requiring a third, softer jet the soft and collinear region is avoided. At smaller $\Delta \phi_{12}$ the pattern is the same as the 2-jet since a recoil from a softer jet ($\sim 30$ GeV) is not enough to produce these $\Delta \phi_{12}$ values (see Fig. 4.2).

8.5 Non-perturbative sources

As discussed in previous chapters, the UE activity and the hadronization lie beyond the perturbative regime and their simulation rely on phenomenological models. Such models depend on a group of parameters that have to be tuned to the UE and hadronization sensitive data. In this section the impact of model dependence of the UE and the hadronization is investigated. Instead of varying the parameters of the models, we compared the predictions obtained from five different tunes determined by the CMS and ATLAS collaborations, which provide a wide range in the parameter space. For this study, the PYTHIA8 event generator was used. Table 8.1 depicts the parameters corresponding to the CMS tunes CUETP8M1 [109], CP1 [110], CP2 [110], Monash [111], 4C [112], and the ATLAS tune A14 [113]. All the tunes make use of the color reconnection MPI model discussed in Sec. 3.2.6. The tunes CUETP8M1, Monash, A14 and 4C employ the exponential $b$ profile in Eq. 3.41 whereas the tunes CP1 and CP2 use the double Gaussian distribution in Eq. 3.42. The parameters which are not listed in Tab. 8.1 are set to the tune CUETP8M1 values. In addition, the PDF set which is employed in the simulations is set to the one used in the tune CUETP8M1 (NNPDF2.3 LO).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CUETP8M1</th>
<th>Monash</th>
<th>4C</th>
<th>A14</th>
<th>CP1</th>
<th>CP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultipartonInteractions:pT0Ref</td>
<td>2.40</td>
<td>2.28</td>
<td>2.40</td>
<td>2.09</td>
<td>2.40</td>
<td>2.30</td>
</tr>
<tr>
<td>MultipartonInteractions:ecmPow</td>
<td>0.252</td>
<td>0.215</td>
<td>0.252</td>
<td>0.215</td>
<td>0.154</td>
<td>0.139</td>
</tr>
<tr>
<td>MultipartonInteractions:expPow</td>
<td>1.60</td>
<td>1.85</td>
<td>1.60</td>
<td>1.85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MultipartonInteractions:coreRadius</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.544</td>
<td>0.376</td>
</tr>
<tr>
<td>MultipartonInteractions:coreFraction</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.68</td>
<td>0.33</td>
</tr>
<tr>
<td>ColourReconnection:range</td>
<td>1.80</td>
<td>1.80</td>
<td>1.71</td>
<td>1.50</td>
<td>2.63</td>
<td>2.32</td>
</tr>
</tbody>
</table>

Table 8.1: The MPI and color reconnection parameters which differ between the different tunes.

In Tab. 8.1 the parameter MultipartonInteractions:pT0Ref corresponds to $p_{\perp 0}^{\text{ref}}$ introduced in Eq. 3.40, while MultipartonInteractions:ecmPow corresponds to the exponent $\text{pow}$ in Eq. 3.40. These two parameters govern the energy dependence of the regularization, at low momentum transfer, of the partonic cross sections for the MPI. The parameter MultipartonInteractions:expPow stands for $a$ in the overlapping function in Eq. 3.41, whereas MultipartonInteractions:coreRadius and MultipartonInteractions:coreFraction characterize the inner core of the double Gaussian and the respective fraction of matter $\beta$ in Eq. 3.42. The additional parameter ColourReconnection:range in Tab. 8.1
Figure 8.4: Ratios of the a) 2-jet, and b) 3-jet distributions obtained evaluating $(\mu_{T}^{\text{ISR}}/p_T, \mu_{T}^{\text{FSR}}/p_T) = (1/2, 1/2), (1/2, 1), (1/2, 2), (1, 1/2), (1, 1), (1, 2), (2, 1/2), (2, 1), (2, 2)$, to the nominal ones obtained by using the central choice $(\mu_{T}^{\text{ISR}}/p_T, \mu_{T}^{\text{FSR}}/p_T) = (1, 1)$. The envelope (solid blue band) represents the PS uncertainty.
8.5. Non-perturbative sources

refers to the reconnection range parameter $R$ (see discussion on Eq. 3.44) which controls the probability of two MPI subsystems to be color reconnected.

The ratios of the inclusive 2- and 3-jet distributions obtained from the parametrization depicted in Tab. 8.1, to that from the tune CUETP8M1 are shown in Fig. 8.5 for two $p_T^{\text{max}}$ ranges. The different tunes show equivalent results. In Fig. 8.5 the predictions at parton level, without considering hadronization and MPI is also shown. The latter illustrates the size of the non-perturbative corrections when compared to the different tunes (not to be confused with non-perturbative uncertainty). For the inclusive 2-jet case the corrections from MPI and hadronization are below 1% whereas no difference is observed for the inclusive 3-jet distributions.

![Graphs showing ratios of inclusive 2- and 3-jet cross sections](image)

Figure 8.5: Ratios of the inclusive a) 2-jet and b) 3-jet cross sections differential in $\Delta \phi_{12}$ obtained with the tunes Monash, 4C, A14, CP1 and CP2, to the distributions obtained using the tune CUETP8M1. Only the parameters governing the MPI and color reconnection (CR) are considered.
8.6 Comparison between theoretical and experimental uncertainties

In Chapter 7 we discussed the different experimental uncertainty sources that affect the measured $\Delta \phi_{12}$ distribution: jet energy scale, jet energy resolution, $\Delta \phi_{12}$ resolution, model dependence as well as the statistical uncertainty of the measurements. For the case of the inclusive 2-jet distributions the resulting total experimental uncertainty ranges from $\sim 1\%$ at low $p_T^{\text{max}}$ to $\sim 3\%$ for high $p_T^{\text{max}}$ ranges. For the inclusive 3-jet distributions we observed a total experimental uncertainty that is generally larger than the one corresponding to the inclusive 2-jet measurements, ranging from $\sim 1.5\%$ to $\sim 4\%$.

Earlier in this Chapter, the different perturbative (Secs. 8.1, 8.2, 8.3, and 8.4) and non-perturbative (Sec. 8.5) theoretical uncertainty sources were investigated. We observed that for the case of the inclusive 2-jet distributions the main theoretical uncertainty source, especially at high $\Delta \phi_{12}$ values, is the PS uncertainty that accounts for sub-leading perturbation terms which are not exponentiated in the Sudakov form factor. The resulting total theoretical error, taken as the quadratic sum of the uncertainties from the different sources, is of the order of 1\% at low $\Delta \phi_{12}$, and reaches values of $\sim 5\%$ close to $180^\circ$ for the inclusive 2-jet distributions. For the case of the inclusive 3-jet distributions, the total theoretical uncertainty ranges from 1\% at low $p_T^{\text{max}}$ to 2\% at high $p_T^{\text{max}}$.

As an example, the ratios of the predictions from Powheg-2j + Pythia8 to data are shown in Fig. 8.6 for a) the inclusive 2-jet and b) inclusive 3-jet distributions as a function of $\Delta \phi_{12}$. The total experimental uncertainty is represented by the solid beige band, and the total theoretical uncertainty is indicated by the hatched band.

In Fig. 8.6a one can observe that for the inclusive 2-jet distributions the total theoretical uncertainty dominates at high $\Delta \phi_{12}$ for all the $p_T^{\text{max}}$ ranges. For the case of the inclusive 3-jet distributions depicted in Fig. 8.6b the total experimental error corresponds to the dominant uncertainty.
Figure 8.6: Ratios of POWHEG-2J + PYTHIA8 predictions to data for the inclusive a) 2-jet, and b) 3-jet distributions as a function of $\Delta \phi_{12}$, for nine $p_T^{\text{max}}$ ranges. The solid beige band indicates the total experimental uncertainty, the hatched band represents the total theoretical uncertainty. The figures are reported in Ref. [165].
CHAPTER 9

RESULTS AND DISCUSSION

Contents

9.1 Measurement of $\Delta \phi_{12}$ in inclusive 2-jet topologies ............ 130
9.2 Studies on the matching of matrix element to parton shower .... 135
  9.2.1 MC@NLO matching ................................................. 135
  9.2.2 POWHEG matching .................................................. 136
  9.2.3 Relation between the POWHEG and MC@NLO .............. 138
  9.2.4 LO multi-jet merging .............................................. 139
  9.2.5 Matrix element correction to parton shower ................. 141
  9.2.6 Application of TMDs to the calculation of the $\Delta \phi_{12}$ distribution . . . 141
9.3 Measurement of $\Delta \phi_{12}$ for nearly back-to-back jet topologies in inclusive 2- and 3-jet events ................................................. 144
9.4 Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies .............................. 151
  9.4.1 MC@NLO matching and impact on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies .................................................. 151
  9.4.2 POWHEG matching and impact on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies .................................................. 151
  9.4.3 Relation between POWHEG and MC@NLO, and the impact on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies .......... 154
  9.4.4 Impact of LO multi-jet merging on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies .................................................. 154
  9.4.5 Impact of the parton shower on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies .................................................. 157
  9.4.6 Application of TMDs to the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies .................................................. 159

129
Chapter 9. Results and discussion

9.5 Prospects and extension of the studies

In this section the measurements of the inclusive 2- and 3-jet normalized cross sections as a function of the azimuthal separation between the leading jets in the event ($\Delta \phi_{12}$) and the $p_T$ of the leading jet ($p_T^{\text{max}}$) are presented. The distributions are normalized to the 2-jet inclusive cross section integrated over $\Delta \phi_{12}$ in the corresponding $p_T^{\text{max}}$ range ($\sigma$):

$$
\frac{1}{\sigma} \frac{d\sigma}{d\Delta \phi_{12}}(\Delta \phi_{12}, p_T^{\text{max}}),
$$

(9.1)

The measurements are compared to predictions from leading order (LO) and next-to-LO (NLO) Monte Carlo (MC) event generators matched to parton showers (PS). Predictions considering both collinear parton distribution functions (PDFs) and transverse momentum dependent parton distributions (TMDs) are discussed. Therefore, the phenomenological studies presented in Chapter 4 are of big importance for the interpretation of the results.

First, the measurement of $\Delta \phi_{12}$ for inclusive 2-jet topologies in the range $90^\circ < \Delta \phi_{12} < 180^\circ$ is presented in Sec. 9.1. This $\Delta \phi_{12}$ range allows to study the tail of the $\Delta \phi_{12}$ distribution which is sensitive to hard radiation in the event (see Chapter 4). For these measurement a bin size of $5^\circ$ was chosen to reduce the impact of the statistical uncertainty in the region $\Delta \phi_{12} \sim 90^\circ$.

Subsequently in Sec. 9.3, the results from the measurement of $\Delta \phi_{12}$ in inclusive 2- and 3-jet topologies in the Sudakov region $170^\circ < \Delta \phi_{12} < 180^\circ$ are presented. This complementary measurement is focused on the back-to-back region, which is sensitive to soft radiation and resummation effects (see Chapter 4).

The related studies in Sec. 9.1 and Sec. 9.3 allow for a consistent investigation of two different regions of phase space: the region affected mostly by hard radiation, and the region where the resummation of soft partonic emissions is necessary.

9.1 Measurement of $\Delta \phi_{12}$ in inclusive 2-jet topologies

In this section, the measurement of the normalized inclusive 2-jet cross section differential in $\Delta \phi_{12}$, in the range $90^\circ < \Delta \phi_{12} < 180^\circ$ is discussed. The bin size corresponds to $5^\circ$, which is around ten times larger than the $\Delta \phi_{12}$ resolution (see Sec. 6.4). This section is based on the investigations I performed on $\Delta \phi_{12}$ azimuthal correlations that are reported in App. A and published in Ref. [6], as well as on the phenomenological studies presented in Chapter 4.

Figure 9.1 shows the normalized inclusive 2-jet cross section as a function of $\Delta \phi_{12}$ compared to the predictions from the Powheg-2j + Pythia8 event generator, for nine $p_T^{\text{max}}$ ranges. The distributions are scaled by multiplicative factors in order to have a better visualization of the results.

One can observe in Fig. 9.1 that the distributions grow steeply as $\Delta \phi_{12}$ approaches $\pi$. Additionally one can notice that as $p_T^{\text{max}}$ increases the distributions become steeper. In the following, the comparisons of the data to the predictions from several theoretical models are studied in detail.

In Fig. 9.2, the ratios of the predictions from the LO MC generators Pythia8, Herwig++, and MadGraph + Pythia8 to data are depicted for several ranges of $p_T^{\text{max}}$. The total experimental uncertainty is represented as a solid band and the statistical error of the predictions is represented by error bars.
9.1. Measurement of $\Delta \phi_{12}$ in inclusive 2-jet topologies

Figure 9.1: Normalized inclusive 2-jet cross section differential in $\Delta \phi_{12}$, for several $p_T^{\text{max}}$ ranges. The distributions are scaled by multiplicative factors for a better visualization. The data points include both statistical and systematic uncertainties, although the uncertainty bars are small and not visible. The data are compared with predictions from the Powheg-2J + Pythia8 event generator. The figure is reported in Ref. [6].

We observe that the predictions from the LO MC generator Pythia8 deviate from the data mainly at $\Delta \phi_{12} \sim 5\pi/6$ with differences ranging from $\sim 15\%$ at low $p_T^{\text{max}}$ to $\sim 30\%$ at high $p_T^{\text{max}}$. The predictions from the LO MC generator Herwig++ show larger discrepancies compared to data mostly at $\Delta \phi_{12} \sim \pi/2$, with differences from $\sim 30\%$ at low $p_T^{\text{max}}$ to $\sim 70\%$ at high $p_T^{\text{max}}$. The calculation provided by both Pythia8 and Herwig++ generators are based on $2 \rightarrow 2$ LO QCD matrix elements (ME). However the prediction of the $\Delta \phi_{12}$ distribution relies completely on the PS description provided by these generators since the fixed-order $2 \rightarrow 2$ LO ME results in the trivial prediction $\Delta \phi_{12} = \pi$ (see Chapter 4). As discussed in Sec. 3.2, the PS is based on the soft and collinear approximation of partonic emissions. Pythia8 uses a $p_T$ ordered PS whereas Herwig++ relies on an angular ordered PS. The discrepancies between the two generators are located mainly in the tail of the $\Delta \phi_{12}$ distribution, where hard radiation plays the main role. The differences observed in Fig. 9.2 between the Pythia8 and Herwig++ predictions are due to the way hard radiation is
Figure 9.2: Ratios of *Pythia8*, *Herwig++*, and *MadGraph + Pythia8* predictions to the normalized inclusive 2-jet cross section differential in $\Delta \phi_{12}$, for several $p_T^{\text{max}}$ ranges. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data. The figure is reported in Ref. [6].
9.1. Measurement of $\Delta \phi_{12}$ in inclusive 2-jet topologies

generated in the two different approaches. As mentioned in App. A, the PYTHIA8 and HERWIG++ parton shower calculations also use different $\alpha_S$ values for initial- and final-state emissions, in addition to a different upper scale for the PS simulation, which is higher in PYTHIA8 than in HERWIG++. In Sec. 9.3 we will observe that these generators give very similar results in the region $\Delta \phi_{12} \sim \pi$, where soft radiation contributions play the main role.

Predictions from MADGRAPH + PYTHIA8 are also shown in Fig. 9.2. A good agreement between the data and the MADGRAPH predictions is observed except for the range $200 < p_T^{\text{max}} < 300$ GeV where differences of $\sim 20\%$ are observed at low $\Delta \phi_{12}$. The MADGRAPH predictions are based on $2 \to 2$, $2 \to 3$, $2 \to 4$ LO ME calculations, and the different multiplicities are merged using the MLM scheme with a merging scale of 20 GeV, which is the default value used for the simulation of QCD event samples in CMS. As a consequence, Fig. 9.2 can be used to compare the effects of hard radiation generated by the PS approach to the hard radiation generated using higher-order LO ME.

However, although a good agreement between the data and the MADGRAPH predictions is observed for $p_T^{\text{max}} > 300$ GeV, a detailed investigation I carried out on LO multi-jet merging resulted in the fact that 20 GeV does not correspond to a proper merging scale choice at high $p_T^{\text{max}}$. The investigations are based on physical observables that are directly sensitive to the merging scale choice like the $p_T$ of the third hardest jet, or the differential jet rate which characterizes the scale at which $(n)$-jet configurations are observed as $(n+1)$-jet configurations. A discontinuity at the merging scale value in these distributions can be observed when the merging scale is not properly chosen. The details of these studies as well as the application to the inclusive jets $p_T$ distribution, which is a crucial distribution in jet physics, are presented App. E. In addition, I applied the LO multi-jet merging investigation results to another observable: the minimum azimuthal separation between any two of the four highest $p_T$ jets in inclusive 4-jet events ($\Delta \phi_{2j}^{\text{min}}$). The details of the $\Delta \phi_{2j}^{\text{min}}$ measurement are reported in App. A whereas the corresponding merging studies are presented in App. F. In Sec. 9.2.4 the comparison of MADGRAPH predictions for the $\Delta \phi_{12}$ distribution, using the default value for the merging scale (20 GeV), and the value resulting from the LO multi-jet merging investigation will be discussed.

The ratios of the NLO predictions from POWHEG-2J matched to PYTHIA8 and HERWIG++, POWHEG-3J + PYTHIA8, and HERWIG7 to the data are shown in Fig. 9.3 for the $\Delta \phi_{12}$ distributions in inclusive 2-jet events for several $p_T^{\text{max}}$ ranges. The total experimental uncertainty is represented as a solid band and the statistical error of the predictions is represented by error bars. The ME level prediction from POWHEG-2J (PH-2J-LHE) is also shown in Fig. 9.3.

The predictions from POWHEG matched to PS differ from the data mainly at low $\Delta \phi_{12}$, where the predictions from POWHEG-2J tend to overestimate the data whereas POWHEG-3J underestimates the data. For the case of the predictions from POWHEG-2J matched to PYTHIA8 the difference with respect to data decreases as $p_T^{\text{max}}$ increases, from $\sim 40\%$ at $p_T^{\text{max}} \sim 200$ GeV down to $\sim 20\%$ at $p_T^{\text{max}} \sim 1000$ GeV. The predictions from POWHEG-2J matched to HERWIG++ show larger deviations, with differences of $\sim 80\%$ at low $p_T^{\text{max}}$ down to $\sim 40\%$ at high $p_T^{\text{max}}$. The POWHEG-2J ME level, with up to three partons in the final state (see Chapter 4) falls steeply at the kinematical limit $\Delta \phi_{12} = 2\pi/3$. Since the ME are the same, the differences of POWHEG-2J + PYTHIA8 and POWHEG-2J + HERWIG++ compared to data have a similar trend to the one observed in Fig. 9.2 where the predictions from the HERWIG++ generator overestimated the data more than the ones from the PYTHIA8
Figure 9.3: Ratios of Powheg-2j + Pythia8, Powheg-2j-LHE, Powheg-2j + Herwig++, Powheg-3j + Pythia8, and Herwig7 predictions to the normalized inclusive 2-jet cross section differential in $\Delta \phi_{1,2}$, for several $p_T^{\text{max}}$ ranges. The solid band indicates the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data. The figure is reported in Ref. [6].
9.2 Studies on the matching of matrix element to parton shower

generator at low $\Delta \phi_{12}$. This supports the argument that the disagreement between the two generators observed at low $\Delta \phi_{12}$ are due to the difference in the generation of hard, well-separated partons by the Pythia8 and Herwig++ PS approaches.

The predictions from POWHEG-3J matched to Pythia8 show differences of 30-40\% at low $\Delta \phi_{12}$. POWHEG-3J provides ME with up to four partons in the final state meaning that the ME level is able to fill the region $\Delta \phi_{12} < 2\pi/3$ unlike the ME level from POWHEG-2J as will be observed in Sec. 9.2.2.

The predictions from POWHEG-2J and POWHEG-3J rely on the POWHEG method for the matching to the PS. The POWHEG method and its phenomenological implications for the $\Delta \phi_{12}$ distribution were discussed in Secs. 3.3.1 and 4.2. In Sec. 9.2.2 the impact on the $\Delta \phi_{12}$ distribution of the POWHEG parameters $\text{ptsqmin}$ and $\text{hdamp}$ introduced in Secs. 3.3.1 and 4.2 will be further discussed, and in addition the comparison to data will be shown.

The predictions from Herwig7 show a good agreement with the data for all the $p_T^{\text{max}}$ ranges. The MC@NLO method was used to match the $2 \rightarrow 2$ NLO ME to the PS for the Herwig7 predictions. We observe a significant difference between the Herwig7 and POWHEG-2J predictions, specifically for $\Delta \phi_{12} \gtrsim 2\pi/3$. One of the main reasons behind this difference is the POWHEG Sudakov that tames the soft and collinear divergence of the real $2 \rightarrow 3$ contributions for POWHEG that is not present for the case of the Herwig7 predictions. The latter are based on the MC@NLO subtraction (see Secs. 3.3.1 and 4.2.1). The other reason is the different showers used for Herwig7 and POWHEG-2J. In Sec. 9.2.2 this conclusion will become clearer by investigating the relation between $2 \rightarrow 2$ MC@NLO and POWHEG-2J matched to the same shower (Pythia8), with the help of the $\text{hdamp}$ parameter (introduced in Sec. 4.2.2) that controls the impact of the POWHEG Sudakov.

9.2 Studies on the matching of matrix element to parton shower

In this section the impact of the matching of ME at LO and NLO accuracy to PS on the calculation of the $\Delta \phi_{12}$ distribution is investigated, especially for the tail of the distribution. The results are compared to the data. The matching is also discussed for the case where TMDs are used for the evolution of the initial state partons and the generation of ISR. In addition, the contribution from higher-order corrections to PS is investigated.

9.2.1 MC@NLO matching

Figure 9.4 shows an example for one $p_T^{\text{max}}$ range, of the NLO predictions from $2 \rightarrow 2$ MC@NLO matched to PS, for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$. The MC@NLO matching is performed using subtraction terms from Pythia8, as well as from Herwig++. The prediction from $2 \rightarrow 2$ MC@NLO + Herwig++ provides a good description of the data for $\Delta \phi_{12} < 2\pi/3$ whereas for higher $\Delta \phi_{12}$ values it shows deviations of up to $\sim 15\%$. On the other hand, The prediction from $2 \rightarrow 2$ MC@NLO + Pythia8 does not provide a good description of the data, showing differences of up to $30\%$ at low $\Delta \phi_{12}$.

Although the ME level calculation from $2 \rightarrow 2$ MC@NLO + Herwig++ and $2 \rightarrow 2$ MC@NLO + Pythia8 agree for $\Delta \phi_{12} < 2.4$ rad (see Fig. 4.4), we observe in Fig. 9.4 that when the PS is included, they differ in $20 - 30\%$ for $\Delta \phi_{12} < 2.4$ rad. This difference is
attributed to the difference between the PS already observed in Fig. 9.2 between the predictions from the standalone Pythia8 and Herwig++ generators. It is interesting to notice that the inclusion of higher-order ME in Fig. 9.4 reduces in half the difference between the Pythia8 and Herwig++ calculation, compared to the difference between the corresponding standalone versions in Fig. 9.2.

9.2.2 POWHEG matching

Figure 9.5 shows the predictions for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$ from Powheg-2j matched to Pythia8, for one $p_T^{\text{max}}$ range as an example. The calculations depicted in Fig. 9.5 were obtained using different values of $\text{ptsqmin}$, including the default $\text{ptsqmin} = (0.8 \text{ GeV})^2$. The parameter $\text{ptsqmin}$ has been introduced in Sec. 3.3.1 and the phenomenological implications of its choice on the $\Delta \phi_{12}$ distribution have been discussed in Sec. 4.2.2.

In Fig. 9.5 one can observe that the tail of the $\Delta \phi_{12}$ distribution prediction is not affected by the different choices of $\text{ptsqmin}$. The similar description of the tail of the $\Delta \phi_{12}$ distribution is in accordance with the fact that $\text{ptsqmin}$ acts like a merging scale between the $2 \to 2$ and $2 \to 3$ partonic multiplicities (see Secs. 3.3.1 and 4.2.2). The tail of the distribution is filled by contributions originating from Powheg $2 \to 3$ partonic configurations. These contributions are the same for the $\text{ptsqmin}$ values considered in Fig. 9.5 (see Fig. 4.6). The
9.2. Studies on the matching of matrix element to parton shower

Figure 9.5: Predictions for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$ from POWHEG-2J + PYTHIA8 using various $\text{pt}^{\text{min}}$ values are compared to data for one $p_T^{\text{max}}$ range. The solid band in the ratio plot represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.

same conclusion does not apply to the back-to-back region where the parton multiplicities can contribute differently depending on the $\text{pt}^{\text{min}}$ choice as we will observe in Sec. 9.4.2.

Effect of the $\text{hdamp}$ parameter choice

Figure 9.6 shows the ratios of the predictions from POWHEG-2J + PYTHIA8 to the normalized inclusive 2-jet cross section differential in $\Delta \phi_{12}$. The predictions were obtained using different values of the $\text{hdamp}$ parameter introduced in Sec. 4.2.3, which controls the impact of the POWHEG Sudakov in the calculation of the exclusive NLO cross section according to the POWHEG method (see discussion in Sec. 3.3.1).

A nice feature shown in Fig. 9.6 is that at high $p_T^{\text{max}}$ the difference between the tail of the distribution using $\text{hdamp} = 200 \text{ GeV}$ agree with the one obtained using $\text{hdamp} = 400 \text{ GeV}$. This can be understood from the expressions in Eqs. 4.1 and 4.2, which imply that when the $p_T$ of the radiation recoil which causes the $\Delta \phi_{12}$ for the leading 2-jet system is much larger than $\text{hdamp}$, the action of the POWHEG Sudakov is suppressed and the tail of the $\Delta \phi_{12}$ distribution is then driven by the $2 \to 3$ real contribution (second term on the right-hand side of Eq. 4.1).

As an example, by using momentum conservation arguments we know that when $\Delta \phi_{12} \sim 2/3\pi$, the radiation against the leading system has $p_T \approx p_T^{\text{max}}$. Then, for $p_T \approx 1000 \text{ GeV}$, Eq. 4.2 results in $D = 0.04$ when $\text{hdamp} = 200 \text{ GeV}$, and $D = 0.14$ for $\text{hdamp} = 400 \text{ GeV}$. This means that, for both choices of $\text{hdamp}$, the effect from the term containing the
9.2.3 Relation between the POWHEG and MC@NLO

In Sec. 3.3.1 we discussed that the MC@NLO and POWHEG methods provide predictions which are NLO accurate for the integrated cross section, however, as discussed in Sec. 4.2, the exclusive, differential in $\Delta \phi_{12}$ distribution calculation depends on the way the soft and collinear singular region ($\Delta \phi_{12} \sim \pi$) is treated. For the case of POWHEG, the Sudakov factor tames the soft and collinear divergence of the real $2 \rightarrow 3$ contributions, however, as a side effect it also affects the tail of the $\Delta \phi_{12}$ distribution (see discussion in Sec. 4.2.2). The impact of the POWHEG Sudakov on the tail is controlled by the $h_{\text{damp}}$ parameter introduced in Sec. 4.2.2. We have checked that an $h_{\text{damp}}$ factor smaller than the scale of the process ($p_T^{\text{max}}$) results in the agreement of the ME predictions from pure $2 \rightarrow 2$ NLO, $2 \rightarrow 2$ MC@NLO, and POWHEG-2J for the tail of the $\Delta \phi_{12}$ distribution (see Fig. 4.9 and the related discussion).
9.2. Studies on the matching of matrix element to parton shower

in Sec. 4.2.3).

Figure 9.7 shows the comparison of the $\Delta \phi_{12}$ distribution predictions from $2 \rightarrow 2$ MC@NLO + PS and POWHEG-2J + PS to the data, where PYTHIA8 is used as the PS.

Figure 9.7: Predictions for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$ from POWHEG-2J + PYTHIA8 using $\text{hdamp} = 145$ GeV and from $2 \rightarrow 2$ MC@NLO + PYTHIA8 are compared to data for one $p_T^{\text{max}}$ range. The solid band in the ratio plot represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.

One can notice in Fig. 9.7 that POWHEG-2J + PYTHIA8 agrees with MC@NLO + PYTHIA8 at for $2.05 < \Delta \phi_{12} < 2.25$ rad. The observed agreement is not a coincidence, it supports the phenomenological investigation results shown in Fig. 4.9 at the ME level, where the fixed-order predictions from $2 \rightarrow 2$ MC@NLO, POWHEG-2J ($\text{hdamp} = 145$ GeV) and also $2 \rightarrow 2$ NLO agreed, in the same $\Delta \phi_{12}$ range.

The previous conclusion is also consistent with the detailed discussions on the POWHEG and MC@NLO matching phenomenology in Secs. 3.3.1, 4.2.1 and 4.2.2.

The differences observed at $\Delta \phi_{12} \sim \pi$ will be discussed in Sec. 9.4.3.

9.2.4 LO multi-jet merging

In Secs. 3.3.2 and 4.2.4 the phenomenology behind LO multi-jet merging was discussed. It was emphasized the importance of a proper merging scale choice for a correct merging of LO jet multiplicities.

As commented earlier during the discussion on the results shown in Fig. 9.2, a merging scale of 20 GeV used for the MADGRAPH predictions is not a consistent choice despite the agreement with the data. The latter conclusion is based on the investigations I carried out...
on LO multi-jet merging, which are reported in Apps. E and A. As mentioned earlier in this section, the studies were focused on the multi-jet merging calculation of the third hardest jet $p_T$ and the differential jet rate distributions, which are observables directly sensitive to the merging scale choice.

As a result, a suitable merging scale (MS) choice for the range $600 < p_T^{\text{max}} < 700$ GeV was found to be 140 GeV. Figure 9.8 shows the ratios of the predictions for the $\Delta \phi_{12}$ distribution from MadGraph + Pythia8 to data, for the range $600 < p_T^{\text{max}} < 700$ GeV. The MadGraph predictions are based on LO $2 \to 2$ and $2 \to 3$ QCD calculations. The samples are merged using the MLM scheme with an MS value of 20 GeV (default setting for QCD MadGraph samples generation in CMS), as well as a MS of 140 GeV using both the MLM and CKKW-L schemes for the merging procedure.

Figure 9.8: Ratios of the predictions from MadGraph + Pythia8 predictions (ME with up to three partons in the final state) to data are shown. The MLM merging scheme with a merging scale (MS) of 20 GeV, as well as both the MLM and CKKW-L merging schemes with an MS of 140 GeV are used for the predictions. The solid band in the ratio plot represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.

As we can observe in Fig. 9.8, there is a significant difference of $20 - 40\%$ between the predictions obtained with MS values of 20 GeV and 140 GeV. It is interesting to notice that both the MLM and CKKW-L calculations with MS of 140 GeV give similar results (less than 15%).

A very important feature of Fig. 9.8 is that the merging scale choice which was found to be consistent (see studies in App. E) is not the value which agrees better with the data and that also happens to be the default option for MC production of MadGraph QCD multi-jet samples in CMS, but it is rather the value that provides physical results for observables.
9.2. Studies on the matching of matrix element to parton shower

In Sec. 9.4.4, the complementary discussion of the impact of the merging scale choice on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies will be presented.

9.2.5 Matrix element correction to parton shower

In Fig. 9.9 the predictions the PS MC generators DIRE, VINCIA and PYTHIA8 are compared to data for one $p_T^{\text{max}}$ range. As mentioned in Secs. 3.3.3 and 3.4, the DIRE PS generator generator is used as a plugin to PYTHIA8 and it provides NLO corrections to the splitting kernels used in the shower. On the other hand, the VINCIA PS generator is also employed as a plugin to PYTHIA8 and, starting from $2 \to 2$ QCD ME, it is able to iteratively correct the up to the third emission of the shower with the corresponding real contribution.

![Figure 9.9: Ratios of the predictions from DIRE, VINCIA and PYTHIA8 to data are shown. The solid band in the ratio plot represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.](image)

One can observe in Fig. 9.9 that, although DIRE includes NLO correction to the shower splitting kernels, it does not provide a better description of the data than PYTHIA8. The differences are of up to 80% compared to the data for the tail of the $\Delta \phi_{12}$ distribution, similar to the differences observed for the case of POWHEG-2j + PS predictions in Fig. 9.3.

For the case of VINCIA we observe that it describes the data similar to PYTHIA8, with differences of $20 - 30\%$ around $\Delta \phi_{12} = 2.5$ rad.

9.2.6 Application of TMDs to the calculation of the $\Delta \phi_{12}$ distribution

In Sec. 4.3, the phenomenology of the use of TMDs for the evolution of the initial state and the generation of ISR was discussed. In this section we use the $\Delta \phi_{12}$ measurement to
Chapter 9. Results and discussion

compare the TMD and collinear PDF approaches.

Similar Sec. 4.3, in this section TMDs fitted to neutral-current and charged-current DIS data are used (see App. C for details). An angular ordering is used for the TMD evolution scale. In Fig. 9.10 the renormalization scale at each branching was chosen as the $p_T$ of the emission whereas in Fig. 9.10, the evolution scale $\mu$ and also the $p_T$ of the emission ($p_T = \mu(1 - z)$) are used as the renormalization scale at the branching vertex. The ISR is generated according to the TMD evolution. More details on the TMD determination, fitting and application can be found in App. C. As mentioned in Sec. 3.4, TMDs can be used in the CASCADE MC generator, which also provides the ISR according to the chosen TMD.

Figure 9.10 shows the predictions from POWHEG-2j + CASCADE and POWHEG-2j + PYTHIA8 compared to the data for two of the low $p_T^{\text{max}}$ ranges as well as two of the high $p_T^{\text{max}}$ ranges as examples. The POWHEG events were generated using the default option for $\text{ptsqmin}$ and $\text{hdamp}$.

![Figure 9.10: Ratio of POWHEG-2j + PYTHIA8 and POWHEG-2j + CASCADE predictions to data for the normalized inclusive 2-jet distributions differential in $\Delta\phi_{12}$, for two low $p_T^{\text{max}}$ ranges and two high $p_T^{\text{max}}$ ranges. The solid band represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.](image)

In Fig. 9.10 one can notice that POWHEG-2j + CASCADE describes the measurement much better than POWHEG-2j + PYTHIA8 at low as well as high $p_T^{\text{max}}$.

An important feature of Fig. 9.10 is that the use of the TMD provides a similar, good description of the $\Delta\phi_{12}$ distribution tail over a wide range of $p_T^{\text{max}}$. We observed in Fig. 9.6...
that a similar performance from \textsc{Powheg-2j} + \textsc{Pythia8} could only be achieved by varying
the arbitrary parameter \texttt{hdamp}.

The uncertainty from the TMD determination was determined in order to better assess
the performance of the TMD approach for the prediction of the $\Delta \phi_{12}$ distribution. There are
two main sources of uncertainty which are associated to the TMD determination. One source
is the experimental uncertainty of the data that is used to fit the TMD, the second source is
the model uncertainty of the parametrization of the initial conditions of the TMD. For the
case in which the $p_T$ of the splitting is used as renormalization scale at the branching vertex,
the arbitrary cut-off scale that is introduced to prevent the non-perturbative region in the
$\alpha_s$ evolution is considered as part of the model uncertainty. The details on the uncertainty
sources of the TMD determination can be found in App. C.

Figure 9.11 shows the predictions for $\Delta \phi_{12}$ in inclusive 2-jet events, compared with the
data for one $p_T^{\text{max}}$ range including the total uncertainty (model and experimental summed
in quadrature) from the TMD determination. The \textsc{Cascade} MC generator was used for
generating the initial state shower according to the TMD evolution whereas the \textsc{Pythia6}
was used for the final state shower. The prediction from \textsc{Powheg-2j} + \textsc{Pythia8} is also
shown in Fig. 9.11. Two TMD sets using the two different choices for the renormalization
scale of the splitting mentioned earlier are employed.

![Dijet azimuthal correlation ak4, 200 < $p_T^{\text{leading}}$ < 300 GeV](image)

Figure 9.11: The $\Delta \phi_{12}$ measurement in inclusive 2-jet topologies compared with the predictions
from TMDs where the angular ordering condition is used for the evolution scale ($\mu$). Two
choices for the renormalization scale at the branching vertex are used: $\mu$ and $p_T = \mu(1 - z)$.
The uncertainty bands correspond to the full (experimental and model) TMD uncertainties.

In Fig. 9.11 one can notice that the tail of the distribution is nicely described by the
\textsc{Powheg-2j} + \textsc{Cascade} predictions, within the total uncertainty, by the two TMD sets.
9.3 Measurement of $\Delta \phi_{12}$ for nearly back-to-back jet topologies in inclusive 2- and 3-jet events

In Sec. 9.1 we have on the measurement of the $\Delta \phi_{12}$ distribution in inclusive 2-jet events with a special interest in the tail of the distribution. We have discussed and interpreted the results in terms of different phenomenological approaches that combines the descriptions given by the ME and the PS. This section is focused on the very important and complementary region $\Delta \phi_{12} \sim 180^\circ$. As we have discussed in Sec. 4.1, this region, also called the Sudakov region, is specially sensitive to contributions from soft radiation.

Figure 9.12 depicts a high-$p_T$ 2-jet event recorded by the CMS experiment in 2016. The jets are central in rapidity ($|y| < 2.5$) and have a $p_T$ of 696 GeV and 694 GeV respectively. The azimuthal separation $\Delta \phi_{12}$ between the jets is around $178^\circ$. Event topologies like the one illustrated in Fig. 9.12 are the main subject of this section. The studies are based on the measurements I performed and are presented in App. B (reported in Ref. [165], submitted for publication to Eur. Phys. J. C), as well as on the phenomenological studies discussed in Chapter 4.

Figure 9.12: Event display showing a nearly back-to-back high $p_T$ 2-jet system. The event was recorded by the CMS experiment during the 2016 data taking period.

Figure 9.13 shows the measurement of the 2-jet inclusive cross section as a function of $\Delta \phi_{12}$ for six $p_T^{\text{max}}$ ranges. The distributions are normalized to the 2-jet inclusive cross section integrated over the full $\Delta \phi_{12}$ range $\sigma$, for the corresponding $p_T^{\text{max}}$. One can observe that as
9.3. Measurement of $\Delta \phi_{12}$ for nearly back-to-back jet topologies in inclusive 2- and 3-jet events

$p_T^{\text{max}}$ increases the importance of the Sudakov region also increases. The data points depicted in Fig. 9.13 are represented by markers and are compared to predictions from the PYTHIA8 (dotted lines) and HERWIG++ (solid lines) MC event generators.

Figure 9.13: Inclusive 2-jet distributions as a function of the azimuthal separation between the two leading jets $\Delta \phi_{12}$ for six $p_T^{\text{max}}$ regions. The data, represented by the markers, is compared to predictions from the HERWIG++ event generator (solid lines) and PYTHIA8 (dotted line). The total experimental uncertainty is depicted as error bars on the measurement. The figures are reported in Ref. [165].

We observe in Fig. 9.13 that the distributions are steeply falling towards low $\Delta \phi_{12}$, and that both the LO PS event generators PYTHIA and HERWIG++ undershoot the data close to 180°. In order to better visualize the differences between data and MC, Fig. 9.14 shows the ratios of the predictions from the LO generators PYTHIA, HERWIG++, and additionally MADGRAPH + PYTHIA8, for nine $p_T^{\text{max}}$ ranges. The statistical uncertainty of the predictions is shown as error bars whereas the total experimental uncertainty is represented by the solid band.

One can notice in Fig. 9.14 that PYTHIA8 describes well the data at low $\Delta \phi_{12}$, whereas at high $\Delta \phi_{12}$ the difference increases to around 10% at $\Delta \phi_{12} = 180^\circ$. One can observe that the PYTHIA8 description is similar for all the $p_T^{\text{max}}$ ranges. On the other hand, the predictions from HERWIG++ show deviations compared to the data of 5% for the low $p_T^{\text{max}}$ ranges, reaching 10% at high $p_T^{\text{max}}$.

In Fig. 9.14, the HERWIG++ and PYTHIA event generators provide a similar description of the data, generally agreeing with the measurement at low $\Delta \phi_{12}$ and undershooting the data at high $\Delta \phi_{12}$. This is an interesting feature since the generators use different approaches for the PS, an angular ordering for the emissions in the case of HERWIG++ and a $p_T$ ordering.
Figure 9.14: Ratios of Pythia8, Herwig++, and MadGraph + Pythia8 predictions to data, of the normalized inclusive 2-jet distributions as a function of the azimuthal angular separation of the two leading jets $\Delta\phi_{12}$, for all the $p_T^{\text{max}}$ regions. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data. The figure is reported in Ref. [165].

For Pythia8 shower. We observed in Fig. 9.2 that this is not the case for the tail of the $\Delta\phi_{12}$ distribution, where differences between the generators of up to 60% were observed. This is in accordance with the observation that small $\Delta\phi_{12}$ values generally require emissions far from the soft and collinear limit approximation of the shower.

The MadGraph + Pythia8 simulation, which is based on the merging of LO ME with up to four partons in the final state with a merging scale of 20 GeV in the MLM scheme, provides a good description of the data for $p_T^{\text{max}} < 600$ GeV, whereas at higher $p_T^{\text{max}}$ the difference increases to up to 10%.

Another interesting feature observed in Fig. 9.14 is the difference between the descriptions provided by Pythia8 and MadGraph + Pythia8, which reaches 20% at $\Delta\phi_{12} \sim 180^\circ$, for high $p_T^{\text{max}}$. This large difference (compared to the PS uncertainty shown in Fig. 8.4a) was a bit unexpected since the resummed contributions from the PS play the main role for $\Delta\phi_{12} \sim 180^\circ$ (see discussion in Chapter 4). As we will observe in Sec. 9.4.4 this difference is due to a non proper merging of the different partonic multiplicities. The studies supporting this conclusion, which are reported in App. E, are based on observables directly sensitive to
the merging scale choice. The impact and application to other crucial QCD observables like the inclusive jets $p_T$ distribution are presented in Apps. E and F.

The ratios of the NLO predictions from Powheg-2j + Pythia8, Powheg-2j + Herwig++, and Powheg-3j + Pythia8 to data are depicted in Fig. 9.15 for the inclusive 2-jet distributions, for nine $p_T^{\text{max}}$ ranges.

A good description of the data is provided by Powheg-3j + Pythia8, except for the last bin where differences of up to 10% are observed. The prediction from Powheg-2j + Herwig++ describes the data similar to Powheg-3j + Pythia8 except for the first $p_T^{\text{max}}$ region, as well as for the last $\Delta \phi_{12}$ bin for all $p_T^{\text{max}}$, where Powheg-2j + Herwig++ underestimates the measurement whereas Powheg-3j + Pythia8 overestimates it. The prediction from Powheg-2j + Pythia8 disagrees with the data with differences of up to 10% at low as well as high $\Delta \phi_{12}$. In Sec. 9.4.2 the Powheg calculation will be further discussed, based on the phenomenological studies presented in Sec. 4.2. Additionally, the NLO matched to PS prediction obtained using the MC@NLO method, as well as its relation with the

Figure 9.15: Ratios of Powheg-2j + Pythia8, Powheg-3j + Pythia8, and Powheg-2j + Herwig++ predictions to data, of the normalized inclusive 2-jet distributions as a function of $\Delta \phi_{12}$, for all the $p_T^{\text{max}}$ regions. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data. The Powheg-3j prediction is not shown for the highest bin in $p_T^{\text{max}}$ due to the large statistical fluctuations. The figure is reported in Ref. [165].
Chapter 9. Results and discussion

description provided by the POWHEG method will be discussed in Secs. 9.4.1 and 9.4.3. In Fig. 9.16, the inclusive 3-jet cross section as a function of $\Delta \phi_{12}$ is compared to the predictions from PYTHIA8 (dotted lines) and HERWIG++ (solid lines) for six $p_T^{\text{max}}$ ranges. The statistical uncertainty of the predictions is shown as error bars whereas the total experimental uncertainty is represented by the solid band. The data points are represented by markers whereas the experimental uncertainty is represented by the error bars on the predictions. The distributions are normalized to the inclusive 2-jet cross section integrated over the full $\Delta \phi_{12}$ range $\sigma$, for the corresponding $p_T^{\text{max}}$ range.

Figure 9.16: Inclusive 3-jet distributions as a function of the azimuthal separation between the two leading jets $\Delta \phi_{12}$ for six $p_T^{\text{max}}$ regions. The data, represented by the markers, is compared to predictions from the HERWIG++ event generator (solid lines) and PYTHIA8 (dotted line). The total experimental uncertainty is depicted as error bars on the measurement. The figures are reported in Ref. [165].

The inclusive 3-jet distributions shown in Fig. 9.16 are flatter than for the 2-jet inclusive case in depicted Fig. 9.13. This is understandable since contributions at $\Delta \phi_{12} \sim 180^\circ$ from exclusive 2-jet events are excluded by the requirement of at least a third jet with $p_T > 30$ GeV.

The ratios of the LO predictions from PYTHIA8, HERWIG++, and MADGRAPH + PYTHIA8 for the inclusive 3-jet distributions to data are shown in Fig. 9.17, for the nine $p_T^{\text{max}}$ ranges. As one can observe in Fig. 9.17 the MADGRAPH + PYTHIA8 generator prediction shows large deviations compared to the data, with differences from 10% at low $p_T^{\text{max}}$ to 15% at high $p_T^{\text{max}}$. The MADGRAPH predictions are based on the merging of ME with up to four partons in the final state, with a merging scale of 20 GeV in the MLM scheme. As commented in Sec. 9.1, this choice for the merging scale is not accurate (see discussion in App. E). In Sec. 9.4.4 the comparison to the data of MADGRAPH + PYTHIA8 predictions using a more
9.3. Measurement of $\Delta \phi_{12}$ for nearly back-to-back jet topologies in inclusive 2- and 3-jet events

Figure 9.17: Ratios of Pythia8, Herwig++, and MadGraph + Pythia8 predictions to data, of the normalized inclusive 3-jet distributions as a function of the azimuthal angular separation of the two leading jets $\Delta \phi_{12}$, for all the $p_T^{\text{max}}$ regions. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data. The figure is reported in Ref. [165].

Proper merging scale choice will be discussed.

One can also observe in Fig. 9.17 that the PS event generators Pythia8 and Herwig++ provide a very good description of the data.

Predictions of the inclusive 3-jet distributions as a function of $\Delta \phi_{12}$ from the NLO MC generators Powheg-2j + Pythia8, Powheg-3j + Pythia8, and Powheg-2j + Herwig++ are also compared to the data. The resulting ratios of these predictions to data are shown in Fig. 9.18.

One can notice in Fig. 9.18 that all the considered NLO matched to PS predictions tend to describe the data at low $\Delta \phi_{12}$, for $p_T^{\text{max}} > 400$ GeV. At high $\Delta \phi_{12}$ they all deviate from the data with differences of 5 – 15%.

Contrary to the trend observed in Fig. 9.15 for the 2-jet inclusive observable, the Powheg-2j + Herwig++ predictions tend to overestimate the data for $\Delta \phi_{12}$ around 180° whereas the Powheg-2j + Pythia8, Powheg-3j + Pythia8 calculations tend to underestimate the measurement in the same $\Delta \phi_{12}$ region.

One should notice that the predictions from Pythia8, Powheg-2j + Pythia8, Powheg-
Chapter 9. Results and discussion

Figure 9.18: Ratios of Powheg-2j + Pythia8, Powheg-3j + Pythia8, and Powheg-2j + Herwig++ predictions to data, of the normalized inclusive 3-jet distributions as a function of $\Delta \phi_{12}$, for all the $p_T^{\text{max}}$ ranges. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data. The Powheg-3j prediction is not shown for the highest bin in $p_T^{\text{max}}$ due to the large statistical fluctuations. The figure is reported in Ref. [165].

It is observed that the $\Delta \phi_{12}$ region close to 180° is not described by any of the MC predictions for the inclusive 2-jet and the inclusive 3-jet measurements simultaneously. This suggests that the observed deviations are related to the way soft partons are simulated within the PS.

3J + Pythia8, and MadGraph + Pythia8 shown in Figs. 9.14 and 9.15 for the inclusive 2-jet distributions, as well as in Figs. 9.15 and 9.18 for the inclusive 3-jet distributions, apparently provide different descriptions of the region $\Delta \phi_{12} > 175^\circ$, despite the fact that they use the same PS (Pythia8). In the following sections these features will be further discussed, based on the phenomenological studies reported in Chapter 4.
9.4 Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies

The impact of the matching of ME at LO and NLO accuracy to PS on the calculation of the $\Delta\phi_{12}$ distribution is investigated, and the results are compared to the data. The studies presented here are focused on the $\Delta\phi_{12}$ Sudakov region, and are therefore a very important, complementary part of the investigations shown in Sec. 9.2. The case where TMDs are used for the evolution of the initial state partons and the generation of ISR is also discussed. Additionally, the effect of higher-order corrections to PS is studied.

9.4.1 MC@NLO matching and impact on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies

The discussion in this sub-section is mainly based on the phenomenological studies presented in Sec. 4.2.1, and the results discussed in Sec. 9.2.1. The comparisons of the predictions from $2 \to 2$ NLO + PYTHIA8 and $2 \to 2$ NLO + HERWIG++ using the MC@NLO method to match the ME and the PS are shown in Fig. 9.19 for three $p_T^{\max}$ ranges. The inclusive 2-jet distributions are shown on the left-hand side of Fig. 9.19 whereas the distributions corresponding to the inclusive 3-jet observables are shown on the right-hand side of the same figure.

In Fig. 9.19 the predictions from $2 \to 2$ NLO + PYTHIA8 fail to describe the 2-jet inclusive distributions with differences of up to 15%. On the other hand, the same predictions describe the data for $\Delta\phi_{12} > 176^\circ$ ($\approx 3.08$ rad) in the case of the inclusive 3-jet observables.

The predictions from $2 \to 2$ MC@NLO + HERWIG++ in Fig. 9.19 describe well the data for $\Delta\phi_{12} < 176^\circ$ ($\approx 3.08$ rad) in the case of the inclusive 2-jet observables whereas for higher $\Delta\phi_{12}$ the differences reach values of around 10%. For the inclusive 3-jet distributions the predictions describe well the data. As one can observe, the use of different counter terms results in very different descriptions of the Sudakov region.

In Fig. 9.19, shown in the previous section, we observed that the prediction from $2 \to 2$ MC@NLO + HERWIG++ described the low $\Delta\phi_{12}$ region ($\Delta\phi_{12} < 2.2$ rad) better than $2 \to 2$ MC@NLO + PYTHIA8 shower.

It is interesting to notice the following feature of the results depicted in Figs. 9.4 and 9.19: the predictions from $2 \to 2$ NLO + HERWIG++ agree with the predictions from PYTHIA8 standalone for both the inclusive 2- and 3-jet distributions shown in Figs. 9.3, 9.14, and 9.17.

9.4.2 POWHEG matching and impact on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies

The discussion in this sub-section is mainly based on the results discussed in Sec. 9.2.2, and the phenomenological studies presented in Sec. 4.2.2. Figure 9.20 shows the predictions from POWHEG-2J + PYTHIA8 using $\text{ptsqmin} = (145 \text{ GeV})^2$, the default $\text{ptsqmin} = (0.8 \text{ GeV})^2$, and an intermediate reference value $\text{ptsqmin} = (50 \text{ GeV})^2$. The predictions are compared to the data.

One can observe in Fig. 9.20c that for the inclusive 3-jet distributions (plots on the right) no big differences are observed (compared to the experimental uncertainty) between the
Figure 9.19: Predictions from $2 \rightarrow 2$ NLO + PYTHIA8 and $2 \rightarrow 2$ NLO + HERWIG++ using the MC@NLO method are compared to data. The inclusive 2-jet (left) and 3-jet (right) distributions and the ratio to data are shown for three $p_T^{max}$ ranges. The solid band in the ratio plots represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.
9.4. Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies

Figure 9.20: Predictions from POWHEG-2J + PYTHIA8 using various ptsqmin values are compared to data. The inclusive 2-jet (left) and 3-jet (right) distributions, as well as the corresponding ratios to data are shown for three $p_T^{max}$ ranges. The solid band in the ratio plots represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.
predictions using the default $\text{pt}^{\text{sqmin}} = (0.8 \text{ GeV})^2$ and the predictions using the additional $\text{pt}^{\text{sqmin}}$ choices.

Another interesting feature to be noticed in Fig. 9.20c is that when $\text{pt}^{\text{sqmin}} = (145 \text{ GeV})^2$ the POWHEG + PYTHIA8 predictions for the inclusive 2-jet distributions (plots on the left) agree with the standalone PYTHIA8 predictions shown in Figs. 9.14 for both the inclusive 2-jet observables. This supports the phenomenological results on the matching of NLO ME to PS using the POWHEG method discussed in Sec. 4.2.2. For instance one can observe in Fig. 4.7 that the lowest multiplicity sample ($2 \rightarrow 2$) matched to the PYTHIA8 PS populates mainly the region $170^\circ < \Delta \phi_{12} < 180^\circ$.

In addition we can observe in Fig. 9.20c that the difference between the different $\text{pt}^{\text{sqmin}}$ values is localized in the region $\Delta \phi_{12} > 178^\circ$ for the 2-jet inclusive distributions. In Fig. 4.7a we observed that when $\text{pt}^{\text{sqmin}} = (0.8 \text{ GeV})^2$ the main contribution to this $\Delta \phi_{12}$ region comes from the POWHEG real emission ($2 \rightarrow 3$ configuration), whereas if $\text{pt}^{\text{sqmin}} = (145 \text{ GeV})^2$ the main contribution to the back-to-back region comes from PS in the $2 \rightarrow 2$ configuration (Fig. 4.7b). As a consequence, the 10% difference observed at $\sim 180^\circ$ between the predictions obtained using $\text{pt}^{\text{sqmin}} = (0.8 \text{ GeV})^2$ and $\text{pt}^{\text{sqmin}} = (145 \text{ GeV})^2$ is due to the difference in the Sudakov factor that is used for the first emission in each case: the PS Sudakov in the case of $\text{pt}^{\text{sqmin}} = (145 \text{ GeV})^2$, and the POWHEG Sudakov in the case of $\text{pt}^{\text{sqmin}} = (0.8 \text{ GeV})^2$ (see Eqs. 3.52, 3.55 and the related discussions in Secs. 3.3.1 and 4.2.2). The intermediate value $\text{pt}^{\text{sqmin}} = (50 \text{ GeV})^2$ provides a combination of both contributions. On the other hand we observed in Fig. 9.4 in the previous section, that for the case of the $\Delta \phi_{12}$ tail in inclusive 2-jet events, the predictions using the different $\text{pt}^{\text{sqmin}}$ values provided equivalent descriptions of the data. This is in accordance to the discussion in Sec. 4.2.2 and the result shown in Fig. 4.6 where we could observe that small values of $\text{pt}^{\text{sqmin}}$ (at least smaller than the scale of the process $p_{\text{T}}^{\text{max}}$) did not affect the $\Delta \phi_{12}$ distribution tail.

Another interesting point is that if one uses a $\text{pt}^{\text{sqmin}}$ value that leaves the back-to-back region to be filled by the PS, like $\text{pt}^{\text{sqmin}} = (145 \text{ GeV})^2$ in the case of Fig. 9.20, then a similar description of the data is achieved between the inclusive 2-jet and the inclusive 3-jet distributions.

### 9.4.3 Relation between POWHEG and MC@NLO, and the impact on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies

In Sec. 4.2.3 we observed how the $\text{hdamp}$ parameter that controls the impact of the Sudakov factor in POWHEG can be used to reproduce the pure NLO and MC@NLO ME description of the $\Delta \phi_{12}$ tail. The studies were used in Sec. 9.2.3, where the predictions for the inclusive 2-jet $\Delta \phi_{12}$ distribution from POWHEG-2j + PYTHIA8 using $\text{hdamp} = 145 \text{ GeV}$ and $2 \rightarrow 2$ MC@NLO + PYTHIA8 were compared to data. Similarly, Fig. 9.21 shows the comparison of the predictions from POWHEG-2j + PYTHIA8 using $\text{hdamp} = 145$ and $2 \rightarrow 2$ MC@NLO + PYTHIA8 to data, but in the range $170^\circ < \Delta \phi_{12} < 180^\circ$. The POWHEG predictions are calculated using the default value $\text{pt}^{\text{sqmin}} = (0.8 \text{ GeV})^2$. The predictions are shown for the inclusive 2-jet distributions (plots on the left) as well as for the inclusive 3-jet distributions (plots on the right), for three $p_{\text{T}}^{\text{max}}$ ranges.

Although $2 \rightarrow 2$ MC@NLO + PYTHIA8 and POWHEG-2j using $\text{hdamp} = 145$ give the same description of the low $\Delta \phi_{12}$ distribution tail at the ME level for inclusive 2-jet topologies (see Fig. 4.9), large discrepancies are observed in the Sudakov region as observed in Fig. 9.21.
9.4. Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies

Figure 9.21: Predictions from $2 \rightarrow 2$ MC@NLO + PYTHIA8 and POWHEG-2J + PYTHIA8 using $\text{hdamp} = 145$ GeV are compared to data. The inclusive 2-jet (left) and 3-jet (right) distributions, as well as the corresponding ratios to data are shown for three $p_T^{\text{max}}$ ranges. The solid band in the ratio plots represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.
Chapter 9. Results and discussion

The two generators differ more than 50% at $\Delta\phi_{12} = 180^\circ$. For the case of the inclusive 3-jet distributions as a function of $\Delta\phi_{12}$, a very good agreement with the data is observed for $\Delta\phi_{12} > 175^\circ$, whereas for $\Delta\phi_{12} > 175^\circ$ the discrepancies are of the order of 10%.

Impact of the $hdamp$ parameter choice on the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies

In Sec. 9.2.2 we investigated the impact of the $hdamp$ parameter on the low tail of the $\Delta\phi_{12}$ distribution in inclusive 2-jet events. In this subsection a similar study is presented for nearly back-to-back jets in inclusive 2-jet topologies.

In Fig. 9.22 the predictions for the inclusive 2-jet distributions as a function of $\Delta\phi_{12}$ from Powheg-2j + Pythia8 using $hdamp = 200$ GeV, $hdamp = 400$ GeV and $hdamp = \infty$ (default value) are compared to data.

![Figure 9.22: Ratios of Powheg-2j matched to Pythia8 predictions to the normalized inclusive 2-jet cross section differential in $\Delta\phi_{12}$, for two low $p_T^{\text{max}}$ ranges and two high $p_T^{\text{max}}$ ranges. The predictions were obtained using $hdamp = 200$ GeV, $hdamp = 400$ GeV and $hdamp = \infty$ (default value). The solid band represents the total experimental uncertainty and the error bars represent the statistical errors of the predictions.](image)

We can observe in Fig. 9.22 that the Sudakov region is very sensitive to the value of $hdamp$. In Sec. 9.2.2 we noticed that $hdamp = 200$ GeV, $hdamp = 400$ GeV provided equivalent descriptions for the low $\Delta\phi_{12}$ distribution tail at high $p_T^{\text{max}}$, however, Fig. 9.22 shows that
9.4. Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies

this is not the case for the Sudakov region where large differences are observed. One can trace this behaviour back to the discussion in Sec. 4.2.3, where we noticed the $h_{\text{damp}}$ parameter, which was introduced to recover the NLO real correction description of the tail (see Eq. 4.1 and related discussion), also modifies the POWHEG Sudakov exponent through the factor $D$, hence the resummation of soft emission contributions.

9.4.4 Impact of LO multi-jet merging on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies

In Sec. 9.2.4 the impact of the merging scale choice on the $\Delta \phi_{12}$ distribution tail was discussed. In this section the effect of the merging scheme and the corresponding merging scale ($\overline{\text{MS}}$) choice on the $\Delta \phi_{12}$ Sudakov region are presented.

For the studies, MadGraph + Pythia8 predictions based on LO $2 \rightarrow 2$ and $2 \rightarrow 3$ QCD calculations are employed. The different multiplicities are merged using the MLM merging scheme, as well as the alternative CKKW-L. An $\overline{\text{MS}}$ of 20 GeV corresponds to the default value for QCD MC samples production in CMS. As commented in Sec. 9.2.4, a $\overline{\text{MS}}$ of 20 GeV is not a consistent choice despite the agreement with the data in Fig. 9.2. This conclusion is based on the LO multi-jet merging studies reported in App. E. Instead, the choice of 140 GeV resulted in a better merging scale choice for $p_T^{\text{max}} > 500$ GeV.

In Fig. 9.23 the predictions from the merged $2 \rightarrow 2$ and $2 \rightarrow 3$ multiplicities using the MLM and CKKW-L schemes are shown, for the inclusive 2-jet (left) and 3-jet (right) observables.

Similar to Figs. 9.14 and 9.17 for high $p_T^{\text{max}}$, one can observe in Fig. 9.23 that the predictions using $\overline{\text{MS}} = 20$ GeV show a pronounced slope specially for the 3-jet observables, overestimating the measurement in $5 - 10\%$ for $\Delta \phi_{12} > 177^\circ$ (3.09 rad) while underestimating the data in the same amount for $\Delta \phi_{12} < 173^\circ$ (3.02 rad).

When a merging scale $\overline{\text{MS}} = 140$ GeV is chosen, the predictions from MLM and CKKW-L merging schemes differ from each other in less than 5% for the inclusive 2-jet as well as the inclusive 3-jet distributions.

It is very interesting to notice that, contrary to the behaviour observed in Figs. 9.14 and 9.17 for the case $\overline{\text{MS}} = 20$ GeV, the predictions obtained using the $\overline{\text{MS}} = 140$ GeV agree with the Pythia standalone results for both the inclusive 2- and 3-jet distributions shown in Figs. 9.14 and 9.17 respectively. This behaviour is reasonable since the PS is the same for both predictions, and it is the shower who gives the main contribution for $\Delta \phi_{12} \sim 180^\circ$ (see Fig. 4.11b).

9.4.5 Impact of the parton shower on the $\Delta \phi_{12}$ distribution for nearly back-to-back jet topologies

In Sec. 9.2.5 we studied the impact of shower corrections on the $\Delta \phi_{12}$ distribution tail, for inclusive 2-jet events. We compared the Dire, Vincia and Pythia8 showers to the data. As commented in Secs. 3.4 and 9.2.5, Dire provides NLO correction to the shower splitting kernels whereas Vincia is able to iteratively correct up to the third shower emission with the corresponding LO ME. We observed earlier in Fig. 9.9 that the Vincia shower provides a very similar description of the data to that from Pythia8 for the $\Delta \phi_{12}$ distribution tail, whereas a large discrepancy was observed between Dire and Pythia8.
Figure 9.23: Predictions from QCD LO $2 \rightarrow 2$ and $2 \rightarrow 3$ merged multiplicities from MADGRAPH using different choices of the merging scale (MS) are shown. Two merging schemes are used: MLM and CKKW-L. The inclusive 2-jet (left) and 3-jet (right) distributions and the ratio to data are shown for three $p_T^{\text{max}}$ ranges. The solid band in the ratio plots represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.
9.4. Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies

In this section we focus on the Sudakov region of the $\Delta\phi_{12}$ distribution for inclusive 2-jet as well as inclusive 3-jet topologies. In Fig. 9.24 the predictions from the Dire and VINCIA generators, used as plugins to PYTHIA8, are shown for the inclusive 2-jet (left) and 3-jet (right) distributions as a function of $\Delta\phi_{12}$, for three $p_T^{\text{max}}$ ranges. The predictions from the standalone PYTHIA8 generator are also shown.

As we can observe in Fig. 9.24, the VINCIA predictions do not show an improvement with respect to PYTHIA8 for the observables considered, and this agrees with the observations from the merging procedure in the last sections: higher-order LO matrix elements do not have a major impact on the Sudakov region.

Dire does not provide an improvement with respect to PYTHIA8. The Dire description compared to PYTHIA8 is different than the behaviour observed in Fig. 9.9, where discrepancies between the two generators of up to 80% were observed for the $\Delta\phi_{12}$ distribution tail. However it is expected that the PS uncertainties are reduced when using the Dire shower since higher-order corrections to the PS are included (see Ref. [166]).

9.4.6 Application of TMDs to the $\Delta\phi_{12}$ distribution for nearly back-to-back jet topologies

In the previous sections we have discussed several approaches that use collinear PDFs for the calculation of the $\Delta\phi_{12}$ distribution. We also discussed in Sec. 9.2.6 the application of TMD to the description of the $\Delta\phi_{12}$ distribution tail. In this section we focus on the $\Delta\phi_{12}$ distribution for nearly back-to-back topologies in inclusive 2- and 3-jet events.

We have observed that the main role in the Sudakov region is driven by the PS. The different theoretical approaches try to take into account as many features as possible like coherence of soft gluons, corrections from higher-order matrix elements, however, it has been shown earlier in this section that none of the models is able to reproduce the data for the 2- and 3-jet observables simultaneously. Perhaps the collinear PDF approach is too rigid and we miss information by not considering the transverse degrees of freedom. Another reason which indicates that this might be the case is the fact that the ISR shower plays the leading role in causing the $\Delta\phi_{12}$ separation as observed in Fig. 4.3, and, as discussed in Sec. 3.2.1 the ISR is driven by the evolution of the parton distribution which is being used.

In this section the CASCADE event generator is used to generate the ISR from ME previously calculated by the POWHEG-2j generator. The ISR is generated according to the TMD evolution discussed in App. C, and PYTHIA6 [167] for the FSR.

In Fig. 9.25 the comparison between POWHEG-2j + PYTHIA8 and POWHEG-2j + CASCADE to data is shown for the inclusive 2-jet distributions as a function of $\Delta\phi_{12}$, for four $p_T^{\text{max}}$ ranges.

It is interesting to notice that POWHEG-2j + CASCADE is able to describe the data with a similar accuracy than POWHEG-2j + PYTHIA8. The CASCADE prediction can be still improved, for example: 1) one could make use of the NLO splitting function machinery included in the latest versions of CASCADE for the ISR (Ref. [107]); 2) the matching of the ME and the ISR can be improved as will be discussed in App. G. At the moment of writing this thesis we are working on the aforementioned improvements.
Figure 9.24: Predictions from PYTHIA8, DIRE, and VINCIA showers are shown. The inclusive 2-jet (left) and 3-jet (right) distributions and the ratio to data are shown for three $p_T^{\text{max}}$ ranges. The solid band in the ratio plots represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.
9.4. Studies on the matching of matrix element to parton shower for nearly back-to-back jet topologies

Figure 9.25: Ratio of Powheg-2J + Pythia8 and Powheg-2J + Cascade predictions to data for the normalized inclusive 2-jet distributions differential in $\Delta\phi_{12}$, for two low $p_T^{\text{max}}$ ranges and two high $p_T^{\text{max}}$ ranges. The solid band represents the experimental uncertainty and the error bars represent the statistical uncertainty of the predictions.
9.5 Prospects and extension of the studies

The investigations presented so far correspond to a detailed study of the $\Delta \phi_{12}$ in QCD events. One can study this observable also in processes involving non-colored objects in the final state, for instance the azimuthal separation between the $Z$ boson and the highest $p_T$ jet ($\Delta \phi_{Zj}$) in inclusive $Z + 1$ jet events. The presence of the non-colored object in the final state implies a very different color flow for the interaction and therefore a different contribution from color correlated initial- and final-state soft gluons. A comparison of $\Delta \phi_{Zj}$ in inclusive $Z + 1$ jet events to the analogous case of $\Delta \phi_{12}$ for inclusive 2-jet event topologies can be an important step towards the understanding of color coherence in QCD.

Figure 9.26 shows the simulation obtained from the LO Pythia8 MC generator of the normalized $\Delta \phi_{12}$ distribution for inclusive 2-jet events, compared to the normalized $\Delta \phi_{Zj}$ distribution for inclusive $Z + 1$ jet events.

As one can observe in Fig. 9.26 the distributions differ in about 10% at low $\Delta \phi_{12}$ while at $\Delta \phi_{12} = 180^\circ$ the difference is of the order of 20%. One of the reasons that contributes to the observed difference is that the LO QCD calculation involves all the possible partonic flavor combinations in the initial state, whereas for the $Z + 1$ parton LO calculation only the combinations $qg \rightarrow Zg$ and $q\bar{q} \rightarrow Zq$ are present where $q(\bar{q})$ represents the quark(anti-quark) flavours and $g$ represents the gluon. This difference between the channels governing the processes is a motivation to investigate, in addition, the $\Delta \phi_{12}$ distribution as a function of the rapidity of the leading jets, as will be discussed in the following.
9.5. Prospects and extension of the studies

$\Delta \phi_{12}$ distribution in bins of rapidity, and the comparison to $\Delta \phi_{Zj}$ for inclusive $Z + 1$ jet events

Fig. 9.27 shows the different LO channels contributions obtained from PYTHIA8 to the $\Delta \phi_{12}$ distribution for inclusive 2-jet events (Fig. 9.27a), as well as for inclusive $Z + 1$ jet events (Fig. 9.27b), for $500 < p_T^{\text{max}} < 600$ GeV.

Figure 9.27: a) Different QCD channels contributing to the $\Delta \phi_{12}$ distribution for inclusive 2-jet events. b) Different channels contributing to the $\Delta \phi_{Zj}$ distribution for inclusive $Z + 1$ jet events. The PYTHIA8 MC generator was used for the simulation.

One can observe in Fig. 9.27 that at LO the main contribution to the inclusive $Z + 1$ jet distribution comes from the $qg$ channel ($80\%-90\%$), while for the inclusive 2-jet distribution the $qg$ channel is less dominant ($50\%$) although it is also the main channel. For the case of the inclusive $Z + 1$ jet distribution $p_T^{\text{max}}$ is defined as the object with the largest $p_T$.

As observed in Eq. 3.3 the contributions to the cross section from the different channels depend on the PDF of the corresponding incoming partons. As an example, Fig. 9.28a shows the PDFs from the NNPDF2.3 LO set [80] as a function of the longitudinal momentum fraction $x$ corresponding to different parton flavors, evaluated at the scale $Q^2 = (500 \text{ GeV})^2$. The TMDPLOTTER [168, 169] interface has been used to obtain Fig. 9.28a. One can see that the quark contribution is dominant at large $x$ whereas the gluon content is the main contribution at low $x$. As a consequence, one could enhance the $qg$ by looking at boosted 2-jet topologies, where one parton is characterized by a large $x$ and therefore it is likely to be a quark, while the other has a smaller $x$ and the probability that it corresponds to a gluon is higher. These jets configurations can be study using the variables $y^* = 1/2|y_1 - y_2|$, $y_b = 1/2|y_1 + y_2|$, where $y_1$, $y_2$ stand for the rapidity of the jets. An illustration of the jet topologies as a function of $y^*$ and $y_b$ is shown in Fig. 9.28b.
Chapter 9. Results and discussion

Figure 9.28: a) Parton distributions from NNPDF2.3 LO with $\alpha_s(M_Z) = 0.130$, evaluated at $Q^2 = (500 \text{ GeV})^2$. The ratio of the different PDFs to the gluon distribution is shown at the bottom. b) Illustration of 2-jet event topologies in the $y^*, y_b$ kinematic plane (taken from Ref. [170]).

The resulting contributions from the different QCD channels to the $\Delta \phi_{12}$ distribution for inclusive 2-jet events is shown in Fig. 9.29, for the boosted topology depicted in Fig. 9.28b ($y^* < 1, 2 < y_b < 3$).

One can observe that the contribution from the $qg$ channel has increased to $70\% - 80\%$ compared to the case inclusive in rapidity shown in Fig. 9.27a. One can now compare the normalized $\Delta \phi_{12}$ distribution in the boosted 2-jet regime to the normalized $\Delta \phi_Z$ distribution for inclusive $Z + 1$ jet events. This comparison is shown in Fig. 9.30.

As observed in Fig. 9.30 the difference between the distributions for $\Delta \phi_{12} = 180^\circ$ has decreased to $5\%$ compared to Fig. 9.26 where the predictions differed in about $20\%$.

The measurement of $\Delta \phi_{12}$ as a function of the rapidity of the leading jets can be an important observable to investigate in combination with the additional measurement of the azimuthal separation between the vector boson and the leading jet in inclusive $Z + 1$ jet topologies.
9.5. Prospects and extension of the studies

Figure 9.29: Different QCD channels contributing to the $\Delta \phi_{12}$ distribution for inclusive 2-jet events where $y^* < 1$ and $2 < y_b < 3$. The PYTHIA8 MC generator was used for the simulation.

Figure 9.30: Comparison of the normalized $\Delta \phi_{Zj}$ distribution for inclusive $Z + 1$ jet events, to the normalized $\Delta \phi_{12}$ observable for inclusive 2-jet events where $y^* < 1$ and $2 < y_b < 3$. The PYTHIA8 MC generator was used for the simulation.
Proton-proton (pp) collision events characterized by jets with high momentum transfer are a direct test for perturbative quantum chromodynamics (pQCD) at the Large Hadron Collider (LHC). Events of this kind are not only important as an expected signature for new physics but they can also improve our understanding of the non-Abelian nature of QCD. The latter is crucial for the development of the theoretical models in modern Monte Carlo (MC) event generators. The azimuthal separation between the two jets with the highest transverse momenta ($\Delta \phi_{12}$) in multi-jet event topologies is an important and unique observable, sensitive to the emission of hard as well as soft partonic radiation. In pQCD, small $\Delta \phi_{12}$ values are characterized by hard, wide-angle partonic emissions, whereas for nearly back-to-back topologies ($\Delta \phi_{12} \approx 180^\circ$) the observable probes the contribution from soft partonic emissions. In modern Monte Carlo (MC) event generators the description of hard radiation relies primarily on fixed-order matrix element (ME) calculations in pQCD, while the large logarithmic contributions from soft partonic radiation are resummed at all orders through the parton shower (PS). In this thesis, the measurement of $\Delta \phi_{12}$ in multijet topologies as well as phenomenological investigations using several theoretical models for generation of PS and its subsequent combination with ME were thoroughly discussed.

The measurements relied on data recorded by the Compact Muon Solenoid (CMS) experiment corresponding to proton-proton (pp) collisions at a center-of-mass energy of 13 TeV. The distributions are measured in bins of the transverse momentum of leading jet in the event ($p_T^{\text{max}}$). Additionally, the measurement has been complemented with a novel, detailed study of $\Delta \phi_{12}$ for nearly back-to-back jet topologies. The latter relied on the excellent performance of the CMS tracker which allowed to use a bin size of $1^\circ$ for the measurement as a compromise with the observed $\Delta \phi_{12}$ resolution ($\sim 0.5^\circ$). The measurements of the $\Delta \phi_{12}$ distribution have been compared to theoretical predictions based on leading-order (LO) and next-to-LO (NLO) matrix elements (ME) calculations match to PS. The LO predictions relied on predictions from Pythia8 and Herwig++ event generators, as well as on merged multi-jet calculations with up to four partons in the final state from MadGraph + Pythia8. The MC@NLO
and POWHEG method has been used to match the NLO ME calculation to the PS.

Differences between data and predictions for the $\Delta\phi_{12}$ distribution are observed to be of less than 80% over several orders of magnitude. The description of the $\Delta\phi_{12}$ distribution tail is generally improved by using higher-order ME. For the case of NLO predictions from POWHEG, the phenomenological studies have shown that a similar description to MC@NLO is obtained when the impact of the POWHEG Sudakov form factor is suppressed for the real ME contribution.

We have observed that the back-to-back region is described similarly by the LO MC generators Pythia8, Herwig++, as well as MadGraph + Pythia8 when the merging of the different multiplicities is properly performed. Specially at high $\Delta\phi_{12}$ and high $p_T^{\text{max}}$ the predictions from the LO generators agree within 5% with the predictions at NLO accuracy using the MC@NLO matching method despite the differences in the accuracy as well as in the PS. The predictions underestimate the data in up to 15% at $\Delta\phi_{12} = 180^\circ$ for the case of the inclusive 2-jet observables while they differ in less than 5% for the inclusive 3-jet distributions. For the case of the NLO predictions obtained using POWHEG one could observe that due to the exponentiation of the real corrections in the POWHEG Sudakov the back-to-back region is predicted differently than in the LO and MC@NLO cases. It was demonstrated than when the contribution from the PS is dominant by imposing a cut on the POWHEG emission the description of the back-to-back region is similar to the LO case.

As one could observed the contributions from the PS are important for the description of the $\Delta\phi_{12}$ distribution, primarily the contributions from initial-state shower (ISR). It was shown in the thesis that for nearly back-to-back event topologies the higher-order contributions to the PS from NLO splitting kernels as well as from higher-order LO ME correction do not affect considerably the predictions. However one could notice that the tail of the $\Delta\phi_{12}$ distribution can be affected in up to 60% by when the higher-order splitting kernels are used.

Two important conclusions can be drawn from the previous discussion on the comparison of the different theoretical predictions to data for nearly back-to-back jet topologies: 1) the inclusive 2- and 3-jet observables are not simultaneously described by any of the models; 2) several theoretical calculations which are formally different give a very similar description (within few percent) of the data.

The impact on the $\Delta\phi_{12}$ distribution of using transverse momentum dependent (TMD) parton distributions for the evolution of the initial state was investigated. As mentioned earlier, ISR is the main contribution from the PS to this observable. It was observed that the corrections introduced by the evolution of the initial state according to the TMD parton distribution and the corresponding generation of ISR significantly improves the description of the $\Delta\phi_{12}$ distribution tail compared to when only POWHEG ME are considered.

The thesis has been supplemented with several studies that allowed to successfully perform the measurements, as well as to interpret the results. A novel method to obtain response matrices was presented. This method can be extended and applied to other observables. The investigations on the proper merging scale choice in LO multi-jet merging were reported, as well as the studies on the matching of PS to ME including TMD parton distributions. The determination as well as the application of TMD parton distributions obtained using the parton branching method were discussed. Additional measurements corresponding to the minimum azimuthal separation between any two of the tree or four jets in inclusive 3- or 4-jet events respectively were reported.

An extension of the studies in the thesis would be to use a colorless particle instead of
one of the jets in the measurement of $\Delta \phi_{12}$. This process involves a very different color flow compared to the pure QCD case. A comparison between these measurements can be an additional important step towards the understanding of correlations between initial and final states caused by the emission of soft partonic radiation.
In this section, the measurements of azimuthal correlations in multi-jet events are discussed. I have significantly contributed to the measurements (specially resolution studies, unfolding, and systematic uncertainty sources), as well as to the theoretical calculations and interpretation of the results. The studies presented here are published in the European Physical Journal C (Eur. Phys. J. C) Ref. [6]. The measurements are performed using data from pp collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV, collected during 2016 with the CMS experiment, corresponding to an integrated luminosity of $35.9 \text{ fb}^{-1}$.

We investigate the azimuthal separation between the leading jets ($\Delta \phi_{12}$) in inclusive 2-, 3-, and 4-jet events. The measurement of $\Delta \phi_{12}$ is performed in the range $90^\circ < \Delta \phi_{12} < 180^\circ$, with a bin size of $5^\circ$. Additionally, we measure the minimum azimuthal separation ($\Delta \phi_{2j}^{\text{min}}$) between any two of the three (four) highest $p_T$ jets in inclusive 3-jet (4-jet) event topologies. The measurements are sensitive to contributions from soft as well as hard radiation in the event. The distributions are compared to theoretical predictions from LO and higher-order (NLO) calculations matched to PS.

The studies are the base for the novel investigations presented in this thesis, specially the ones reported in Apps. B and F.
Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at \( \sqrt{s} = 13 \) TeV

CMS Collaboration
CERN, 1211 Geneva 23, Switzerland

Received: 14 December 2017 / Accepted: 28 June 2018 / Published online: 10 July 2018
© CERN for the benefit of the CMS collaboration 2018

Abstract Azimuthal correlations between the two jets with the largest transverse momenta \( p_T \) in inclusive 2-, 3-, and 4-jet events are presented for several regions of the leading jet \( p_T \) up to 4 TeV. For 3- and 4-jet scenarios, measurements of the minimum azimuthal angles between any two of the three or four leading \( p_T \) jets are also presented. The analysis is based on data from proton–proton collisions collected by the CMS Collaboration at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 fb\(^{-1}\). Calculations based on leading-order matrix elements supplemented with parton showering and hadronization do not fully describe the data, so next-to-leading-order calculations matched with parton showering and hadronization models are needed to better describe the measured distributions. Furthermore, we show that azimuthal jet correlations are sensitive to details of the parton showering, hadronization, and multiparton interactions. A next-to-leading-order calculation matched with parton showers in the MC@NLO method, as implemented in HERWIG 7, gives a better overall description of the measurements than the POWHEG method.

1 Introduction

Particle jets with large transverse momenta \( p_T \) are abundantly produced in proton–proton collisions at the CERN LHC through the strong interactions of quantum chromodynamics (QCD) between the incoming partons. When the momentum transfer is large, the dynamics can be predicted using perturbative techniques (pQCD). The two final-state partons at leading order (LO) in pQCD are produced back-to-back in the transverse plane, and thus the azimuthal angular separation between the two highest-\( p_T \) jets, \( \Delta \phi_{1,2} = |\phi_{\text{jet1}} - \phi_{\text{jet2}}| \), equals \( \pi \). The production of additional high-\( p_T \) jets leads to a deviation of the azimuthal angle from \( \pi \). The measurement of azimuthal angular correlations (or decorrelation from \( \pi \)) in inclusive 2-jet topologies is a useful tool to test theoretical predictions of multijet production processes. Previous measurements of azimuthal correlation in inclusive 2-jet events were reported by the D0 Collaboration in \( pp \) collisions at \( \sqrt{s} = 1.96 \) TeV at the Fermilab Tevatron [1,2], and by the ATLAS Collaboration in pp collisions at \( \sqrt{s} = 7 \) TeV [3] and the CMS Collaboration in pp collisions at \( \sqrt{s} = 7 \) and 8 TeV [4,5] at the LHC. Multijet correlations have been measured by the ATLAS Collaboration at \( \sqrt{s} = 8 \) TeV [6,7].

This paper reports measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections as a function of the azimuthal angular separation between the two highest \( p_T \) (leading) jets, \( \Delta \phi_{1,2} \),

\[
\frac{1}{\sigma} \frac{d\sigma}{d\Delta \phi_{1,2}}
\]

for several regions of the leading jet \( p_T \), \( p_T^{\text{max}} \), for the rapidity region \( |y| < 2.5 \). The measurements cover the region \( \pi/2 < \Delta \phi_{1,2} \leq \pi \); the region \( \Delta \phi_{1,2} \leq \pi/2 \) includes large backgrounds due to \( t \bar{t} \) and \( Z/W+\)jet(s) events. Experimental and theoretical uncertainties are reduced by normalizing the \( \Delta \phi_{1,2} \) distribution to the total dijet cross section within each region of \( p_T^{\text{max}} \).

For 3- and 4-jet topologies, measurements of the normalized inclusive 3- and 4-jet cross sections are also presented as a function of the minimum azimuthal angular separation between any two of the three or four highest \( p_T \) jets, \( \Delta \phi_{2j}^{\text{min}} \),

\[
\frac{1}{\sigma} \frac{d\sigma}{d\Delta \phi_{2j}^{\text{min}}}
\]

for several regions of \( p_T^{\text{max}} \), for \( |y| < 2.5 \). This observable, which is infrared safe (independent of additional soft radiation), is especially suited for studying correlations amongst the jets in multijet events: the maximum value of \( \Delta \phi_{2j}^{\text{min}} \) is \( 2\pi/3 \) for 3-jet events (the “Mercedes star” configuration), while it is \( \pi/2 \) in the 4-jet case (corresponding to the “cross” configuration). The cross section for small angular separations is suppressed because of the finite jet sizes for a particular jet algorithm. The observable \( \Delta \phi_{2j}^{\text{min}} \) is sensitive to the...
contributions of jets with lower $p_T$ than the leading jet, i.e. the subleading jets, and one can distinguish nearby (nearly collinear) jets (at large $\Delta \phi_{ij}^{\text{min}}$) from other additional high $p_T$ jets (small $\Delta \phi_{ij}^{\text{min}}$), yielding information additional to that of the $\Delta \phi_{ij}^{\text{min}}$ observable. The 4-jet cross section differential in $\Delta \phi_{ij}^{\text{min}}$ has also been measured by the ATLAS Collaboration [7].

The measurements are performed using data collected during 2016 with the CMS experiment at the LHC, and the event sample corresponds to an integrated luminosity of 35.9 fb$^{-1}$ of proton-proton collisions at $\sqrt{s} = 13$ TeV.

2 The CMS detector

The central feature of the CMS detector is a superconducting solenoid, 13 m in length and 6 m in inner diameter, providing an axial 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. Charged-particle trajectories are measured by the tracker with full azimuthal coverage within $|\eta| < 3.0$. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors to the region $3.0 < |\eta| < 5.2$. Finally, muons are measured up to $|\eta| < 2.4$ by gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A 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. [8].

3 Theoretical predictions

Predictions from five different Monte Carlo (MC) event generators are compared with data. The PYTHIA 8 [9] and HERWIG++ [10] event generators are used, both based on LO $2 \to 2$ matrix element calculations. The PYTHIA 8 event generator simulates parton showers ordered in $p_T$ and uses the Lund string model [11] for hadronization, while HERWIG++ generates parton showers through angular-ordered emissions and uses a cluster fragmentation model [12] for hadronization. The contribution of multiparton interactions (MPI) is simulated in both PYTHIA 8 and HERWIG++, but the number of generated MPI varies between PYTHIA 8 and HERWIG++ MPI simulations. The MPI parameters of both generators are tuned to measurements in proton-proton collisions at the LHC and proton-antiproton collisions at the Tevatron [13], while the hadronization parameters are determined from fits to LEP data. For PYTHIA 8 the CUETP8M1 [13] tune, which is based on the NNPDF2.3LO PDF set [14, 15], is employed, while for HERWIG++ the CUETHppS1 tune [13], based on the CT14QED PDF set [16], is used.

The MADGRAPH [17, 18] event generator provides LO matrix element calculations with up to four outgoing partons, i.e. $2 \to 2, 2 \to 3$, and $2 \to 4$ diagrams. It is interfaced to PYTHIA 8 with tune CUETP8M1 for the implementation of parton showers, hadronization, and MPI. In order to match with PYTHIA 8 the $k_T$-MLM matching procedure [19] with a matching scale of 14 GeV is used to avoid any double counting of the parton configurations generated within the matrix element calculation and the ones simulated by the parton shower. The NNPDF2.3LO PDF set is used for the hard-process calculation.

Predictions based on next-to-leading-order (NLO) pQCD are obtained with the POWHEGBOX library [20–22] and the HERWIG 7 [23] event generator. The events simulated with POWHEG are matched to PYTHIA 8 or to HERWIG++ parton showers and MPI, while HERWIG 7 uses similar parton shower and MPI models as HERWIG++, and the MC@NLO [24, 25] method is applied to combine the parton shower with the NLO calculation. The POWHEG generator is used in the NLO dijet mode [26], referred to as PH-2J, as well as in the NLO three-jet mode [27], referred to as PH-3J, both using the NNPDF3.0NLO PDF set [28]. The POWHEG generator, referred to as PH-2J-LHE, is also used in the NLO dijet mode without parton showers and MPI. A minimum $p_T$ for real parton emission of 10 GeV is required for the PH-2J predictions, and similarly for the PH-3J predictions a minimum $p_T$ for the three final-state partons of 10 GeV is imposed. To simulate the contributions due to parton showers, hadronization, and MPIs, the PH-2J is matched to PYTHIA 8 with tune CUETP8M1 and HERWIG++ with tune CUETHppS1, while the PH-3J is matched only to PYTHIA 8 with tune CUETP8M1. The matching between the POWHEG matrix element calculations and the PYTHIA 8 underlying event (UE) simulation is performed using the shower-veto procedure, which rejects showers if their transverse momentum is greater than the minimal $p_T$ of all final-state partons simulated in the matrix element (parameter PTHARD = 2 [26]). Predictions from the HERWIG 7 event generator are based on the MMHT2014 PDF set [29] and the default tune H7-UE-MMHT [23] for the UE simulation. A summary of the details of the MC event generators used for comparisons with the experimental data is shown in Table 1.

Uncertainties in the theoretical predictions of the parton shower simulation are illustrated using the PYTHIA 8 event generator. Choices of scale for the parton shower are expected to have the largest impact on the azimuthal distributions. The parton shower uncertainty is calculated by independently varying the renormalization scales ($\mu_R$) for initial- and final-state radiation by a factor 2 in units of the $p_T$ of the emitted
Appendix A. Azimuthal correlations for inclusive 2-jet, 3-jet and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV

Table 1 Monte Carlo event generators used for comparison in this analysis. Version of the generators, PDF set, underlying event tune, and corresponding references are listed

<table>
<thead>
<tr>
<th>Matrix element generator</th>
<th>Simulated diagrams</th>
<th>PDF set</th>
<th>Tune</th>
</tr>
</thead>
<tbody>
<tr>
<td>MadGraph5_aMC@NLO 2.3.3 [17,18] + pythia 8.219 [9]</td>
<td>$2 \rightarrow 2$, $2 \rightarrow 3$, $2 \rightarrow 4$ (LO)</td>
<td>NNPDF2.3LO [14,15]</td>
<td>CUETP8M1 [13]</td>
</tr>
<tr>
<td>herwig 7.0.4 [23]</td>
<td>$2 \rightarrow 2$ (NLO), $2 \rightarrow 3$ (LO)</td>
<td>MMHT2014 [29]</td>
<td>H7-UE-MMHT [23]</td>
</tr>
</tbody>
</table>

Table 2 The integrated luminosity for each trigger sample considered in this analysis

<table>
<thead>
<tr>
<th>HLT $p_T$ threshold (GeV)</th>
<th>$p_{T}^{2\text{nd}}$ region (GeV)</th>
<th>$L$ (fb$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>200–300</td>
<td>0.024</td>
</tr>
<tr>
<td>200</td>
<td>300–400</td>
<td>0.11</td>
</tr>
<tr>
<td>320</td>
<td>400–500</td>
<td>1.77</td>
</tr>
<tr>
<td>400</td>
<td>500–600</td>
<td>5.2</td>
</tr>
<tr>
<td>450</td>
<td>$&gt;600$</td>
<td>36</td>
</tr>
</tbody>
</table>

Fig. 1 Normalized inclusive 2-jet cross section differential in $\Delta \phi_{1,2}$ for nine $p_{T}^{2\text{nd}}$ regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2j + PYTHIA 8 event generator

Fig. 2 Normalized inclusive 3-jet cross section differential in $\Delta \phi_{1,2}$ for eight $p_{T}^{2\text{nd}}$ regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH- 2j + PYTHIA 8 event generator
partons of the hard scattering. The maximum deviation found is considered a theoretical uncertainty in the event generator predictions.

4 Jet reconstruction and event selection

The measurements are based on data samples collected with single-jet high-level triggers (HLT) [30,31]. Five such triggers are considered that require at least one jet in an event with \( p_T > 140, 200, 320, 400, \) or 450 GeV in the full rapidity coverage of the CMS detector. All triggers are prescaled except the one with the highest threshold. Table 2 shows the integrated luminosity \( \mathcal{L} \) for the five trigger samples. The relative efficiency of each trigger is estimated using triggers with lower \( p_T \) thresholds. Using these five jet energy thresholds, a 100% trigger efficiency is achieved in the region of \( p_T^{\text{max}} > 200 \) GeV.

Particles are reconstructed and identified using a particle-flow (PF) algorithm [32], which uses an optimized combination of information from the various elements of the CMS detector. Jets are reconstructed by clustering the Lorentz vectors of the PF candidates with the infrared- and collinear-safe anti-\( k_T \) clustering algorithm [33] with a distance parameter \( R = 0.4 \). The clustering is performed with the \textsc{FastJet} package [34]. The technique of charged-hadron subtraction [35] is used to remove tracks identified as originating from additional pp interactions within the same or neighbouring bunch crossings (pileup). The average number of pileup interactions observed in the data is about 27.

The reconstructed jets require energy corrections to account for residual nonuniformities and nonlinearities in the detector response. These jet energy scale (JES) corrections [35] are derived using simulated events that are generated with \textsc{pythia} 8.219 [9] using tune CUETP8M1 [13] and processed through the CMS detector simulation based on \textsc{Geant4} [36]; they are confirmed with in situ measurements with dijet, multijet, photon+jet, and leptonic \( Z \)-jet events. An offset correction is required to account for the extra energy clustered into jets due to pileup. The JES corrections, which depend on the \( \eta \) and \( p_T \) of the jet, are applied as multiplicative factors to the jet four-momentum vectors. The typical overall correction is about 10% for central jets having \( p_T = 100 \) GeV and decreases with increasing \( p_T \).

Resolution studies on the measurements of \( \Delta \phi_{1,2} \) and \( \Delta \phi_{2j}^{\text{min}} \) are performed using \textsc{pythia} 8.219 with tune
Appendix A. Azimuthal correlations for inclusive 2-, 3-jet and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV

CUETP8M1 processed through the CMS detector simulation. The azimuthal angular separation is determined with an accuracy from $1^\circ$ to $0.5^\circ$ (0.017 to 0.0087 in radians) for $p_T^{\text{max}} = 200$ GeV to 1 TeV, respectively.

Events are required to have at least one primary vertex candidate [37] reconstructed offline from at least five charged-particle tracks and lies along the beam line within 24 cm of the nominal interaction point. The reconstructed vertex is the objects determined by a jet finding algorithm [33, 34] applied to all charged tracks associated with the vertex plus the corresponding associated missing transverse momentum. Additional selection criteria are applied to each event to remove spurious jet-like signatures originating from isolated noise patterns in certain HCAL regions. Stringent criteria [38] are applied to suppress these nonphysical signatures; each jet should contain at least two particles, one of which is a charged hadron, and the jet energy fraction carried by neutral hadrons and photons should be less than 90%. These criteria have a jet selection efficiency greater than 99% for genuine jets.

For the measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections as a function of $\Delta\phi_{1,2}$ or $\Delta\phi_{2j}^{\text{min}}$ all jets in the event with $p_T > 100$ GeV and a rapidity $|y| < 5$ are considered and ordered in $p_T$. Events are selected where the two highest-$p_T$ jets have $|y| < 2.5$, (i.e. events are not counted where one of the leading jets has $|y| > 2.5$). Also, events are only selected in which the highest-$p_T$ jet has $|y| < 2.5$ and exceeds 200 GeV. The inclusive 2-jet event sample includes events where the two leading jets lie within the tracker coverage of $|y| < 2.5$. Similarly the 3-jet (4-jet) event sample includes those events where the three (four) leading jets lie within $|y| < 2.5$, respectively. In this paper results are presented in bins of $p_T^{\text{max}}$, corresponding to the $p_T$ of the leading jet, which is always within $|y| < 2.5$.

5 Measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections in $\Delta\phi_{1,2}$ and $\Delta\phi_{2j}^{\text{min}}$

The normalized inclusive 2-, 3-, and 4-jet cross sections differential in $\Delta\phi_{1,2}$ and $\Delta\phi_{2j}^{\text{min}}$ are corrected for the finite detector resolution to better approximate the final-state particles, a procedure called “unfolding”. In this way, a direct comparison of this measurement to results from other experiments and to QCD predictions is possible. Particles are considered stable if their mean decay length is $c\tau > 1$ cm.
The bin width used in the measurements of $\Delta \phi_{1,2}$ and $\Delta \phi_{\text{min}}^2$ is set to $\pi/36 = 0.087$ rads ($5^\circ$), which is five to ten times larger than the azimuthal angular separation resolution. The corrections due to the unfolding are approximately a few per cent.

The unfolding procedure is based on the matrix inversion algorithm implemented in the software package RooUnfold \([39]\) using a 2-dimensional response matrix that correlates the modeled distribution with the reconstructed one. The response matrix is created by the convolution of the $\Delta \phi$ resolution with the generator-level inclusive 2-, 3-, and 4- cross section distributions from \textsc{pythia} 8 with tune CUETP8M1. The unfolded distributions differ from the distributions at detector level by 1–4%. As a cross-check, the above procedure was repeated by creating the response matrix with event samples obtained with the full \textsc{geant4} detector simulation, and no significant difference was observed.

We consider three main sources of systematic uncertainties that arise from the estimation of the JES calibration, the jet energy resolution (JER), and the unfolding correction. The JES uncertainty is estimated to be 1–2% for PF jets using charged-hadron subtraction \([35]\). The resulting uncertainties in the normalized 2-, 3-, and 4-jet cross sections differential in $\Delta \phi_{1,2}$ range from 3% at $\pi/2$ to 0.1% at $\pi$. For the normalized 3- and 4-jet cross sections differential in $\Delta \phi_{\text{min}}^2$ the resulting uncertainties range from 0.1 to 1%, and 0.1–2%, respectively.

The JER \([35]\) is responsible for migration of events among the $p_T^{\text{max}}$ regions, and its parametrization is determined from a full detector simulation using events generated by \textsc{pythia} 8 with tune CUETP8M1. The effect of the JER uncertainty is estimated by varying its parameters within their uncertainties \([35]\) and comparing the normalized inclusive 2-, 3-, and 4-jet cross sections before and after the changes. The JER-induced uncertainty ranges from 1% at $\pi/2$ to 0.1% at $\pi$ for the normalized 2-, 3-, and 4-jet cross sections differential in $\Delta \phi_{1,2}$ and is less than 0.5% for the normalized 3- and 4-jet cross sections differential in $\Delta \phi_{\text{min}}^2$.

The above systematic uncertainties in the JES calibration and the JER cover the effects from migrations due to the $p_T^{\text{max}}$ thresholds, i.e. migrations between the 2-, 3-, and 4-jet samples and migrations between the various $p_T^{\text{max}}$ regions of the measurements.

The unfolding procedure is affected by uncertainties in the parametrization of the $\Delta \phi$ resolution. Alternative response matrices, generated by varying the $\Delta \phi$ resolution by $\pm 10\%$, are used to unfold the measured spectra. This variation is

\[ \pm 10\% \]
motivated by studies on the $\Delta \phi$ resolution for simulated dijet events [32]. The uncertainty in the unfolding correction factors is estimated to be about 0.2%. An additional systematic uncertainty is obtained by examining the dependence of the response matrix on the choice of the MC generator. Alternative response matrices are constructed using the HERWIG++ event generator [10] with tune EE5C [40]; the effect is <0.1%. A total systematic unfolding uncertainty of 0.2% is considered, which accounts for all these various uncertainty sources.

6 Comparison with theoretical predictions

6.1 The $\Delta \phi_{1,2}$ measurements

The unfolded, normalized, inclusive 2-, 3-, and 4-jet cross sections differential in $\Delta \phi_{1,2}$ are shown in Figs. 1, 2, 3 for the various $p_{T}^{\text{max}}$ regions considered in this analysis. In the 2-jet case the $\Delta \phi_{1,2}$ distributions are strongly peaked at $\pi$ and become steeper with increasing $p_{T}^{\text{max}}$. In the 3-jet case, the $\Delta \phi_{1,2}$ distributions become flatter at $\pi$, since by definition dijet events do not contribute, and in the 4-jet case they become even flatter. The data points are overlaid with the predictions from the PH-2J + PYTHIA 8 event generator.

The ratios of the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 event generator predictions to the normalized inclusive 2-, 3-, and 4-jet cross sections differential in $\Delta \phi_{1,2}$ are shown in Figs. 4, 5, and 6, respectively, for all $p_{T}^{\text{max}}$ regions. The solid band around unity represents the total experimental uncertainty and the error bars on the points represent the statistical uncertainties in the simulated data. Among the LO dijet event generators, HERWIG++ exhibits the largest deviations from the experimental measurements, whereas PYTHIA 8 behaves much better than HERWIG++, although with deviations of up to 30–40%, in particular around $\Delta \phi_{1,2} = 5\pi/6$ in the 2-jet case and around $\Delta \phi_{1,2} < 2\pi/3$ in the 3- and 4-jet case. Predictions from HERWIG++ tend to overestimate the measurements as a function of $\Delta \phi_{1,2}$ in the 2-, 3-, and 4-jet cases, especially at $\Delta \phi_{1,2} < 5\pi/6$ for $p_{T}^{\text{max}} > 400$ GeV. However, it is remarkable that predictions based on the 2 $\rightarrow$ 2 matrix element calculations supplemented with parton showers, MPI, and hadronization describe the $\Delta \phi_{1,2}$ distributions rather well, even in regions that are sensitive to hard jets not included in the matrix element calculations. The MADGRAPH + PYTHIA 8 calculation using up to 4 partons in the
Fig. 11 Normalized inclusive 4-jet cross section differential in $\Delta\phi_{2j}$ for eight $p_T^{\text{max}}$ regions, scaled by multiplicative factors for presentation purposes. The size of the data symbol includes both statistical and systematic uncertainties. The data points are overlaid with the predictions from the PH-2J + PYTHIA 8 event generator matrix element calculations provides the best description of the measurements.

Figures 7, 8 and 9 show the ratios of the PH-2J matched to PYTHIA 8 and HERWIG++, PH-3J + PYTHIA 8, and HERWIG 7 event generators predictions to the normalized inclusive 2-, 3-, and 4-jet cross section differential in $\Delta\phi_{1,2}$, for all $p_T^{\text{max}}$ regions. The solid band around unity represents the total experimental uncertainty and the vertical bars on the points represent the statistical uncertainties in the simulated data. The predictions of PH-2J and PH-3J exhibit deviations from the measurement, increasing towards small $\Delta\phi_{1,2}$. While PH-2J is above the data, PH-3J predicts too few events at small $\Delta\phi_{1,2}$. These deviations were investigated in a dedicated study with parton showers and MPI switched off. Because of the kinematic restriction of a 3-parton state, PH-2J without parton showers cannot fill the region $\Delta\phi_{1,2} < 2\pi/3$, shown as PH-2J-LHE with the dashed line in Fig. 7, whereas for PH-3J the parton showers have little impact. Thus, the events at low $\Delta\phi_{1,2}$ observed for PH-2J originate from leading-log parton showers, and there are too many of these. In contrast, the PH-3J prediction, which provides $2 \rightarrow 3$ jet calculations at NLO QCD, is below the measurement. The NLO PH-2J calculation and the LO POWHEG three-jet calculation are equivalent when initial- and final-state radiation are not allowed to occur.

The predictions from PH-2J matched to PYTHIA 8 describe the normalized cross sections better than those where PH-2J is matched to HERWIG++. Since the hard process calculation is the same, the difference between the two predictions might be due to the treatment of parton showers in PYTHIA 8 and HERWIG++ and to the matching to the matrix element calculation. The PYTHIA 8 and HERWIG++ parton shower calculations use different $\alpha_s$ values for initial- and final-state emissions, in addition to a different upper scale for the parton shower simulation, which is higher in PYTHIA 8 than in HERWIG++. The dijet NLO calculation of HERWIG 7 provides the best description of the measurements, indicating that the MC@NLO method of combining parton showers with the NLO parton level calculations has advantages compared to the POWHEG method in this context.

For $\Delta\phi_{1,2}$ generator-level predictions in the 2-jet case, parton shower uncertainties have a very small impact (< 5%) at values close to $\pi$ and go up to 40–60% for increasing $p_T^{\text{max}}$ at $\Delta\phi_{1,2} \sim \pi/2$. For the 3- and 4-jet scenarios, parton shower uncertainties are less relevant, not exceeding ~20% for $\Delta\phi_{1,2}$.

6.2 The $\Delta\phi_{2j}^{\text{min}}$ measurements

The unfolded, normalized, inclusive 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{\text{min}}$ are shown in Figs. 10 and 11, respectively, for eight $p_T^{\text{max}}$ regions. The measured distributions decrease towards the kinematic limit of $\Delta\phi_{2j}^{\text{min}} \rightarrow 2\pi/3(\pi/2)$ for the 3-jet and 4-jet case, respectively. The data points are overlaid with the predictions from the PH-2J + PYTHIA 8 event generator. The size of the data symbol includes both statistical and systematic uncertainties.

Figures 12 and 13 show, respectively, the ratios of the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 event generators predictions to the normalized inclusive 3- and 4-jet cross sections differential in $\Delta\phi_{2j}^{\text{min}}$, for all $p_T^{\text{max}}$ regions. The PYTHIA 8 event generator shows larger deviations from the measured $\Delta\phi_{2j}^{\text{min}}$ distributions in comparison to HERWIG++, which provides a reasonable description of the measurement. The MADGRAPH generator matched to PYTHIA 8 provides a reasonable description of the measurements in the 3-jet case, but shows deviations in the 4-jet case.

The predictions from MADGRAPH + PYTHIA 8 and PYTHIA 8 are very similar for the normalized cross sections as a function of $\Delta\phi_{2j}^{\text{min}}$ in the four-jet case. It has been checked that predictions obtained with the MADGRAPH matrix element with up to 4 partons included in the calculation without contribution of the parton shower are able to reproduce the data very well. Parton shower effects increase the number of events with low values of $\Delta\phi_{2j}^{\text{min}}$.

$\text{Springer}$
Figures 14 and 15 illustrate the ratios of predictions from PH-2J matched to PYTHIA 8 and HERWIG++, PH-3J + PYTHIA 8, and HERWIG 7 to the normalized inclusive 3- and 4-jet cross sections differential in $\Delta\phi_{\min}^{2j}$, for all $p_T^{\text{max}}$ regions. Due to an unphysical behavior of the HERWIG 7 prediction (which has been confirmed by the HERWIG 7 authors), the first $\Delta\phi_{\min}^{2j}$ and last $\Delta\phi_{1,2}$ bins are not shown in Figs. 8, 9, 14 and 15. An additional uncertainty is introduced to the prediction of HERWIG 7, that is evaluated as the difference between this prediction and the prediction when the first bin is replaced with the result from HERWIG++. The additional uncertainty ranges from 2 to 10%. Among the three NLO dijet calculations PH-2J matched to PYTHIA 8 or to HERWIG++ provides the best description of the measurements.

For the two lowest $p_T^{\text{max}}$ regions in Figs. 13 and 15, which correspond to the 4-jet case, the measurements become statistically limited because the data used for these two regions were collected with highly prescaled triggers with $p_T$ thresholds of 140 and 200 GeV (cf. Table 2).

The PH-3J predictions suffer from low statistical accuracy, especially in the highest interval of $p_T^{\text{max}}$, because the same $p_T$ threshold is applied to all 3 jets resulting in low efficiency at large $p_T$. Nevertheless, the performance of the PH-3J simulation on multijet observables can already be inferred by the presented predictions, especially in the low $p_T$ region.

The effect of parton shower uncertainties in the event generator predictions of $\Delta\phi_{\min}^{2j}$ is estimated to be less than 10% over the entire phase space.

7 Summary

Measurements of the normalized inclusive 2-, 3-, and 4-jet cross sections differential in the azimuthal angular separation $\Delta\phi_{1,2}$ and of the normalized inclusive 3- and 4-jet cross sections differential in the minimum azimuthal angular separation between any two jets $\Delta\phi_{\min}^{2j}$ are presented for several regions of the leading-jet transverse momentum $p_T^{\text{max}}$. The measurements are performed using data collected during 2016 with the CMS detector at the CERN LHC corresponding to an integrated luminosity of 35.9 fb$^{-1}$ of proton–proton collisions at $\sqrt{s} = 13$ TeV.

The measured distributions in $\Delta\phi_{1,2}$ and $\Delta\phi_{\min}^{2j}$ are compared with predictions from PYTHIA 8, HERWIG++, MADGRAPH + PYTHIA 8, PH-2J matched to PYTHIA 8 and HERWIG++, PH-3J + PYTHIA 8, and HERWIG 7 event generators.
The leading order (LO) PYTHIA 8 dijet event generator exhibits small deviations from the $\Delta \phi_{1,2}$ measurements but shows significant deviations at low-$p_T$ in the $\Delta \phi_{2j}$ distributions. The HERWIG++ generator exhibits the largest deviations of any of the generators for the $\Delta \phi_{1,2}$ measurements, but provides a reasonable description of the $\Delta \phi_{2j}$ distributions. The tree-level multijet event generator MadGraph in combination with PYTHIA provides a good overall description of the measurements, except for the $\Delta \phi_{2j}$ distributions in the 4-jet case, where the generator deviates from the measurement mainly at high $p_T^{\text{max}}$.

The dijet next-to-leading order (NLO) PH-2j event generator deviates from the $\Delta \phi_{2j}$ measurements, but provides a good description of the $\Delta \phi_{2j}^{\text{min}}$ observable. The predictions from the three-jet NLO PH-3j event generator exhibit large deviations from the measurements and describe the considered multijet observables in a less accurate way than the predictions from PH-2j. Parton shower contributions are responsible for the different behaviour of the PH-2j and PH-3j predictions. Finally, predictions from the dijet NLO HERWIG 7 event generator matched to parton shower contributions with the MC@NLO method provide a very good description of the $\Delta \phi_{1,2}$ measurements, showing improvement in comparison to HERWIG++.

All these observations emphasize the need to improve predictions for multijet production. Similar observations, for the inclusive 2-jet cross sections differential in $\Delta \phi_{1,2}$, were reported previously by CMS [5] at a different centre-of-mass energy of 8 TeV. The extension of $\Delta \phi_{1,2}$ correlations, and the measurement of the $\Delta \phi_{2j}^{\text{min}}$ distributions in inclusive 3- and 4-jet topologies are novel measurements of the present analysis.

Acknowledgements We thank Simon Plätzer and Simone Alioli for discussion and great help on setting up, respectively, the HERWIG and the PH-3j simulation. We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centres and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWFW and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Fin-
Appendix A. Azimuthal correlations for inclusive 2-jet, 3-jet and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV


land); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NIH (Hungary); DAE and DST (India); IPM (Iran); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (Italy); INFN (I...


Appendix A. Azimuthal correlations for inclusive 2-jet, 3-jet and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV
The measurements I performed of the azimuthal separation between the leading jets $\Delta\phi_{12}$ are reported in this section for inclusive 2- and 3-jet event topologies. The investigations are focused on the region $170^\circ < \Delta\phi_{12} < 180^\circ$ with a fine bin size of $1^\circ$. The studies are also reported in Ref. [165] and have been submitted for publication to the European Physical Journal C (Eur. Phys. J. C).

The investigations are a novel extension of the measurements reported in App. A, and correspond to a detailed study of the Sudakov region which is sensitive to contributions from soft radiation. The study of azimuthal correlations in nearly back-to-back jet topologies allows for a more precise test of different resummation strategies, and it is a first step towards an improved understanding of the correlation effects of initial and final state soft radiation. Predictions from calculations at LO and NLO accuracy combined with PS are compared to the data. The measurements are performed using data from pp collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV, collected during 2016 with the CMS experiment, corresponding to an integrated luminosity of 35.9 fb$^{-1}$. 

APPENDIX

B

AZIMUTHAL SEPARATION IN NEARLY BACK-TO-BACK JET TOPOLOGIES IN INCLUSIVE 2- AND 3-JET EVENTS IN PP COLLISIONS AT $\sqrt{S} = 13$ TEV
Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration*

Abstract

A measurement for inclusive 2- and 3-jet events of the azimuthal correlation between the two jets with the largest transverse momenta, $\Delta \phi_{12}$, is presented. The measurement considers events where the two leading jets are nearly collinear (“back-to-back”) in the transverse plane and is performed for several ranges of the leading jet transverse momentum. Proton-proton collision data collected with the CMS experiment at a center-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 35.9 fb$^{-1}$ are used. Predictions based on calculations using matrix elements at leading-order and next-to-leading-order accuracy in perturbative quantum chromodynamics supplemented with leading-log parton showers and hadronization are generally in agreement with the measurements. Discrepancies between the measurement and theoretical predictions are as large as 15%, mainly in the region $177^\circ < \Delta \phi_{12} < 180^\circ$. The 2- and 3-jet measurements are not simultaneously described by any of models.

Submitted to the European Physical Journal C

© 2019 CERN for the benefit of the CMS Collaboration. CC-BY-4.0 license
1 Introduction

Collimated streams of particles (jets) can be produced in highly energetic parton-parton interactions in proton-proton (s) collisions, and their properties are described by the theory of strong interactions, quantum chromodynamics (QCD). In the lowest order perturbative QCD (pQCD), two jets with high transverse momenta \( p_T \) are produced "back-to-back" in the transverse plane. Higher order corrections lead to deviations from this configuration. Experimentally, this can be investigated by the measurement of the azimuthal separation, \( \Delta \phi_{12} = |\phi_{\text{jet}1} - \phi_{\text{jet}2}| \), between the two leading \( p_T \) jets in the transverse plane. Within the framework of pQCD, a final state with three or more partons is required for significant deviations from \( \Delta \phi_{12} = 180^\circ \). However, when deviations of \( \Delta \phi_{12} \) from 180° are small, a pQCD calculation at a fixed order in the strong coupling \( \alpha_S \) becomes unstable and a resummation of soft parton emissions to all orders in \( \alpha_S \) has to be performed. This resummation is approximated through the use of parton showers in Monte Carlo (MC) event generators.

Azimuthal correlations in inclusive 2-jet events have been measured previously by the D0 Collaboration in \( p\bar{p} \) collisions at a center-of-mass energy of \( \sqrt{s} = 1.96 \text{ TeV} \) [1, 2], in pp collisions by the ATLAS Collaboration at \( \sqrt{s} = 7 \text{ TeV} \) [3], and by the CMS Collaboration at \( \sqrt{s} = 7, 8, \) and 13 TeV [4–6], but none of the measurements considered in detail the region close to the back-to-back configuration. A detailed study of azimuthal correlations close to the back-to-back configuration allows a more precise test of different resummation strategies, and it is a first step towards an improved understanding of the effects of soft initial and final state gluons [7, 8].

In this article measurements are reported of the normalized inclusive 2-jet distribution as a function of the azimuthal separation \( \Delta \phi_{12} \) between the two leading \( p_T \) jets (jets 1 and 2),

\[
\frac{1}{\sigma_{p_{\text{max}}}} \frac{d\sigma}{d\Delta \phi_{12}},
\]

in several intervals of the leading jet \( p_T \) \( (p_{\text{max}}^T) \) within the rapidity range \( |y| < 2.5 \). The total dijet cross section \( \sigma_{p_{\text{max}}} \) is measured within each range of \( p_{\text{max}}^T \) integrated over the full range in \( \Delta \phi_{12} \). The binning of the measurement presented here is much finer than that of Ref. [6]. We consider \( \Delta \phi_{12} \) in the range \( 170^\circ < \Delta \phi_{12} \leq 180^\circ \).

The inclusive 3-jet distributions, differential in \( \Delta \phi_{12} \) and \( p_{\text{max}}^T \), with the \( p_T \) of third highest \( p_T \) jet typically being 1-2 orders of magnitude smaller than \( p_{\text{max}}^T \), are also suitable to test resummation effects arising from the presence of multiple scales in the interaction. Measurements of the inclusive 3-jet distribution normalized to \( \sigma_{p_{\text{max}}} \) are also presented, for several ranges of \( p_{\text{max}}^T \), and within \( |y| < 2.5 \).

The measurements are performed using data collected from pp collisions at \( \sqrt{s} = 13 \text{ TeV} \) during 2016 with the CMS experiment at the CERN LHC, corresponding to an integrated luminosity of 35.9 \( \text{fb}^{-1} \).

2 The CMS Detector

The central feature of the CMS detector is a superconducting solenoid, 13 m in length and 6 m in inner diameter, providing an axial 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. Charged-particle trajectories are measured by the tracker with full azimuthal
coverage within pseudorapidities $|\eta| < 2.5$. The ECAL, which is equipped with a preshower detector in the endcaps, and the HCAL cover the region $|\eta| < 3.0$. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors to the region $3.0 < |\eta| < 5.2$. Finally, muons are measured up to $|\eta| < 2.4$ by gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A 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. [9].

3 Theoretical predictions

Leading-order (LO) and next-to-LO (NLO) predictions are investigated. Among the LO event generators, both PYTHIA 8 [10] (version 8.219) and HERWIG++ [11] (version 2.7.1) are used for predictions because they feature different parton showering (PS) algorithms for soft and collinear parton radiation at leading-log accuracy. In PYTHIA 8 the PS emissions cover a region of phase space ordered in $x$ (fraction of the proton momentum carried by the parton) and the $p_T$ of the emitted parton, whereas in HERWIG++ the parton emissions are ordered in $x$ and the angle of the radiated parton (angular ordering). The Lund string model [12] is used for hadronization in PYTHIA 8 [10], whereas in HERWIG++ the cluster fragmentation model [13] is applied. Multiparton interactions (MPI) are simulated in PYTHIA 8 (tune CUETP8M1 [14] with the parton distribution function (PDF) set NNPDF2.3LO [15, 16]) and in HERWIG++ (tune CUETHppS1 [14] with the PDF set CTEQ6L1 [17]) with parameters tuned to measurements in pp collisions at the LHC and $p\bar{p}$ collisions at the Tevatron.

The MadGRAPH5_aMC@NLO [18] version 2.3.3 event generator (labelled as MadGraph in the following) interfaced with PYTHIA 8 with tune CUETP8M1 is also used in the analysis. Processes with up to 4 final-state partons at LO accuracy are calculated using the NNPDF2.3LO PDF set. The $k_T$-MLM matching procedure [19] is used with a matching scale of 10 GeV.

Among the NLO event generators, predictions obtained using the POWHEG BOX library [20–22] (version 2) with the PDF set NNPDF3.0NLO [23] are considered. The event generators PYTHIA 8 (tune CUETP8M1) and HERWIG++ (tune CUETHppS1) are used to simulate PS, hadronization, and MPI. The POWHEG generator in dijet mode [24], referred to as PH-2j, provides an NLO $2 \rightarrow 2$ calculation, and the POWHEG generator in three-jet mode [25] (using the MiNLO scheme [26, 27]), referred to as PH-3j, provides an NLO $2 \rightarrow 3$ calculation. For the PH-2j matrix elements (ME), a minimum $p_T$ of 100 GeV is required on the partons in the Born process, while for the PH-3j ME the minimum is lowered to 10 GeV to ensure coverage of the full phase space. These thresholds are applied to optimize the generation of events in the phase space of interest. The matching between the POWHEG matrix element calculations and the PYTHIA 8 underlying event (UE) [14] simulation is performed by using the shower-veto procedure (User-Hook option 2 [10]). The matching between the POWHEG matrix element calculations and the HERWIG++ UE [14] is performed by using a truncated shower [20].

Events generated by PYTHIA 8 (tune CUETP8M1), HERWIG++ (tune CUETHppS1), and MadGraph interfaced with PYTHIA 8 (tune CUETP8M1) are passed through a full detector simulation based on GEANT4 [28]. The simulated events are reconstructed with standard CMS programs.

Table 1 summarizes the theoretical predictions used in the present analysis.
Table 1: Monte Carlo event generators, parton densities, and underlying event tunes used for comparison with measurements.

<table>
<thead>
<tr>
<th>Matrix element generator</th>
<th>Simulated diagrams</th>
<th>PDF set</th>
<th>Tune</th>
</tr>
</thead>
<tbody>
<tr>
<td>MadGraph [18, 19] + PYTHIA 8.219 [10]</td>
<td>2→2, 2→3, 2→4 (LO)</td>
<td>NNPDF2.3LO [15, 16]</td>
<td>CUETP8M1 [14]</td>
</tr>
</tbody>
</table>

4 Jet reconstruction and event selection

The measurements are based on data samples collected with single-jet high-level triggers [29, 30]. The five single-jet triggers require at least one jet in the event with $p_T > 140, 200, 320, 400, \text{ or } 450$ GeV within the full rapidity coverage of the CMS calorimetry. Table 2 shows the various $p_{T_{\text{jet}}}^{\text{max}}$ regions accessed by the various triggers and the integrated luminosity for each trigger in the analysis. Each trigger is fully efficient for jets in the corresponding $p_T$ range in Table 2.

Table 2: The integrated luminosity for each trigger sample in the analysis, and trigger used for each $p_{T_{\text{jet}}}^{\text{max}}$ range.

<table>
<thead>
<tr>
<th>HLT $p_T$ threshold (GeV)</th>
<th>140</th>
<th>200</th>
<th>320</th>
<th>400</th>
<th>450</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{L}$ (fb$^{-1}$)</td>
<td>0.024</td>
<td>0.11</td>
<td>1.77</td>
<td>5.2</td>
<td>36</td>
</tr>
<tr>
<td>$p_{T_{\text{jet}}}^{\text{max}}$ region (GeV)</td>
<td>200–300</td>
<td>300–400</td>
<td>400–500</td>
<td>500–600</td>
<td>&gt;600</td>
</tr>
</tbody>
</table>

Particles are reconstructed and identified using a particle-flow (PF) algorithm [31], which utilizes an optimized combination of information from the various elements of the CMS detector. Jets are reconstructed by clustering the four-vectors of the PF candidates with the infrared- and collinear-safe anti-$k_T$ clustering algorithm [32] with a distance parameter $R = 0.4$. The clustering is performed with the FASTJET package [33]. To reduce the contribution to the reconstructed jets from additional pp interactions within the same bunch crossing (pileup), the charged-hadron subtraction technique [34] is used to remove tracks identified as originating from pileup vertices. The average number of pileup interactions per single bunch crossing observed in the data is about 27. The pileup contribution from neutral hadrons is corrected using a jet-area-based correction technique [35].

For this analysis, jets with rapidity $|y| < 5.0$ are reconstructed. For both the inclusive 2- and 3-jet samples, the events are selected by requiring the two highest $p_T$ jets to have $|y| < 2.5$ and $p_T > 100$ GeV. For the inclusive 3-jet events a third jet with $p_T > 30$ GeV and $|y| < 2.5$ is required. Contributions from pileup are negligible because the pileup removal algorithm has an efficiency of $\sim 99\%$ for jets with $30 < p_T < 50$ GeV and $|y| < 2.5$ [36].

5 Measurements of the normalized inclusive 2- and 3-jet distributions

The normalized inclusive 2- and 3-jet distributions as a function of $\Delta \phi_{12}$ are corrected for detector resolution. We achieve this by unfolding the observables to the level of stable final-state particles. In this way, a direct comparison of these measurements to results from other exper-
iments and to QCD predictions is possible. Particles are considered stable if their mean decay length is larger than 1 cm.

The unfolding procedure is based on the D’Agostini algorithm [37], which is implemented in the RooUNFOLD package [38], by using a response matrix that maps the generated jets onto the jets reconstructed by the CMS detector. The regularization (number of iterations) of the unfolding procedure is chosen by comparing the difference in $\chi^2$ between data and MC at detector level to that between data and MC at particle level. The consistency of the unfolding procedure is checked against the alternative TUNFOLD package [39, 40], which uses a least square minimization with Tikhonov regularization. Both methods provide equivalent results. The unfolding is performed in $\Delta \phi_{12}$. The response matrices are obtained using simulated events from the PYTHIA 8 event generator with the tune CUETP8M1. The difference between the unfolded distributions and the distributions at detector level range from $\sim$1% for the low $p_T$ regions up to $\sim$5% for the high $p_T$ regions.

The sources of systematic uncertainties arise primarily from the jet energy scale calibration (JES), the jet energy resolution (JER), the $\Delta \phi_{12}$ resolution, and the model dependence of the unfolding matrix. The effect of migrations between $p_T$ regions is very small because of the normalization of the cross sections in each $p_T$ range and therefore is neglected.

The $\Delta \phi_{12}$ resolution is $\sim$0.5°, as obtained from fully simulated event samples from PYTHIA 8 and MADGRAPH. A bin size of 1° is a compromise between the ability to study the back-to-back region and the impact of the unfolding correction of $\sim$2%. In Ref. [6] the study is focused on a different $\Delta \phi_{12}$ region, and a coarser bin size is chosen to account for the smaller size of the data sample.

Alternative response matrices are used to unfold the measured spectra by varying the $\Delta \phi_{12}$ resolution by $\pm$10%, an amount motivated by the observed difference between data and simulation in the $\Delta \phi_{12}$ bins.

An additional systematic uncertainty is caused by the dependence of the response matrix on the choice of the MC generator. Alternative response matrices are built using the HERWIG++ and MADGRAPH + PYTHIA 8 event generators. Because this analysis uses a finer binning compared with that of Ref. [6], the sensitivity to the uncertainty in the unfolding is increased. The observed effect from bin migration is less than 2%.

The JER and shifts in the JES can cause events to migrate between the $p_T$ regions. The JES uncertainties on the energy measurement are estimated to be 1–2% [34]. The resulting JES uncertainties in the normalized inclusive 2-jet distributions due to bin migrations are less than 2%, whereas for the normalized inclusive 3-jet distributions they are less than 3%. The effect of the JER uncertainties [34] is estimated by varying the JER parameters by one standard deviation up and down and comparing the results before and after the changes. The JER-induced uncertainties are less than 0.2% for the inclusive 2-jet $\Delta \phi_{12}$ measurement and below 0.4% for the normalized inclusive 3-jet measurement.

6 Comparison to theoretical predictions

In this section the measurements are compared with different theoretical predictions introduced in Section 3. In all figures displaying ratios, the solid band indicates the total experimental uncertainty and the error bars represent the statistical uncertainties from the simulation. In the figures displaying the normalized distributions, the statistical uncertainties of the measurements are often so small that the uncertainty bars are smaller than the symbol size.
The unfolded normalized inclusive 2-jet distribution as a function of the azimuthal separation of the two leading jets $\Delta\phi_{12}$ is shown in Fig. 1, and compared with the predictions from HERWIG++, (solid lines) and PH-2J + PYTHIA 8 (dotted lines) for different $p_T^{\text{max}}$ regions. The distributions are strongly peaked at 180° and become steeper with increasing $p_T^{\text{max}}$. The ratio of the PYTHIA 8 and HERWIG++ predictions to data are depicted in Fig. 2 for the inclusive 2-jet distributions in the nine $p_T^{\text{max}}$ ranges. Among the event generators, PYTHIA 8 and HERWIG++ show the largest deviations from the measurements for the $p_T^{\text{max}} < 800$ GeV regions in the inclusive 2-jet case, and the MADGRAPH + PYTHIA 8 event generator gives the best description in the same regions. The three generators show large deviations from the measurements in the $p_T^{\text{max}} > 800$ GeV regions.

The ratios of the NLO predictions to data for the unfolded normalized inclusive 2-jet distributions for the different $p_T^{\text{max}}$ regions are shown in Fig. 3. The NLO calculations considered are PH-2J + PYTHIA 8, PH-2J + HERWIG++, and PH-3J + PYTHIA 8. Among these NLO predictions PH-3J + PYTHIA 8 agrees better with the data. The PH-2J + HERWIG++ prediction is similar to the one of PH-3J + PYTHIA 8, except for the lowest $p_T^{\text{max}}$ region.

In Fig. 4 the unfolded normalized inclusive 3-jet distribution as a function of $\Delta\phi_{12}$ are compared with the predictions from HERWIG++ (solid lines) and PH-2J + PYTHIA 8 (dotted lines) for different $p_T^{\text{max}}$ regions. The ratios of the normalized inclusive 3-jet distributions for the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to data are shown in Fig. 5 for the different $p_T^{\text{max}}$ regions. In contrast to the 2-jet case, MADGRAPH + PYTHIA 8 shows the largest deviations from the measurements close to 180°, whereas PYTHIA 8 and HERWIG++ give a good description of the data.

The ratios of the NLO predictions from PH-2J + PYTHIA 8, PH-2J + HERWIG++, and PH-3J + PYTHIA 8 to data for the normalized inclusive 3-jet distributions are shown in Fig. 6. All the considered NLO+PS predictions fail to describe the measurements close to 180°. The predictions from PH-3J and MADGRAPH (Fig. 5) behave very differently, in contrast to their similar
Figure 2: Ratios of the normalized inclusive 2-jet distributions for the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to data as a function of the azimuthal separation of the two leading jets $\Delta \phi_{12}$, for all the $p_T^{\text{max}}$ regions. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data.

Since PYTHIA 8, PH-2J + PYTHIA 8, PH-3J + PYTHIA 8, and MADGRAPH + PYTHIA 8 use the same parton shower, the observed differences in the predictions can be attributed to the treatment of the additional partons present in the POWHEG and MADGRAPH ME. In general we observe that the $\Delta \phi_{12}$ region close to 180° is not well described by the predictions. The predictions agree better with the measurements for increasing $p_T^{\text{max}}$ and moving further away from the back-to-back region in $\Delta \phi_{12}$, where the contribution of resummation effects becomes smaller. The fact that none of the generators is able to describe the 2- and 3-jet measurements simultaneously suggests that the observed differences (of the order of 10%) are related to the way soft partons are simulated within the PS. The observed differences between $p_T$ and angular ordered PS for the LO generators PYTHIA 8 and HERWIG++ are small (Figs. 2 and 5) compared to the MADGRAPH predictions, which can be attributed to the presence of higher order ME.

The theoretical calculations have an intrinsic uncertainty arising from the freedom of choice of the renormalization and factorization scales ($\mu_R$ and $\mu_F$), the choice of the PDF and $\alpha_S(m_Z)$, and the modeling of nonperturbative effects and PS. The total theoretical uncertainty is the quadratic sum of the uncertainties from the scale, PDF, $\alpha_S$, and PS variations. Despite the better agreement of PH-3J, the PH-2J event generator is used instead for the estimation of the scale, PDF, and $\alpha_S$ uncertainties, because of the larger event sample. For the estimation of the PS uncertainty PYTHIA 8 is utilized. The following four sources of theoretical uncertainties are analyzed:

- The uncertainties due to the renormalization and factorization scales of the hard pro-
The nonperturbative contributions (MPI and hadronization) are included in the calculations above. The uncertainty from these contributions are estimated from the different choices of the UE tune and found to be negligible.

The uncertainty from PS dominates for the normalized inclusive 2-jet distributions. It is one order of magnitude larger than the rest of the sources near $\Delta\phi_{12} = 180^\circ$. On the other hand, for the normalized inclusive 3-jet distributions, the main contributions come from PS and PDF uncertainties.
Appendix B. Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV

Figure 4: Normalized inclusive 3-jet distributions as a function of the azimuthal separation of the two leading jets $\Delta\phi_{12}$ for different $p_T^{\text{max}}$ regions (left and right). The data are represented by the markers and the theory by histograms. Overlaid with the data are predictions from the HERWIG++ event generator (solid lines) and PH-2J + PYTHIA 8 (dotted lines). The total experimental uncertainty is depicted as error bars on the predictions.

Figure 5: Ratios of the normalized inclusive 3-jet distributions for the PYTHIA 8, HERWIG++, and MADGRAPH + PYTHIA 8 predictions to data as a function of the azimuthal separation of the two leading jets $\Delta\phi_{12}$, for all the $p_T^{\text{max}}$ regions. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data.

The predictions from PH-2J + PYTHIA 8 and PH-2J + HERWIG++ (Fig. 3) show the differences from using different PS models together with different matching procedures.
Figure 6: Ratios of the normalized inclusive 3-jet distributions for the PH-2J + PYTHIA8, PH-3J + PYTHIA8, and PH-2J + HERWIG++ predictions to data as a function of the azimuthal separation of the two leading jets $\Delta \phi_{12}$, for all $p_T^{\text{max}}$ regions. The solid band indicates the total experimental uncertainty and the error bars on the MC points represent the statistical uncertainty of the simulated data. The PH-3J prediction is not shown for the highest bin in $p_T^{\text{max}}$ because of the large statistical fluctuations.

Figs. 7(8) show the ratios of the PH-2J predictions to data for the normalized inclusive 2(3)-jet distributions for the different $p_T^{\text{max}}$ regions. The solid beige band indicates the total experimental uncertainty, and the hatched band represents the total theoretical uncertainty.

For the inclusive 2-jet distributions, the theoretical uncertainty is larger than the experimental one in the region close to $\Delta \phi_{12} = 180^\circ$ (Fig. 7). This is because the contribution from PS dominates in this region, and its uncertainty is large. For the inclusive 3-jet distributions (Fig. 8), the theoretical uncertainty is smaller in the region close to 180°. In this case, the region close to 180° is not filled by the partons from the PS, but by the third parton from PH-2J, leading to a smaller PS uncertainty.

7 Summary

Measurements of the normalized inclusive 2- and 3-jet distributions as a function of the azimuthal separation $\Delta \phi_{12}$ between the two jets with the highest transverse momentum $p_T$, in the collinear back-to-back region, are presented for several $p_T^{\text{max}}$ ranges of the leading jet. The measurements are performed using data collected with the CMS experiment at the LHC, corresponding to an integrated luminosity of 35.9 fb$^{-1}$ of pp collisions at a center-of-mass energy of 13 TeV.

The measured $\Delta \phi_{12}$ distributions generally agree with predictions from PYTHIA8, HERWIG++, MADGRAPH + PYTHIA8, PH-2J + HERWIG++, and POWHEG (PH-2) and PH-3J) matched to PYTHIA8. Discrepancies between the measurement and theoretical predictions are as large as 15%, mainly in the region $177^\circ < \Delta \phi_{12} < 180^\circ$. The predictions agree better with the measure-
Appendix B. Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV

Figure 7: Ratios of the normalized inclusive 2-jet distributions for the PH-2J + PYTHIA 8 predictions to data as a function of the azimuthal separation of the two leading jets $\Delta\phi_{12}$, for all $p_{T}^{\text{max}}$ regions. The solid beige band indicates the total experimental uncertainty and the hatched band represents the total theoretical uncertainty.

Figure 8: Ratios of the normalized inclusive 3-jet distributions for the PH-2J + PYTHIA 8 predictions to data as a function of the azimuthal separation of the two leading jets $\Delta\phi_{12}$, for all $p_{T}^{\text{max}}$ regions. The solid beige band indicates the total experimental uncertainty, the hatched band represents the total theoretical uncertainty.
ments for larger $p_T^{\text{max}}$ and smaller $\Delta \phi_{12}$, where the contribution of resummation effects becomes smaller. The 2- and 3-jet measurements are not simultaneously described by any of models.

The tree-level multijet event generator MADGRAPH in combination with PYTHIA 8 for showering, hadronization, and multiparton interactions, shows deviations from the measured $\Delta \phi_{12}$ for the inclusive 2-jet case, and even larger deviations for the 3-jet case. The PYTHIA 8 and HERWIG++ predictions show deviations (up to 10%) for the 2-jet inclusive distributions, whereas their predictions are in reasonable agreement with the inclusive 3-jet distributions.

The next-to-leading-order PH-2J + PYTHIA 8 prediction does not describe the data and a different trend compared to PYTHIA 8 and HERWIG++ towards $\Delta \phi_{12} = 180^\circ$ is observed. The PH-3J + PYTHIA 8 predictions agree with the measurements except for the last bin in the low $p_T^{\text{max}}$ intervals. The PH-2J + HERWIG++ prediction agrees well with the measurement in the highest $p_T^{\text{max}}$ ranges. For the inclusive 3-jet case, PH-2J + PYTHIA 8 performs similarly to PYTHIA 8 and HERWIG++ in the whole $\Delta \phi_{12}$ range for high $p_T^{\text{max}}$ intervals. MADGRAPH + PYTHIA 8, PH-3J + PYTHIA 8, and PH-2J + HERWIG++ show deviations from the measurements of up to 15%.

The measurement of correlations for collinear back-to-back dijet configurations probes the multiple scales involved in the event and, therefore, the differences observed between predictions and the measurements illustrate the importance of improving the models of soft parton radiation accompanying the hard process.

**Acknowledgments**

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMBWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); NKFIA (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); MES (Latvia); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MOS (Montenegro); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, RFBR, and NRC KI (Russia); MESTD (Serbia); CPAN, PCTI, and FEDER (Spain); MOSTR (Sri Lanka); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, contract No. 675440 (European Union); the Leventis Foundation; the A.P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the F.R.S.-FNRS and FWO (Belgium) under the “Excellence of Science – EOS” – be.h project n. 30820817; the Beijing Municipal Science & Technology Commission, No.
Z181100004218003; the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Lendület (“Momentum”) Program and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the New National Excellence Program ÚNKP, the NKFIÁ research grants 123842, 123959, 124845, 124850, and 125105 (Hungary); the Council of Science and Industrial Research, India; the HOMING PLUS program of the Foundation for Polish Science, cofinanced from European Union, Regional Development Fund, the Mobility Plus program of the Ministry of Science and Higher Education, the National Science Center (Poland), contracts Harmonia 2014/14/M/ST2/00428, Opus 2014/13/B/ST2/02543, 2014/15/B/ST2/03998, and 2015/19/B/ST2/02861, Sonata-bis 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia María de Maeztu, grant MDM-2015-0509 and the Programa Severo Ochoa del Principado de Asturias; the Thalis and Aristeia programmes cofinanced by EU-ESF and the Greek NSRF; the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University and the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand); the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

Appendix B. Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV
References


Appendix B. Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV


Appendix B. Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV
APPENDIX

C

COLLINEAR AND TMD PARTON DENSITIES FROM FITS TO PRECISION DIS MEASUREMENTS IN THE PARTON BRANCHING METHOD

In this section a novel investigation is presented, in which the Parton Branching (PB) method has been employed to obtain transverse momentum dependent (TMD) parton densities at NLO accuracy. The studies are reported in Ref. [52] and have been accepted for publication in Physical Review D (Phys. Rev. D).

The integrated TMD parton densities are fitted to precision data from deep inelastic scattering cross section measurements at HERA, over a large range in longitudinal momentum fraction $x$ and scale $Q^2$. Two TMD sets differing in the renormalization scale at which the strong coupling $\alpha_s$ is evaluated were obtained and applied to the prediction of the Drell-Yan transverse momentum spectrum. The application to the $\Delta\phi_{12}$ distribution was discussed in Secs. 4.3, 9.2.6 and 9.4.6, and it is also reported in Ref. [171].
Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

A. Bermudez Martinez\textsuperscript{1}, P. Connor\textsuperscript{1}, F. Hautmann\textsuperscript{2,3,4}, H. Jung\textsuperscript{1}, A. Lelek\textsuperscript{1}, V. Radescu\textsuperscript{3,5,6}, R. Žlebčík\textsuperscript{1}

\textsuperscript{1}DESY, Hamburg, FRG
\textsuperscript{2}RAL, Chilton OX11 0QX and University of Oxford, OX1 3NP
\textsuperscript{3}Elementary Particle Physics, University of Antwerp, B 2020 Antwerp
\textsuperscript{4}UPV/EHU, University of the Basque Country, E 48080 Bilbao
\textsuperscript{5}CERN, CH-1211 Geneva 23

Abstract

Collinear and transverse momentum dependent (TMD) parton densities are obtained from fits to precision measurements of deep inelastic scattering (DIS) cross sections at HERA. The parton densities are evolved by DGLAP evolution with next-to-leading-order (NLO) splitting functions using the parton branching method, allowing one to determine simultaneously collinear and TMD densities for all flavors over a wide range in $x$, $\mu^2$ and $k_t$, relevant for predictions at the LHC. The DIS cross section is computed from the parton densities using perturbative NLO coefficient functions.

Parton densities satisfying angular ordering conditions are presented. Two sets of parton densities are obtained, differing in the renormalization scale choice for the argument in the strong coupling $\alpha_s$. This is taken to be either the evolution scale $\mu$ or the transverse momentum $q_t$. While both choices yield similarly good $\chi^2$ values for the fit to DIS measurements, especially the gluon density turns out to differ between the two sets.

The TMD densities are used to predict the transverse momentum spectrum of $Z$-bosons at the LHC.

1 Introduction

Parton density functions (PDFs) play an essential role for precise predictions of production processes in hadronic collisions, obtained from the factorization of the cross sections in hard-scattering process and PDFs, containing a non-perturbative input with perturbatively calculable evolution. The most advanced determination of parton densities come from the application of DGLAP \cite{1-4} evolution with next-to-leading-order (NLO) \cite{5,6} and next-to-next-to-leading order (NNLO) \cite{7,8} splitting functions. The collinear parton densities as a function of the longitudinal momentum fraction $x$ and the evolution scale $\mu^2$ are obtained by several groups, for example ABM \cite{9}, CTEQ \cite{10}, HERAPDF \cite{11}, NNPDF \cite{12} and MSTW \cite{13,14}. The different groups use the same DGLAP evolution, with ordering in virtuality and the

\textsuperscript{1}Now at IBM Germany
same choice of the renormalization scale, but they differ in, for example, the treatment of heavy flavors, and the experimental data sets which are used for the determination of the starting distributions.

In Refs. [15, 16] a new method, the Parton Branching method (PB), was introduced to treat DGLAP evolution. The method applies at exclusive level, and provides an iterative solution of the evolution equations. It agrees with the usual methods to solve the DGLAP equations for inclusive distributions, but it provides also additional features: in addition to the standard ordering in virtuality, angular ordering can be applied with the necessary change in the argument of $\alpha_s$ [17, 18]. The transverse momentum at every branching vertex can be calculated, leading to a natural determination of the transverse momentum dependent (TMD) parton densities. The PB method uses the unitarity formulation of QCD evolution equations [19] and is close in spirit to the works in [20–25]. As shown in Refs. [16, 26], it can be applied to NLO and NNLO splitting functions.

In this article we present a determination of collinear and TMD parton densities at NLO applying the PB method for the parton evolution. The initial parton distributions are determined from a fit to HERA I+II inclusive DIS cross section measurements [11]. An early fit was presented in Ref. [16]. Here, we present results obtained with angular ordering, both for collinear (integrated, iTMD) and TMD parton densities, and for different choices of the renormalization scale in $\alpha_s$ including a full treatment of experimental and model dependent uncertainties. We show an application of these TMDs to the calculation of the transverse momentum of the $Z$-boson in Drell-Yan (DY) production at the Large Hadron Collider (LHC).

2 Parton Branching method and evolution equation

The PB method has been described in detail in Refs. [15, 16]. Here we limit ourselves to recalling its main elements.

2.1 General features

The method is based on introducing a soft-gluon resolution scale $z_M$ into the QCD evolution equations to separate resolvable and non-resolvable emissions, and treating these via, respectively, the resolvable splitting probabilities $P_{ba}^{(R)}(\alpha_s, z)$ and the Sudakov form factors

$$\Delta_a(z_M, \mu^2, \mu_0^2) = \exp \left( -\sum_b \int_{\mu_0^2}^{\mu^2} \frac{d\mu}{\mu^2} \int_0^{z_M} dz \int_{\mu_0^2}^{z_M} d\mu' \frac{d\mu'}{\mu'^2} \int_{0}^{\frac{z_M}{\mu_0^2}} dz' \frac{d\mu'}{\mu'^2} P_{ba}^{(R)}(\alpha_s, z) \right) .$$

Here $a, b$ are flavor indices, $\alpha_s$ is the strong coupling at a scale being a function of $\mu^2$ to be specified in Section 3, $z$ is the longitudinal momentum splitting variable, and $z_M < 1$ is the soft-gluon resolution parameter. For easier reading we use the notation $\Delta_a(\mu^2) = \Delta_a(z_M, \mu^2, \mu_0^2)$. The form factors eq. (1) have the interpretation of probabilities for non-resolvable branchings between the evolution scales $\mu_0$ and $\mu$. The functions $P_{ba}^{(R)}(\alpha_s, z)$ have
Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

the structure

\[ P_{ba}^{(R)}(\alpha_s, z) = \delta_{ba} k_b(\alpha_s) \frac{1}{1 - z} + R_{ba}(\alpha_s, z), \]  

(2)

where the first term on the right hand side contains the pole singularity in the soft-gluon radiation region \( z \to 1 \) and the second term contains logarithmic terms and analytic terms for \( z \to 1 \). The coefficients \( k_b \) and \( R_{ba} \) in eq. (2) have the perturbation series expansions

\[ k_b(\alpha_s) = \sum_{n=1}^{\infty} \left( \frac{\alpha_s}{2\pi} \right)^n k_b^{(n-1)}, \quad R_{ba}(\alpha_s, z) = \sum_{n=1}^{\infty} \left( \frac{\alpha_s}{2\pi} \right)^n R_{ba}^{(n-1)}(z). \]  

(3)

The explicit expressions for the \( n = 1 \) (LO) and \( n = 2 \) (NLO) contributions in the expansions in eq. (3) are given in [16]. The \( n = 3 \) (NNLO) contributions can be read from [7, 8] and are used for NNLO calculations in the PB method in [26]. The integrals appearing in the Sudakov form factors eq. (1) are positive at LO, NLO and NNLO, while the functions eq. (2) can be negative at NLO and NNLO. The positivity of the integrals in eq. (1) is essential for the application of the PB method.

The PB method allows one to take into account simultaneously soft-gluon emission in the region \( z \to 1 \) and transverse momentum \( q' \) recoils in the parton branchings along the QCD cascade. Its advantage is twofold: on one hand, in collinear distributions additional QCD features can be studied such as the color radiation’s angular ordering, determined by soft-gluon interferences, and its effects on factorization and renormalization scales; on the other hand, the method can be applied to obtain transverse momentum dependent (TMD) distributions.

The PB evolution equations for TMD parton densities \( A_a(x, k, \mu^2) \) are given by [16]

\[ A_a(x, k, \mu^2) = \Delta_a(\mu^2) A_a(x, k, \mu_0^2) + \sum_b \int \frac{d^2q'}{\pi q'^2} \frac{\Delta_a(\mu^2)}{\Delta_a(q'^2)} \Theta(\mu^2 - q'^2) \Theta(q'^2 - \mu_0^2) \]

\[ \times \int_x^{zM} \frac{dz}{z} P_{ab}^{(R)}(\alpha_s, z) A_b \left( \frac{x}{z}, k + (1 - z)q', q'^2 \right), \]  

(4)

in terms of the \( \Delta_a \) form factors, eq. (1), and \( P_{ba}^{(R)} \) functions, eq. (2). The scale in \( \alpha_s \) is a function of \( q'^2 \), as discussed in Section 3. These equations can be solved by an iterative Monte Carlo method. In this method every resolvable branching is reconstructed explicitly and the full kinematics at each branching is taken into account. The PB method allows to solve eq. (4) in an easy and direct way, with the possibility to include, for example, also heavy quark masses and soft-gluon coherence conditions.

The collinear parton densities \( f_a(x, \mu^2) \) are related to the TMD densities by

\[ f_a(x, \mu^2) = \int A_a(x, k, \mu^2) \frac{d^2k}{\pi}, \]  

(5)
and are described as integrated TMD (iTMD). The evolution equations for iTMD densities analogous to eq. (4) can be written as

\[ f_a(x, \mu^2) = \Delta_a(\mu^2) f_a(x, \mu_0^2) + \sum_b \int_{\mu_0^2}^{\mu^2} \frac{d\mu^2}{\mu^2} \Delta_a(\mu^2) \int_x^{z_M} \frac{dz}{z} P_{ab}(z, \alpha_s) f_b \left( \frac{x}{z}, \mu^2 \right) . \]  
(6)

These equations have been shown to be equivalent to DGLAP evolution equations at NLO [15, 16, 20, 21] and NNLO [26] for \( \alpha_s = \alpha_s(\mu^2) \) and \( z_M \to 1 \).

### 2.2 PB method and determination of initial distribution

The PB method has been implemented in the \texttt{xFitter} package [27] to allow fits to be made to cross section measurements. A full Monte Carlo solution of the evolution equation for every set of initial parameters would be too time consuming to be efficient. Instead, a method developed already in [28–30] is applied: first, a kernel \( K_{ba}^{int} \left( x'', \mu_0^2, \mu^2 \right) \) is determined from the Monte Carlo solution of the evolution equation for any initial parton \( a \) evolving to a final parton of flavor \( b \); then this kernel is folded with the non-perturbative starting distribution \( f_{0,b}(x, \mu_0^2) \),

\[ x f_a(x, \mu^2) = x \int dx' \int dx'' f_{0,b}(x', \mu_0^2) K_{ba}^{int} \left( x'', \mu_0^2, \mu^2 \right) \delta(x'' - x) \]

\[ = \int dx' f_{0,b}(x', \mu_0^2) \frac{x}{x'} K_{ba}^{int} \left( \frac{x}{x'}, \mu_0^2, \mu^2 \right) . \]  
(7)

The kernel \( K_{ba}^{int} \) includes the full parton evolution from \( \mu_0^2 \) to \( \mu^2 \), as in eq. (6), with Sudakov form factors and splitting probabilities, and is determined with the PB method. In eq. (7) the kernel \( K_{ba}^{int} \) depends on \( x, \mu_0^2 \) and \( \mu^2 \) for the \( k_t \)-integrated (iTMD) distributions.

To include also the transverse momentum \( k_t \), we define a new kernel \( K_{ba} \left( x'', k_{t,0}^2, k_t^2, \mu_0^2, \mu^2 \right) \) for the TMD distributions, with \( k_t^2 = k^2 \),

\[ x A_a(x, k_t^2, \mu^2) = x \int dx' \int dx'' A_{0,b}(x', k_{t,0}^2, \mu_0^2) K_{ba} \left( x'', k_{t,0}^2, k_t^2, \mu_0^2, \mu^2 \right) \delta(x'' - x) \]

\[ = \int dx' A_{0,b}(x', k_{t,0}^2, \mu_0^2) \frac{x}{x'} K_{ba} \left( \frac{x}{x'}, k_{t,0}^2, k_t^2, \mu_0^2, \mu^2 \right) . \]  
(8)

The evolution of the kernel starts at \( x_0 = 1 \) at \( \mu_0^2 \). In general, the starting distribution \( A_0 \) can have flavor and \( x \) dependent \( k_{t,0} \) distributions, for simplicity we use here a factorized form:

\[ A_{0,b}(x, k_{t,0}^2, \mu_0^2) = f_{0,b}(x, \mu_0^2) \cdot \exp(-|k_{t,0}^2|/\sigma^2) \]  
(9)

\(^{1}\)In practice, since the initial state partons can be only light quarks or gluons, it is enough to determine the kernel \( K \) only for one initial state quark and a gluon.
where the intrinsic $k_{t,0}$ distribution is given by a Gauss distribution with $\sigma^2 = q_0^2/2$ for all flavors and all $x$ with a constant value $q_0 = 0.5 \text{ GeV}$.

Technically, the results of the kernel evolution are stored in a grid of size $50 \times 50 (\times 50)$ (for the TMD densities). The grid spacing is logarithmic ($\mu_0 < \mu < 14000 \text{ GeV}$ and $0.01 < k_t < 14000 \text{ GeV}$), the $x$ range is divided into 5 subregions with logarithmic spacing: subregions of 10 bins are defined with the boundaries $10^{-6}, 0.01, 0.1, 0.4, 0.9, 1$ which is optimized to ensure appropriate behavior for large $x$, where the parton densities (and the kernel) are varying rapidly.

In Fig. 1 we show the result of convoluting the starting distribution (here taken to be the benchmark parameterization of Ref. [31]) with the kernel as given in eq. (7) for the integrated distribution, and compare this with the prediction from a standard evolution program (QCDNUM) for different values of the evolution scale $\mu^2$. The kernel is evolved using NLO splitting functions with resolution scale parameter $z_M$, separating resolvable from non-resolvable branchings, set to the value $z_M = 0.99999$. Very good agreement is observed over the whole range. Only the quark distribution shows differences at very large $x$ of the order of a few percent, which come from the finite grid spacing in $x$ when storing the kernel (changing to a uniform logarithmic grid spacing in $x$ leads to significantly larger deviations at large $x$). In most of the phase space region relevant for high precision physics at HERA and the LHC the differences are at the per mille level.

![Figure 1: Comparison of the results from the convolution in eq.(7) with the prediction from QCDNUM [32] using the same input distributions, for d-quarks (left) and gluons (right) at different values of the evolution scale $\mu^2$ starting from $\mu_0^2 = 1.9 \text{ GeV}^2$ with $\alpha_s(\mu^2)$. The lower panels show the ratio of the parton density with the one predicted by QCDNUM. The evolution is performed with NLO DGLAP splitting functions and using $z_M = 0.99999$.]

Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method
3 Parton densities obtained from fits to inclusive HERA DIS measurements

The most recent and most precise measurements of the lepton-proton DIS cross section over a wide range in $x$ and $Q^2$ were performed at HERA with a combination of the measurements from the H1 and ZEUS collaborations [11]. These measurements are the basis for any determination of parton densities. In Ref. [11] a fit to the inclusive DIS measurements was performed using DGLAP at LO, NLO and NNLO, resulting in the HERAPDF2.0 parton distributions. These fits were performed with QCDNUM [32] within the xFitter framework [27] using a starting scale $\mu_0 = 1.9$ GeV$^2$ and the renormalization and factorization scales set to $\mu^2_f = Q^2$. The light quark matrix elements were taken from QCDNUM, the heavy-quark contributions were obtained within the general-mass variable-flavor scheme RTOPT [33–35] for neutral current, while for charged current interactions the zero-mass approximation from QCDNUM was used. The mass of the charm quark is set $m_c = 1.47$ GeV, and $m_b = 4.5$ GeV is used for the bottom quark mass. The strong coupling is set to $\alpha_s(M_Z^2) = 0.118$.

The parameterized PDFs are the gluon distribution, $xg$, the valence-quark distributions, $xu_v, xd_v$, and the $u$-type and $d$-type anti-quark distributions, $x\bar{U}, x\bar{D}$. The relations $x\bar{U} = x\bar{u}$ and $x\bar{D} = x\bar{d} + x\bar{s}$ are assumed at the starting scale $\mu_0$.

The following parameterizations are used for the different parton flavors:

$$
\begin{align*}
  xg(x) &= A_g x^{B_g} (1 - x)^{C_g} - A'_g x^{B'_g} (1 - x)^{C'_g}, \\
  xu_v(x) &= A_{u_v} x^{B_{u_v}} (1 - x)^{C_{u_v}} \left(1 + E_{u_v} x^2\right), \\
  xd_v(x) &= A_{d_v} x^{B_{d_v}} (1 - x)^{C_{d_v}}, \\
  x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}} (1 - x)^{C_{\bar{U}}} \left(1 + D_{\bar{U}} x\right), \\
  x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}} (1 - x)^{C_{\bar{D}}}. 
\end{align*}
$$

The quark-number sum rules and the momentum sum rule can be used to constrain the normalization parameters, $A_{u_v}, A_{d_v}, A_g, A'_g$. The $B$ parameters are set $B_{\bar{U}} = B_{\bar{D}}$ for the sea distributions. The strange-quark distribution is parameterized as a $d$-type sea with an $x$-independent fraction, $f_s x\bar{s} = f_s x\bar{D}$ at $\mu_0^2$ with $f_s = 0.4$. A further constraint was applied by setting $A_{\bar{D}} = A_{\bar{D}}(1 - f_s)$.

A total of 1145 data points of neutral-current and charged-current deep-inelastic cross section measurements were used in the range of $3.5 < Q^2 < 50000$ GeV$^2$ and $4 \cdot 10^{-5} < x < 0.65$.

The same data sets, kinematic ranges and hard-scattering coefficient functions, including the heavy-quark treatment, are used for the fits described here. We use NLO DGLAP splitting functions [5,6] as well as NLO coefficient functions [36] for light quarks. For heavy quarks we apply the general-mass variable-flavor scheme RTOPT [33–35] for neutral current, while for charged current interactions the zero-mass approximation is used.

In the next section we determine the free parameters of the initial distributions given by
Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

eq. (10) via fits to the HERA DIS data in the range of $Q^2 > 3.5$ GeV$^2$ using NLO DGLAP splitting functions within the PB method using $z_M = 0.99999$.

\[
\begin{align*}
x_{ip}, k_{it} &
\end{align*}
\]

Figure 2: Left: Branching process $b \rightarrow a + c$. Right: Schematic view of a parton branching process.

<table>
<thead>
<tr>
<th>PB NLO Set1 $\alpha_s(\mu_i^2)$</th>
<th>$\chi^2$</th>
<th>d.o.f</th>
<th>$\chi^2$/d.o.f</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0 = 1.9$ GeV$^2$</td>
<td>1363.37</td>
<td>1131</td>
<td>1.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PB NLO Set 2 $\alpha_s(q_{li}^2)$</th>
<th>$\chi^2$</th>
<th>d.o.f</th>
<th>$\chi^2$/d.o.f</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0 = 1.4$ GeV$^2$</td>
<td>1369.80</td>
<td>1131</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Table 1: Values of $\chi^2$ for the different fits at NLO.

The PB method allows the explicit calculation of the kinematics at every branching vertex (see Fig. 2 left). Once the physical meaning of the evolution scale is specified in terms of kinematic variables, the transverse momenta of the propagating and emitted partons can be calculated. In Ref. [15] it was pointed out that angular ordering gives transverse momentum distributions which are stable with respect to variations of the resolution parameter $z_M$. In angular ordering, the angles of the emitted partons increase from the hadron side towards the hard scattering, as shown in Fig. 2 right. The transverse momentum $q_{li}$ of the emitted parton with respect to the beam directions from $q_{li} = (1 - z_i)E_i \sin \Theta_i$. Associating the “angle” $E_i \sin \Theta_i$ with $\mu_i$ gives

\[
q_{li}^2 = (1 - z_i)^2 \mu_i^2.
\]

In the following, we use the PB method to determine collinear (iTMD) and transverse momentum dependent (TMD) parton densities using NLO DGLAP splitting functions for two different scenarios: first we only apply the angular ordering condition for the calculation of the transverse momentum and keep the evolution scale $\mu_i^2$ as the argument in $\alpha_s$ (Set 1); in a second scenario (Set 2), we use (in eqs. (1,4,6)) the transverse momentum $|q_{li}^2|$ as the argument in $\alpha_s$, as suggested in Ref. [17, 18]. An additional parameter $q_{cut}$ needs to
Figure 3: Parton densities for different values of the scale $\mu^2 = Q^2$. The different choices for the renormalization scale in $\alpha_s$ are shown. The red band shows the experimental uncertainty, the yellow band the model dependence. The green band shows the uncertainty coming from the variation of the parameter $q_{cut}$ in Set 2.

be introduced in $\alpha_s(\max(q_{cut}^2, |q_{t,i}^2|))$ to avoid the non-perturbative region, since with large $z$ the scale $|q_{t,i}^2| = (1 - z_i)^2 \mu_i^2$ can become very small. We take the default choice for this parameter to be $q_{cut} = 1$ GeV, and we estimate the model dependence with a variation around the default choice.

In the first case, the integrated parton density, and the initial parameters, will be the same (up to numerical precision) as the ones obtained by HERAPDF2.0, and we use this as a benchmark for the whole method. In the second case, even the integrated parton distributions differ, because of the different scale in $\alpha_s$. In both cases a reasonably good fit is obtained with $\chi^2/ndf \sim 1.2$, as for HERAPDF2.0. In Tab. 1 results of the fits are given. The starting scale $\mu_0^2$ is chosen differently for the 2 scenarios: for Set 1 we chose (as in HERAPDF) $\mu_0^2 = 1.9$ GeV$^2$ while for Set 2 we chose $\mu_0^2 = 1.4$ GeV$^2$, which gave the best $\chi^2/ndf$. In the appendix we show results obtained from a fit when $\mu_0^2 = 1.9$ GeV$^2$ is chosen instead of $\mu_0^2 = 1.4$ GeV$^2$. The distributions agree within their uncertainties. The values of the parameters at the starting scale $\mu_0^2$ are given in Tab. 2.
Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

PB NLO Set1 $\alpha_s(\mu_t^2)$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>$A'$</th>
<th>$B'$</th>
<th>$C'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$xg$</td>
<td>4.32</td>
<td>-0.015</td>
<td>9.15</td>
<td>13.5</td>
<td>1.040</td>
<td>-0.166</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>$xu_v$</td>
<td>4.07</td>
<td>0.714</td>
<td>4.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$xd_v$</td>
<td>3.15</td>
<td>0.806</td>
<td>4.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x\bar{U}$</td>
<td>0.107</td>
<td>-0.173</td>
<td>8.05</td>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x\bar{D}$</td>
<td>0.178</td>
<td>-0.173</td>
<td>4.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PB NLO Set 2 $\alpha_s(q_t^2)$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>$A'$</th>
<th>$B'$</th>
<th>$C'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$xg$</td>
<td>0.42</td>
<td>-0.047</td>
<td>0.96</td>
<td>13.7</td>
<td>0.008</td>
<td>-0.58</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>$xu_v$</td>
<td>2.49</td>
<td>0.65</td>
<td>3.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$xd_v$</td>
<td>2.02</td>
<td>0.75</td>
<td>2.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x\bar{U}$</td>
<td>0.14</td>
<td>-0.16</td>
<td>5.29</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x\bar{D}$</td>
<td>0.24</td>
<td>-0.16</td>
<td>5.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Parameter values of the initial distributions at NLO. The parameter $C' = 25$ was fixed, as in HERAPDF2.0. The parameters correspond to a starting scale $\mu^2_0 = 1.9(1.4)$ GeV$^2$ for Set 1 (Set 2).

3.1 Collinear Parton Densities (ITMD)

The fits to HERA measurements are performed using $\chi^2$ minimization, as in the case of the HERAPDF fits, implemented in xFitter [27]. The definition of $\chi^2$ includes systematic shifts, a treatment of correlated and uncorrelated systematic uncertainties. In total 162 systematic uncertainties plus procedural uncertainties from the combination of H1 and ZEUS are treated as correlated uncertainties.

<table>
<thead>
<tr>
<th></th>
<th>Central value</th>
<th>Lower value</th>
<th>Upper value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB NLO Set1 $\mu^2_0$ (GeV$^2$)</td>
<td>1.9</td>
<td>1.6</td>
<td>2.2</td>
</tr>
<tr>
<td>PB NLO Set 2 $\mu^2_0$ (GeV$^2$)</td>
<td>1.4</td>
<td>1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>PB NLO Set 2 $q_{cut}$ (GeV)</td>
<td>1.0</td>
<td>0.9</td>
<td>1.1</td>
</tr>
<tr>
<td>$m_c$ (GeV)</td>
<td>1.47</td>
<td>1.41</td>
<td>1.53</td>
</tr>
<tr>
<td>$m_b$ (GeV)</td>
<td>4.5</td>
<td>4.25</td>
<td>4.75</td>
</tr>
</tbody>
</table>

Table 3: Central values and change ranges of parameters for model dependence

The experimental uncertainties of the resulting parton densities are determined with the Hessian method [37] (as implemented in xFitter) with $\Delta \chi^2 = 1$. The model dependence of the PDF fits is obtained by varying charm and bottom masses and the starting scale of the evolution $\mu^2_0$. For Set 2 also the parameter $q_{cut}$ is varied. The central values and the range of variation is given in Tab. 3.

In Fig. 3 the $\bar{U}$-type quark and gluon densities are shown as functions of $x$ for different values of the evolution scale $\mu^2 = Q^2$ including the experimental uncertainties (red band)
Figure 4: Total uncertainties (experimental and model uncertainties) for the two different sets at different values of the evolution scale $\mu^2$.

and the uncertainties coming from the model dependence (yellow band). For Set 2 the uncertainty of the parameter $\alpha_{\text{cut}}$ is shown as the green band. The results of Set 1 are identical to the ones obtained in HERAPDF 2.0. Although the fits (Set 1 and Set 2) to HERA I+II data are of similar quality, the resulting parton distributions, especially for the gluon, are significantly different. With increasing evolution scale, however, they become more and more similar.

In Fig. 4 the total uncertainties (experimental and model) of the parton densities are shown. The uncertainties of Set 2 for the gluon distribution at large $x$ become large. We have investigated a possible bias coming from the chosen form of the parameterization by including additional terms for the gluon density:

$$xg(x) = A_g x^{B_g} (1 - x)^{C_g} \left(1 + D_g x + E_g x^2\right) - A'_g x^{B'_g} (1 - x)^{C'_g}.$$ 

The obtained $\chi^2$ of the fit does change by at most 1 unit, the resulting gluon distribution does not change visibly. Details of the bias study are given in the appendix.

In Fig. 5 we show predictions for the inclusive DIS cross section and the inclusive charm cross section obtained from the two different parton distributions, and compare them with the measurements from HERA [11, 38]. While the inclusive DIS cross section is well described, the prediction using Set 2 differs from inclusive charm measurement at low $Q^2$ and
Figure 5: Measurement of the reduced cross section obtained at HERA compared to predictions using Set 1 and Set 2. Upper row: inclusive DIS cross section [11], lower row: inclusive charm production [38]. The dashed lines include the systematic shifts in the theory prediction.

small $x$. For values $x > 0.001$ all predictions agree reasonably well with the data. It has been checked explicitly that including thecharm measurements in the fits does not significantly change the fit result (the charm data have too large an uncertainty compared to the precise inclusive measurements). In Fig. 5 the predictions including the systematic shifts are also shown, visually showing that the quality of the two different fits is similar.

### 3.2 Transverse Momentum Dependent Parton Densities (TMD)

Within the PB method both collinear and TMD densities can be determined, as the transverse momentum is calculated at every step of the branching process. TMD parton densities can be obtained via the PB method once the relationship between kinematical variables and evolution scale $\mu$ is specified, and the transverse momentum at each individual branching is calculated with eq. (11). The parameters for the starting distributions are obtained for the collinear parton densities by a fit to inclusive DIS cross section measurements, as described previously. The TMD parton densities are then obtained from a convolution of the TMD kernel with the starting distribution as given in eq. (8). The starting distribution is taken from the collinear iTMD described in Sec. 3.1.
In Fig. 6 we show the TMD parton densities for $\bar{u}$-quarks and gluons as a function of the transverse momentum $k_t = \sqrt{k_x^2}$ for different values of the evolution scale $\mu = 10, 100, 1000$ GeV and different values of $x$ for Set 1 and Set 2. One can clearly see that both sets give identical results for larger $k_t$, while they are different for small $k_t$, a consequence of the different scale choices for the argument of $\alpha_s$.

In Fig. 7 the parton densities for all flavors are shown as a function of $k_t$ at $x = 0.01$ and for different values of the evolution scale $\mu = 10, 100, 1000$ GeV. The large scales are relevant for phenomenology at the LHC, and it is interesting to observe that the transverse momenta extend to very large values, up to the values of the factorization scales (for $\mu = 1$ TeV the transverse momenta extend to $k_t \sim 1$ TeV). However, the large $k_t$ values are suppressed compared to smaller ones. The different quark flavors show a different behavior at small $k_t$, coming essentially from the no-branching probability times the starting distribution (first
Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

term in eq. (4)), while they are very similar at larger \( k_t \), a result of perturbative splittings (second term in eq. (4)).

![Figure 7: Transverse Momentum Dependent parton densities (PB-NLO-2018-Set1 upper row and PB-NLO-2018-Set2 lower row) as a function of \( k_t \) for different scales \( \mu \) at \( x = 0.01 \) for all flavors.](image)

In Fig. 8 the gluon and \( \bar{u} \) densities as a function of the transverse momentum are shown for \( \mu = 100 \) GeV and \( x = 0.01 \) together with the uncertainty bands obtained from the fits. The panels show the uncertainties coming from the experimental sources as well as the total uncertainty coming from experimental and model sources separately. Although only collinear splitting functions are used, and the fit was obtained with collinear parton densities, a \( k_t \) dependence of the uncertainties is obtained. At small \( k_t \) essentially the first term in eq. (4) contributes without any resolvable branching and the uncertainty comes from the starting distribution at \( x \), while at large \( k_t \) several branching may have occurred and therefore the uncertainty comes from the starting distribution at \( x/z \gg x \). The experimental uncertainties are small over the whole range, while the model dependent uncertainties dominate.

The parametrization of the intrinsic transverse momentum distribution is another uncertainty. With the fit to inclusive DIS data, this distribution cannot be further constrained. In Fig. 9 we show the TMD distribution for gluon and \( \bar{u} \) for Set 1 and Set 2 at \( \mu = 10(100) \) GeV and \( x = 0.01 \) when \( q_0 \) in \( B(k^2, \mu^2) \) is varied from \( q_0 = 0.25 \) GeV to \( q_0 = 1 \) GeV. We do not include the variation of \( q_0 \) as a systematic uncertainty, since it is not constrained by the fit (in future we plan to use also Z-boson transverse momentum spectra, which would constrain \( q_0 \)).

The resulting TMD parton densities, PB-NLO-2018-Set1 and PB-NLO-2018-Set2, including uncertainties (as well as with variation of \( q_0 \)) are available in TMDLIB [39]. The TMD-
Figure 8: Transverse Momentum Dependent parton densities for $\bar{u}$ and gluon from Set 1 and Set 2 as a function of $k_t$ for $\mu = 100$ GeV at $x = 0.01$. In the lower panels we show the relative uncertainties coming from experimental uncertainties as well as the total of experimental and model uncertainties.

PLOTTER [40, 41] interface allows easy and fast comparison to other TMDs, once they are made publicly accessible and available in TMDlib.

4 Application to $Z$-boson production at the LHC

The transverse momentum spectrum of $Z$ bosons in Drell-Yan (DY) production at small values of transverse momentum $q_T$ cannot be described by fixed-order perturbative calculations, and resummation of soft gluon emissions to all orders in $\alpha_s$ is needed. See e.g. [42] for a recent discussion. The DY $q_T$ spectrum can be described by the CSS method [43–46] using TMD factorization at small $q_T$ [47, 48], or by parton showers within Monte Carlo event generators [25]. The ATLAS and CMS experiments at the LHC have measured the $q_T$ spectrum of the $Z$-boson [49–51].

The TMD distributions obtained from HERA DIS measurements can be used to predict the DY $q_T$ spectrum of the $Z$-boson at LHC energies. Since we are interested in the low-$q_T$
Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method

Figure 9: Transverse Momentum Dependent parton densities ($\bar{u}$ and gluon) from Set 1 and Set 2 as a function of $k_t$ for $\mu = 10(100)$ GeV at $x = 0.01$, when the width of the intrinsic transverse momentum distribution is varied by a factor of two.

region, we use the LO expression for $Z$ production matrix elements.\(^1\) The transverse momentum of the initial state partons is calculated according to the TMDs and added to the event record in such a way that the mass of the produced DY pair is conserved, while the longitudinal momenta are changed accordingly. This procedure is common in standard parton shower approaches [53,54] and is implemented in the CASCADE package [55,56] (version newer than 2.4.x) where events in HEPMC [57] format are produced, for further processing with Rivet [58]. The importance of the proper inclusion of transverse momentum effects from parton showers has been pointed out in Ref. [59,60]. With the TMD distributions described here, these effects can be included already at the level of the cross section calculation.

In Fig. 10 (left) we show the predictions for the transverse momentum spectrum of the $Z$-boson obtained with the two TMD distributions, compared with the measurements of ATLAS [51]. The uncertainties coming from experimental and model sources are shown for both Set 1 and Set 2 with the colored bands (Fig. 10 left); the experimental and full uncertainties are shown for Set 2 in Fig. 10 (right). The difference between the full and experimental uncertainties from the fit is very small.

In general the shape of the spectrum is described by both TMD fits. The TMD Set 2, applying the transverse momentum as the renormalisation scale (instead of the evolution scale $\mu$), provides a significantly better description of the transverse momentum spectrum of the $Z$-boson, coming from the different $k_t$ spectrum of the TMD already visible in Fig. 6. One should note that no adjustment of any parameter is made, and that the TMDs are entirely constrained by the fits to inclusive DIS data. The description of the transverse mo-

\(^1\)In practical terms we use an LHE (Les-Houches Event) file [52] for $qq \to Z$ obtained from the PYTHIA event generator [53] with on-shell initial partons.
Figure 10: Transverse momentum $q_T$ spectrum of $Z$-bosons obtained from the two TMDs, compared with measurements from [51]. Left: comparison of predictions using Set 1 and Set 2 including the full (experimental and model) uncertainties. Right: prediction using Set 2, with experimental and full uncertainties separated (the difference is very small).

5 Conclusion

The parton branching method has been used to determine a first complete set of collinear and TMD parton densities from fits to precision DIS data over a large range in $x$ and $Q^2$ as measured at HERA. The parton densities are obtained with NLO DGLAP splitting functions and 2-loop $\alpha_s$ with $\alpha_s(M_Z) = 0.118$. The renormalisation scale in the evolution has been chosen to be the evolution scale $\mu_i$ (Set 1) or the transverse momentum $q_T$ (Set 2). Two different collinear and TMD sets are obtained for these different choices, both giving a similar $\chi^2/ndf = 1.2$. The obtained parton densities are valid over a wide range in $x$ and scale $\mu$, up to the multi-TeV scale, relevant for LHC physics.

Experimental uncertainties of the fit are obtained using the Hessian method with $\Delta \chi^2 = 1$ and model dependent uncertainties are determined.

The obtained TMDs are applied to calculate the transverse momentum spectrum of the $Z$-boson in DY production at LHC energies. Good agreement with the measurement is observed if angular ordering is applied. The uncertainties of the prediction come only from the
TMD uncertainties determined in the fit to HERA measurements.

For the first time, precision DIS measurements have been used to obtain both collinear and TMD parton densities, including uncertainties, over a wide range in $x$ and $\mu$ values, which are relevant for LHC and future collider phenomenology as well as for low-energy and small-$k_t$ physics.

**Acknowledgments.** We are grateful for many discussions with the xFitter developers team, in particular with R. Placakyte and A. Glazov. FH acknowledges the support and hospitality of the CERN Theory Division, of DESY, Hamburg, while part of this work was being done. HJU thanks the Polish Science and Humboldt Foundations for the Humboldt Research fellowship during which part of this work was completed.

**Appendix**

In Fig. 11 we show a comparison of the gluon density of Set 2 ($\mu_0^2 = 1.4 \text{ GeV}^2$) with a gluon density obtained using starting scale $\mu_0^2 = 1.9 \text{ GeV}^2$ (all other settings are the same as in Set 2) at a scale of $Q^2 = 3 \text{ GeV}^2$. The fit with a starting scale $\mu_0^2 = 1.9 \text{ GeV}^2$ gives a $\chi^2 = 1402.4$ compared to $\chi^2 = 1369.8$ when using $\mu_0^2 = 1.4 \text{ GeV}^2$. The uncertainties for the new fit include only the uncertainties from experimental sources, the uncertainties for Set 2 are the same as in Fig 4. Both sets agree within uncertainties.

**Figure 11:** Comparison of gluon density of Set 2 type obtained at $\mu_0^2 = 1.4 \text{ GeV}^2$ and $\mu_0^2 = 1.9 \text{ GeV}^2$ at a scale of $Q^2 = 3 \text{ GeV}^2$. The ratio of the gluon densities is shown with respect to the default Set 2. The uncertainties for the new fit include only those from experimental sources, the uncertainties for Set 2 are the same as in Fig 4.

---

**Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method**
A potential bias of the form of the parameterization was checked by extending the original parameterization 

\[ xg(x) = A_g x B_g (1 - x)^C_g - A'_g x B'_g (1 - x)^C'_g \]

with additional parameters:

\[ xg(x) = A_g x B_g (1 - x)^C_g (1 + D_g x + E_g x^2) - A'_g x B'_g (1 - x)^C'_g . \]

In Fig. 12 we show the gluon distribution after fitting \( C'_g \) and including the additional factors \( D_g \) and \( E_g \) one after the other. The starting scale is \( \mu_0^2 = 1.4 \text{ GeV}^2 \) (as for the original fit Set 2). The obtained \( \chi^2 \) is larger by 1 unit after including additional terms, the shape of the distribution does not change significantly. The uncertainty band of Set 2 corresponds to the uncertainties coming form the experimental sources, no model or parameterization uncertainty is included. The parton distributions agree within the uncertainties shown, excluding a significant bias from the chosen form of the parametrization.

Figure 12: Comparison of gluon densities after fit when additional terms in the gluon parameterization are included. The uncertainty band of Set 2 corresponds to the uncertainties coming form the experimental sources.

References


Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method


Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method


Appendix C. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method


In this section I present an alternative approach, which I call response matrix parametrization method, for constructing response matrices starting from the CMS fully simulated MC samples. The method was already checked and used in the resolution studies for the measurements reported in Apps. A and B.

This method allows to reproduce the response matrices used to unfold the data, with the important features that the fluctuations in the response matrix which usually affect the stability of the unfolding are drastically reduced, and also that the characteristics of the new response matrices are controlled by a few parameters. The method can therefore be used to easily estimate the resolution systematic uncertainties by varying the matrices parametrization (Sec. 7.3.1).

This new procedure has important advantages with respect to the method which is currently used in CMS called toy MC method or forward smearing. The toy MC method is based on the use of the resolution curves obtained from the CMS fully simulated MC samples to smear the GEN level distribution, and obtain the corresponding toy response matrix. The main disadvantage of this method is the non-trivial form of the resolution curves, generally asymmetric and varying differentially in the observable, which makes the fitting difficult specially for the tail of the distributions. This can be observed in Fig. D.1, where the response matrix obtained from CMS fully simulated MC samples is shown. The resolution curves are projections of the response matrix onto the DET level axis. One can notice in Fig. D.1 that the resulting resolution curves (represented by the dashed region) are not symmetric since the there are clearly more events on the right of the diagonal.

As a consequence of the difficult fitting of the resolution curves, especially the tails, the
Appendix D. New method for migration matrix determination and resolution effects estimation

Figure D.1: Original response matrix. The projection of the slice onto the $x_1$ axis represents the resolution curve.

response matrices obtained with the toy MC method usually can only describe the migration to few neighbouring bins.

The new method presented in this section uses the symmetry properties of the response matrices in order to reproduce the migration and correlation between bins far away from the diagonal.

The general procedure is presented below followed by a discussion of its implementation and advantages:

• One starts by projecting the response matrix from the CMS fully simulated MC samples onto the diagonal axis and the axis orthogonal to it (represented on the left plot in Fig. D.2).

• From the projection of the response matrix onto the symmetry axis one obtains what I call the rotated GEN level distribution (represented by the top right distribution in Fig. D.2).

• From the projections of the response matrix onto the axis orthogonal to the symmetry axis one obtains what I call the rotated resolution curves (represented by the bottom right distribution in Fig. D.2).

• The new response matrix is then generated by smearing the rotated GEN level distribution using the rotated resolution curves.

• The DET level distribution is then obtained by folding the original GEN level with the new response matrix determined in the previous bullet.

• Variations of the migration effects can be easily implemented as variations of the parametrization of the rotated resolution curve. These effects can be then consistently propagated to the observable by folding the original GEN level with the matrix variations.

As illustrated in Fig. D.2, the rotated resolution curve is symmetric and generally easy to fit (as will be observed in the next section).
Figure D.2: Example of a generic response matrix (left) and its projections onto its symmetry axis (represented by a gray band) and the axis orthogonal to it (represented by a black band). The projected distributions are indicated on the right of the figure.

**Implementation of the method**

As an example, Fig. D.4b shows the official response matrix from the CMS fully simulated MC samples for the inclusive 2-jet cross section as a function of $\Delta\phi_{12}$ for $500 < p_T^{\text{max}} < 600$ GeV. The response matrices is symmetric with respect to the diagonal axis.

By rotating the matrix by 45° counterclockwise one obtains the corresponding rotated matrix shown in Fig. D.4a.

The projections onto the $x$ and $y$ axes of the rotated matrix in Fig. D.3b clearly correspond to the projections onto the diagonal axes of the original matrix in Fig. D.3a. One can now implement the items stated in the previous section corresponding to the determination of the new response matrices. Figure D.4a shows the rotated GEN level distribution obtained by projecting the matrix in Fig. D.4b onto the axis represented by the black dashed line.

Similarly, Fig. D.5a shows the rotated resolution distribution obtained by projecting the dashed region of the matrix in Fig. D.5b onto its $x$-axis.

The rotated resolution distribution shown in Fig. D.5a was fitted using the simple functional form in Eq. 6.5. It is interesting to notice the very good fit by looking at the corresponding $\chi^2/\text{ndf}$. The result of the fit is impressive given the fact that the distribution spans many orders of magnitude. The reason for this is that the use of the symmetry axes for projecting the matrix allowed a simple symmetric form for the projections of the rotated matrix. One can then use MC methods to obtain the new response matrix by smearing the rotated
Appendix D. New method for migration matrix determination and resolution effects estimation

Figure D.3: a) Original response matrix for the inclusive 2-jet distribution as a function of $\Delta \phi_{12}$, for $500 < p_T^{\text{max}} < 600$ GeV; b) same response matrix after a counterclockwise $45^\circ$ rotation.

Figure D.4: a) Rotated GEN level distribution obtained as the projection of the rotated matrix shown in b). The projection axis is represented by the dashed line.

GEN distribution using the rotated resolution curve. Finally the new response matrix can be rotated back to the initial position in order to be used for the unfolding. In the next section the new response matrices corresponding to the inclusive 2- and 3-jet distributions as a function of $\Delta \phi_{12}$ will be shown and compared to the original matrices.

The resulting new matrix response matrix only depends on the parameters of the fit shown in Fig. D.5a. An important consequence of this is that, by varying the width of the Gaussian distributions one can estimate the impact of the $\Delta \phi_{12}$ resolution on the observables. The effects of the parameter variations are propagated to the DET level by folding the original GEN level prediction with the variations of the new response matrix. In this way the correlations between the GEN and DET levels are preserved. The matrix variations and the corresponding DET level distributions can then be used to estimate the systematic uncertainty due to the limited knowledge of the real $\Delta \phi_{12}$ resolution (see Sec. 7.3.1).
Figure D.5: a) Rotated resolution distribution obtained as the projection onto the $x$-axis of the rotated matrix shown in b). The projected region is represented by the dashed lines. The rotated resolution distribution is fitted using the functional form in Eq. 6.5

New response matrices

The new response matrices and the original response matrices from the official MC Pythia8 tune CUETP8M1 are shown in Figs. D.6 and D.7 for the inclusive 2-jet observables as a function of $\Delta\phi_{12}$, as well as in Figs. D.8 and D.9 for the inclusive 3-jet observables. In Figs. D.6, D.7, D.8 and D.9, the new response matrices correspond to the distributions in the middle column and the original response matrices correspond to the distributions in the left column. The resulting ratios of the new matrices to the official matrices are depicted in the right column of Figs. D.6, D.7, D.8 and D.9. The ratios are around 1 not only near the symmetry axis, but also away from it. The regions where the ratio is away from 1 coincide with the sectors in the original matrices where the statistics is limited.
Appendix D. New method for migration matrix determination and resolution effects estimation

Figure D.6: The response matrices for the inclusive 2-jet observables as a function of $\Delta \phi_{12}$, derived using the official CMS simulated MC samples from PYTHIA8 tune CUETM1 MC are shown in the left column. The response matrices obtained with the new method (middle column) are also shown as well as the ratios (right column) of the new matrices to the official matrices.
Figure D.7: Same as figure D.6 for the remaining $p_T^{\text{max}}$ ranges.
Appendix D. New method for migration matrix determination and resolution effects estimation

![Figure D.8](image-url)

Figure D.8: The response matrices for the inclusive 3-jet observables as a function of $\Delta\phi_{12}$, derived using the official CMS simulated MC samples from PYTHIA8 tune CUETM1 MC are shown in the left column. The response matrices obtained with the new method (middle column) are also shown as well as the ratios (right column) of the new matrices to the official matrices.
Figure D.9: Same as figure D.8 for the remaining $p_T^{\text{max}}$ ranges.
Appendix D. New method for migration matrix determination and resolution effects estimation
In this section the studies I performed on LO multi-jet merging are presented, specifically the investigations on the proper choice of the merging scale in observables which involve multiple scales. Part of the studies presented here corresponds to a summer student project which I supervised (see report in Ref. [172]).

The studies were motivated by the measurements of the inclusive jets $p_T$ spectrum in CMS, where large differences have been observed between data and the LO predictions from MadGraph + Pythia8 multi-jet samples. The inclusive jets $p_T$ spectrum is a benchmark QCD observable. The most recent measurement of this observable in CMS have been performed at the center-of-mass energy of 13 TeV [173]. The measurements have been compared with theoretical predictions from several MC generators at both LO and NLO. However, over the years it was observed that the predictions from LO multi-jet calculations from MadGraph showed large discrepancies compared to the data, both in shape and normalization, that were not observed for the other calculations. This issue was not totally understood, and as a consequence the LO MadGraph predictions were not included in the past CMS inclusive jet $p_T$ measurement publications.

As an example, Fig. E.1 shows the predictions from MadGraph + Pythia8 compared to the data for the inclusive jets $p_T$ spectrum at 13 TeV, for two different rapidity ranges. The predictions from the Pythia8 MC generator are also included in Fig. E.1 as a reference.

The Pythia8 predictions rely on LO $2 \rightarrow 2$ QCD calculation. The MadGraph + Pythia8 predictions are based on merged LO $2 \rightarrow 2$ and $2 \rightarrow 3$ QCD calculations. The MadGraph multiplicities are merged using the same setting which has been used in CMS: MLM merging scheme with a merging scale $q_{cut} = 20$ GeV.
Appendix E. Matching/merging and the inclusive jets transverse momentum cross section

Figure E.1: Double-differential inclusive jet cross section as function of jet $p_T$ for the rapidity ranges a) $0 < |y| < 0.5$, and b) $0.5 < |y| < 1.0$. The predictions from Pythia8 (LO $2 \rightarrow 2$ ME) are shown, as well as the predictions from MadGraph + Pythia8 (LO $2 \rightarrow 2, 3$ ME) where $q_{cut} = 20$ GeV (default in CMS) is used. The solid band in the ratio plot represents the experimental uncertainty.

One can observe in Fig. E.1 that the MadGraph + Pythia8 calculations differ both in shape and normalization with the data, with differences of up to 45%. On the other hand, one can notice that the Pythia8 predictions differ only in normalization with differences of less than 20%. Similarly, it was observed in Ref. [173] that predictions from Herwig++, Powheg-2J + Pythia8 and fixed-order NLO calculations differed only in normalization with the data for the same rapidity ranges used in Fig. E.1.

As discussed in Secs. 3.3.2, 4.2.4, 9.2.4 and 9.4.4, the merging scale corresponds to an arbitrary cut which is set to separate the regions of phase space that will be filled by the different jet multiplicities. The physical observable which is calculated should not depend on the choice of the arbitrary cut. However, we observed in Sec. 3.3.2 that Eq. 3.51 and Eq. 3.52 are not equal when they are evaluated at the same merging scale value $q_{cut}^2$, meaning that if we take one of the expressions to generate events below the cut and the other expression to generate events above the cut we can end up with an unphysical discontinuity at the cut value.

By looking at observables which are sensitive to the merging scale one can test the smoothness of the calculation for the specific merging scale choice. One of the observables which can be used is the $p_T$ distribution of the third hardest jet ($pT_3$), because if the $pT_3$ was below the merging scale it would be generated by the PS from the LO $2 \rightarrow 2$ ME, whereas if it was above the merging scale the third jet would correspond to the LO $2 \rightarrow 3$ contribution.

In Fig. E.2 the prediction from MadGraph + Pythia8 for the $p_T$ of the third jet is shown, using ME with up to three partons in the final state with a merging scale $q_{cut} = 20$ GeV. The hardest scale of the process is restricted to be high by requiring $H_T > 800$ GeV, where $H_T$ is the scalar sum of the $p_T$ of the jets in the event.
Figure E.2: The prediction from MadGraph + Pythia8 (LO $2 \rightarrow 2,3$ ME) is shown for the third hardest jet $p_T$ spectrum (pt3), using $q_{\text{cut}} = 20$ GeV (default in CMS). The corresponding contributions from the different multiplicities are also shown. The error bars represent the statistical error of the predictions.

In Fig. E.2, jet sample $n$ corresponds to events generated by the multiplicity with $n$ partons in the final state. One can observe that the transition between the contributions from the $2 \rightarrow 2$ and $2 \rightarrow 3$ multiplicity samples is not smooth, which results in an unphysical distribution for the pt3 spectrum. This is a clear sign that the merging scale choice is not correct. I observed this pattern for small values of the merging scale compared to the hardest scale of the event ($q_{\text{cut}}/(2H_T) < 10$).

When a higher merging scale is used, a smooth pt3 distribution is achieved. This can be observed in Fig E.3a where $q_{\text{cut}} = 90$ GeV. An additional observable, which is sensitive to the merging scale value corresponds to the differential jet rate (DJR), which characterizes the scale at which the transition between the (n)-jet and (n+1)-jet configurations occurs. In Fig. E.3b the differential jet rate distribution characterizing the transition between 2-jet and 3-jet configurations (d23) is shown. As observed for the pt3 spectrum in Fig. E.3a, a smooth transition between the multiplicities is achieved.

Based on the studies discussed earlier, I recalculated the prediction for the inclusive jets $p_T$ using a merging scale choice that preserves the smoothness of the physical observables pt3 and d23. Since the inclusive jets $p_T$ spectrum covers a wide range in $p_T$ (see Fig. E.1) I used two different $q_{\text{cut}}$ values, depending on the scale of the process in the following way: if $H_T < 1000$ GeV $\Rightarrow q_{\text{cut}} = 50$ GeV, and if $H_T > 1000$ GeV $\Rightarrow q_{\text{cut}} = 140$ GeV. The resulting prediction is compared to data in Fig. E.4b for two rapidity ranges. The corresponding
Figure E.3: The prediction from MadGraph + Pythia8 (with LO $2 \rightarrow 2,3$ ME) using $q_{\text{cut}} = 90$ GeV is shown for a) the third hardest jet $p_T$ spectrum ($pT3$) as well as for b) the differential jet rate distribution of the transition from 2-jet to 3-jet configurations ($d23$). The corresponding contributions from the different multiplicities are also shown. The error bars represent the statistical error of the predictions, and the solid band represents a $\pm 10$ GeV variation of the merging scale.
contributions obtained by requiring HT < 1000 GeV and HT > 1000 GeV are also shown.

As one can observe, with the new choice of merging scales, the inclusive jet measurements in Fig. E.4 are described by the predictions with less than 15% accuracy, differing mainly in normalization.

Choosing a merging scale which is not too small compared to the hardest scale of the process is of big importance. In order to see this, Fig. E.5 shows the predictions using qcut = 20 GeV (default choice in CMS), qcut = 50 GeV, as well as the two-scale choice commented earlier (HT < 1000 GeV ⇒ qcut = 50 GeV, and HT > 1000 GeV ⇒ qcut = 140 GeV).

In Fig. E.5 one can clearly observe that the choice qcut = 50 GeV starts to differ with the data (in shape) at high p_T, where qcut is small compared to the scale of the process. Since for the two-merging scale choice a larger value qcut = 140 GeV is used at high scales, the flat shape of the ratio is recovered at high p_T.

Finally, Fig. E.6 shows the same comparison shown in Fig. E.1 for the inclusive jets p_T spectrum for two rapidity ranges, but this time including the predictions from MADGRAPH + PYTHIA8 using the new scale choice: HT < 1000 GeV ⇒ qcut = 50 GeV, and HT > 1000 GeV ⇒ qcut = 140 GeV.
Appendix E. Matching/merging and the inclusive jets transverse momentum cross section

Data

$q_{\text{cut}} = 20 \text{ GeV (used by CMS)}$
$q_{\text{cut}} = 50 \text{ GeV}$
$q_{\text{cut}} = q_{\text{cut}}(\text{HT})$

CMS, $\sqrt{s} = 13 \text{ TeV}$, Inclusive AK$_4$ jets, $0.0 < |y| < 0.5$

$\frac{d^2\sigma}{dp_T^2} \frac{d\hat{y}}{d\hat{y}} [\text{pb/GeV}]$ $10^3$
$10^2$
$10^1$
$10^0$

MC/Data

(a)

Figure E.5: Double-differential inclusive jet cross section as function of jet $p_T$ for the rapidity ranges a) $0 < |y| < 0.5$, and b) $0.5 < |y| < 1.0$. The predictions from MadGraph + Pythia8 (LO $2 \rightarrow 2, 3$ ME) are shown with the choices $q_{\text{cut}} = 20 \text{ GeV (default in CMS)}$, $q_{\text{cut}} = 50 \text{ GeV}$, and the additional new choice: $HT < 1000 \text{ GeV} \Rightarrow q_{\text{cut}} = 50 \text{ GeV}$, and $HT > 1000 \text{ GeV} \Rightarrow q_{\text{cut}} = 140 \text{ GeV}$. The solid band in the ratio plot represents the experimental uncertainty, and the error bars represent the statistical uncertainty of the predictions.
Figure E.6: Double-differential inclusive jet cross section as function of jet $p_T$ for the rapidity ranges a) $0 < |y| < 0.5$, and b) $0.5 < |y| < 1.0$. The predictions from Pythia8 (LO $2 \rightarrow 2$ ME), and MadGraph + Pythia8 (LO $2 \rightarrow 2, 3$ ME) are shown. The MadGraph + Pythia8 predictions are obtained using $q_{\text{cut}} = 20$ GeV (default in CMS), as well as the new choice: $HT < 1000$ GeV $\Rightarrow q_{\text{cut}} = 50$ GeV, and $HT > 1000$ GeV $\Rightarrow q_{\text{cut}} = 140$ GeV. The solid band in the ratio plot represents the experimental uncertainty, and the error bars represent the statistical uncertainty of the predictions.
Appendix E. Matching/merging and the inclusive jets transverse momentum cross section
Among the investigations reported in App. A we discussed the measurement of the minimum azimuthal separation between any two of the four highest $p_T$ jets in inclusive 4-jet events ($\Delta \phi_{2j}^{\text{min}}$). In this section I use the results of the LO multi-jet merging studies reported in App. E to explain the large discrepancies observed in App. A between the predictions from MadGraph + Pythia8 and the data, for the $\Delta \phi_{2j}^{\text{min}}$ distribution in inclusive 4-jet events.

We performed the important measurement of $\Delta \phi_{2j}^{\text{min}}$ using data from pp collisions at the center-of-mass energy $\sqrt{s} = 13$ TeV, collected during 2016 with the CMS experiment. The 4-highest $p_T$ jets in the event are required to have a $p_T$ larger than 100 GeV, and to be central in rapidity ($|y| < 2.5$). The measurement was done in bins of the leading jet transverse momentum $p_T^{\text{max}}$, and the resulting distributions were normalized to the 4-jet inclusive cross section integrated over the full $\Delta \phi_{2j}^{\text{min}}$ range, in the corresponding $p_T^{\text{max}}$ bin. A high-$p_T$ 4-jet event recorded by the CMS experiment in 2016 is shown in Fig. F.1.

We observed in App. A that the predictions for the $\Delta \phi_{2j}^{\text{min}}$ distribution in inclusive 4-jet events from MadGraph + Pythia8 (using ME with up to four partons in the final state), differed from the data in up to 40% at high $\Delta \phi_{2j}^{\text{min}}$, for high $p_T^{\text{max}}$ ranges. Moreover, the description of the data, provided by Pythia8, is very similar to the one from MadGraph + Pythia8, even though the Pythia8 generator employs only LO $2 \to 2$ ME while the MadGraph calculation includes the LO $2 \to 3$ and $2 \to 4$ multiplicities in addition to the LO $2 \to 2$. At first, this behaviour was not very well understood, due to the fact that the observable involves four hard jets, meaning that it should be better described by predictions which include $2 \to 4$ ME.

As commented in App. A, the different multiplicities in the MadGraph + Pythia8
Figure F.1: Event display showing a high-$p_T$ 4-jet event system. The event was recorded by the CMS experiment during the 2016 data taking period.

calculation were combined using the MLM merging scheme with a merging scale $q_{\text{cut}} = 20$ GeV (default settings in CMS).

Figure F.2 shows the comparison of the MadGraph + Pythia8 prediction (using the aforementioned settings) to the data for the $\Delta\phi_{2j}^{\text{min}}$ distribution in inclusive 4-jet events, for $700 < p_T^{\text{max}} < 800$ GeV. The fixed-order $2 \to 4$ ME level calculation is also shown in Fig. F.2 in order to show the contribution to the $\Delta\phi_{2j}^{\text{min}}$ distribution from the PS, as well as from the merging with lower multiplicities once the PS is included.

In Fig. F.2 one can observe that the fixed-order 24 ME level calculation provides a good description of the data over the full $\Delta\phi_{2j}^{\text{min}}$ range. However, when the PS is included and the subsequent merging with the lower multiplicities is carried out, the prediction deviates in up to 30% with respect to the ME level calculation. This important result suggests that the calculation which includes the PS might have not been properly done, since the PS should not change the configuration of events with hard, well separated jets when the corresponding ME is available.

Based on the studies reported in App. E it is easy to notice that the problem can be related to the smallness of the merging scale ($q_{\text{cut}} = 20$ GeV) compared to the scale of the process ($\sim p_T^{\text{max}}$). Using observables that are directly sensitive to the merging scale choice (see discussion on the $pT3$ and $d23$ distributions in App. E), I found that $q_{\text{cut}} = 140$ GeV corresponds to a merging scale choice which is valid also for $700 < p_T^{\text{max}} < 800$ GeV.

In Fig. F.3, the prediction of the $\Delta\phi_{2j}^{\text{min}}$ distribution in inclusive 4-jet events from Mad-
Figure F.2: The prediction for the $\Delta\phi_{2j}^{\text{min}}$ distribution in inclusive 4-jet events from MADGRAPH + PYTHIA8, and from standalone PYTHIA8 are compared to data, for $700 < p_T^{\text{max}} < 800$ GeV. The MADGRAPH predictions are based on LO QCD ME with up to four partons in the final state, and the MLM merging scheme is used with the merging scale choice $q_{\text{cut}} = 20$ GeV (default settings in CMS). The solid band in the ratio plot represents the experimental uncertainty, and the error bars represent the statistical uncertainty of the predictions.

Additionally, Fig. F.3 shows the calculation using MADGRAPH + PYTHIA8 with the default value $q_{\text{cut}} = 20$ GeV, as well as the prediction from the standalone PYTHIA8 MC generator. As one can observe in Fig. F.3, the correction from the PS is small when the new merging scale choice $q_{\text{cut}} = 140$ GeV is used. This agrees with the fact that the leading contribution to an observable that is sensitive to four high-$p_T$ jets is provided by the $2 \rightarrow 4$ ME. The results presented in this section also support the conclusions from the LO multi-jet investigations reported in App. E.
Appendix F. Matching/merging and azimuthal correlations in 4-jet events

Figure F.3: The prediction for the $\Delta \phi_{qj}^{\text{min}}$ distribution in inclusive 4-jet events from MadGraph + Pythia8, as well as from the standalone Pythia8 MC generator are compared to data, for $700 < p_T^{\text{max}} < 800$ GeV. The MadGraph predictions are based on LO QCD ME with up to four partons in the final state, and the MLM merging scheme is used with the merging scale choice $q_{\text{cut}} = 20$ GeV (default settings in CMS), as well as the new merging value $q_{\text{cut}} = 140$ GeV. The solid band in the ratio plot represents the experimental uncertainty, and the error bars represent the statistical uncertainty of the predictions.
In Secs. 4.3, 9.2.6 and 9.4.6 we discussed the impact on the \( \Delta \phi_{12} \) distribution of the use of TMDs for the initial state evolution, where Powheg-2j was employed to calculate the ME. In those studies the matching of the ISR generated according to the TMD evolution, to the ME was done by simply choosing the starting scale of the ISR as the scale characterizing the hardness of the real emission (\textsc{calup}), characterized by the relative \( p_T \) of the radiated parton. However, a detailed study of the matching procedure must be carried out in order to prevent regions of phase space with double counting of emissions from ME and ISR, or possible gaps in the phase space which are not filled neither by the ME nor the PS.

In this section, the studies I performed on the matching of NLO ME from Powheg, to ISR generated by Cascade according to the TMD evolution, are presented.

**POWHEG matching of initial state shower generated from TMD evolution to NLO matrix elements**

In order to test the validity of the phenomenological framework, one can start by checking the matching between the ME from Powheg-2j and the shower from Pythia8, before moving on to the matching of the ME to Cascade.

Let us express the radiation phase space in terms of \((\theta, E)\), where \( \theta \) is the angle with respect to the longitudinal axis in the center-of-mass frame of the partonic collision, and \( E \) corresponds to the energy of the parton in the same frame. For simplicity, the partons are
Appendix G. Matching NLO matrix element with initial transverse momentum from TMD evolution to parton shower

Let us as well restrict the scale for the partonic process to be high by requiring the \( p_T \) of the leading parton to be larger than 400 GeV, and the \( p_T \) of the sub-leading parton to be larger than 200 GeV. As commented earlier, the border between the ME and the PS is set by the relative \( p_T \) of the real emission \( \text{Scalup} \). In terms of the angle and energy of the real emission one has: \( \text{Scalup} = E \sin(\theta) \). As a consequence, if \( \text{Scalup} \) is fixed, the border between the ME and PS in the \((\theta, E)\) plane will be given by the curve: \( E(\theta) = \text{Scalup}/\sin(\theta) \).

Following this line of thoughts I restrict \( \text{Scalup} \) to lie within the small range (compared to the scale of the process) \( 100 < \text{Scalup} < 102 \) GeV.

Figure G.1 shows the resulting ISR partons distribution in the \((\theta, E)\) plane, when \( \text{Pythia8} \) is used for the shower, and \( \text{Powheg-2j} \) is used for the ME with \( 100 < \text{Scalup} < 102 \) GeV. The corresponding distribution for the real emission is shown in black color.

![Figure G.1: ISR partons distribution as a function of the angle \( \theta \) and energy \( E \) of the emission in the center-of-mass frame of the partonic collision. \( \text{Pythia8} \) is used for the ISR generation and \( \text{Powheg-2j} \) is used for the ME with \( 100 < \text{Scalup} < 102 \) GeV. The corresponding distribution for the real emission is shown in black color.](image)

One can observe the following important features in Fig. G.1: 1) the distribution in black obeys the relation \( E(\theta) = \text{Scalup}/\sin(\theta) \) where \( 100 < \text{Scalup} < 102 \) GeV; 2) emissions from the \( \text{Pythia8} \) ISR are located below the \( \text{Scalup} \) border (black curve), therefore the real emission represents indeed the border in phase space for the PS. It is important to notice that the number PS emissions fill up to the \( \text{Scalup} \) border, implying that there is no gap or double counting of emissions, therefore the matching was properly done.

One can then calculate the ISR emission distribution as a function of \((\theta, E)\) for the case in which the ISR is generated by \( \text{Cascade} \) according to the TMD evolution. I will use a TMD obtained using the parton branching method (discussed in App. C), for which the evolution scale is required to satisfy the angular ordering condition. The renormalization scale at the branching vertex is taken as the \( p_T \) of the emission. Further details on the TMD
determination, fitting and application can be found in App. C.

The resulting ISR emission distribution as a function of \((\theta, E)\), where CASCADE is used to generate ISR, is shown in Fig. G.2.

![ISR emission distribution](image)

**Figure G.2:** ISR partons distribution as a function of the angle \(\theta\) and energy \(E\) of the emission in the center-of-mass frame of the partonic collision. CASCADE is used for the ISR generation according to the TMD evolution and POWHEG-2J is used for the ME with \(100 < \text{Scalup} < 102\) GeV, and maximum cumulative transverse momentum from the TMD \(k_t\)\(_{\text{max}} =\) Scalup. The corresponding distribution for the real emission is shown in black color.

The value \(k_t\)\(_{\text{max}}\) characterizes the maximum cumulative transverse momentum that the event is allowed to get from the TMD evolution. As indicated in Fig. G.2 \(k_t\)\(_{\text{max}} =\) Scalup, which is the default choice in CASCADE. Similar to Fig. G.1, the black region in Fig. G.2 represents the distribution of the real emission from POWHEG. The important feature which can be noticed in Fig. G.2 is the gap in the distribution, a region which is not filled by the ISR below the Scalup border. This is an indication that, by default, the matching is not properly done in the POWHEG + CASCADE calculation.

One way of overcome this problem is to increase the phase space for ISR radiation by increasing \(k_t\)\(_{\text{max}}\). In Fig. G.3, the ISR emission distribution as a function of \((\theta, E)\) is shown, where the new value \(k_t\)\(_{\text{max}} = \sqrt{2} \text{ Scalup}\) is used as the maximum cumulative transverse momentum that the event is allowed to get from the TMD evolution.

As can be observed in Fig. G.3, the gap has been filled by increasing \(k_t\)\(_{\text{max}}\). One can also notice that there can be events with emissions above the Scalup border. Nevertheless, these events can be easily vetoed in order to prevent possible double counting.
Appendix G. Matching NLO matrix element with initial transverse momentum from TMD evolution to parton shower

Figure G.3: ISR partons distribution as a function of the angle $\theta$ and energy $E$ of the emission in the center-of-mass frame of the partonic collision. CASCADE is used for the ISR generation according to the TMD evolution and POWHEG-2J is used for the ME with $100 < \text{Scalup} < 102$ GeV, and maximum cumulative transverse momentum from the TMD $k_t^{\text{max}} = \sqrt{2} \text{Scalup}$. The corresponding distribution for the real emission is shown in black color.
In this section the resolution curves as well as the purity, stability, acceptance, and background distributions are shown for the inclusive 2-jet observables as a function of $\Delta \phi_{12}$ (Figs. H.1, H.3), as well as for the 3-jet $\Delta \phi_{12}$ distributions as a function of $\Delta \phi_{12}$ (Figs. H.2, H.4), for ten $p_T^{\text{max}}$ regions. The distributions are an important part of the resolution and bin migration studies discussed in Secs. 6.4 and 6.5.
Figure H.1: The difference between the generated and reconstructed $\Delta \phi_{12}$ for the inclusive 2-jet observables, for ten $p_T^{\text{max}}$ ranges.
Figure H.2: The difference between the generated and reconstructed $\Delta \phi_{12}$ for the inclusive 3-jet observables, for ten $p_T^{max}$ ranges.
Figure H.3: Purity, stability, acceptance and background distributions for the inclusive 2-jet observables as a function of $\Delta\phi_{12}$, for ten $p_T^{\text{max}}$ ranges.
Figure H.4: Purity, stability, acceptance and background distributions for the inclusive 3-jet observables as a function of $\Delta \phi_{12}$, for ten $p_T^{max}$ ranges.
APPENDIX

I

RESPONSE MATRICES, FRACTIONAL ERRORS AND CORRELATION MATRICES

In this section the response matrices, fractional errors, as well as the correlation matrices are shown for the inclusive 2-jet observables as a function of $\Delta \phi_{12}$ (Figs. I.1, I.3, I.5 respectively) as well as for the 3-jet $\Delta \phi_{12}$ distributions as a function of $\Delta \phi_{12}$ (Figs. I.2, I.4, I.6 respectively), for ten $p_T^{\text{max}}$ regions. The distributions shown in this section are crucial for the unfolding studies discussed in Sec. 6.8.
Appendix I. Response matrices, fractional errors and correlation matrices

Figure I.1: Response matrices obtained from the official PyTHIA8 (tune CUETP8M1) CMS MC samples for the inclusive 2-jet observables as a function of $\Delta \phi_{12}$.
Figure I.2: Response matrices obtained from the official Pythia8 (tune CUETP8M1) CMS MC samples for the inclusive 3-jet observables as a function of $\Delta \phi_{12}$. 
Appendix I. Response matrices, fractional errors and correlation matrices

Figure I.3: Fractional statistical errors for the unfolded and the measured inclusive 2-jet observables as a function of $\Delta\phi_{12}$. 
Figure I.4: Fractional statistical errors for the unfolded and the measured inclusive 3-jet observables as a function of $\Delta\phi_{12}$. 
Appendix I. Response matrices, fractional errors and correlation matrices

Figure I.5: Correlation matrices (lower half) after unfolding for the inclusive 2-jet observables as a function of $\Delta \phi_{12}$. 
Figure I.6: Correlation matrices (lower half) after unfolding for the inclusive 3-jet observables as a function of $\Delta \phi_{12}$. 

- **Correlation coefficients**
- **Inclusive 3-jet**
- **200 < $p_T^{3\text{max}}$ < 300 GeV**
- **300 < $p_T^{3\text{max}}$ < 400 GeV**
- **400 < $p_T^{3\text{max}}$ < 500 GeV**
- **500 < $p_T^{3\text{max}}$ < 600 GeV**
- **600 < $p_T^{3\text{max}}$ < 700 GeV**
- **700 < $p_T^{3\text{max}}$ < 800 GeV**
- **800 < $p_T^{3\text{max}}$ < 1000 GeV**
- **1000 < $p_T^{3\text{max}}$ < 1200 GeV**
- **$p_T^{3\text{max}} > 1200$ GeV**
- **$p_T^{3\text{max}} > 2000$ GeV**
Appendix I. Response matrices, fractional errors and correlation matrices
In this section, the results from four different tests to the unfolding procedure used to correct the inclusive 3-jet distributions as a function of $\Delta \phi_{12}$ from detector effects, are presented. The unfolding tests corresponding to the inclusive 2-jet distributions are discussed in Sec. 6.8.6.
Appendix J. Tests for the unfolding of the inclusive 3-jet $\Delta \phi_{12}$ distributions

Internal closure test

For the case of the internal closure test, the detector (reconstructed) level predictions obtained from the official PYTHIA8 (tune CUETP8M1) MC samples are unfolded using the same MC, and the result is then compared to the corresponding stable particle level from the simulation. The corresponding ratios are shown in Fig. J.1 for the inclusive 3-jet distributions as a function of $\Delta \phi_{12}$. Similar to the case of the inclusive 2-jet distributions shown in Fig. 6.27, the ratios are consistent with 1.

![Graph showing internal closure test results](image)

Figure J.1: Internal closure test for the unfolding of the inclusive 3-jet distributions as a function of $\Delta \phi_{12}$ using the official PYTHIA8 (tune CUETP8M1) MC samples. The detector level predictions from the MC simulation are unfolded using the same MC sample and compared to the corresponding stable particle level predictions.
Backfolding test

Figure J.2 corresponds to the backfolding test for the inclusive 3-jet distributions differential in $\Delta \phi_{12}$. The unfolded distributions are folded back and then compared to the corresponding distributions at reconstructed level. The official MC Pythia8 (tune CUETP8M1) samples are used. Similar to the case of the inclusive 2-jet distributions shown in Fig. 6.28, the resulting ratios are consistent with a flat behaviour around 1, within the statistical uncertainties.

![Graph showing backfolding test results for inclusive 3-jet distributions](image)

Figure J.2: Backfolding test for the inclusive 3-jet distributions differential in $\Delta \phi_{12}$. The unfolded distributions are folded back and then compared to the distributions at reconstructed level. The official Pythia8 (tune CUETP8M1) MC samples are used.
Appendix J. Tests for the unfolding of the inclusive 3-jet $\Delta\phi_{12}$ distributions

**Bottom line test**

The results corresponding to the bottom line test to the unfolding of the inclusive 3-jet distributions, using the official MC PyTHIA8 tune CUETP8M1 samples, are shown in Tab. J.1. Similar to the case of the inclusive 2-jet distributions shown in Tab. 6.3, the difference between the model and the data at DET level is small and generally larger than the same difference at GEN level.

<table>
<thead>
<tr>
<th>$p_T^{\text{max}}$ [GeV]</th>
<th>$\chi^2$/ndf Unfolding</th>
<th>$\chi^2$/ndf Folded back</th>
</tr>
</thead>
<tbody>
<tr>
<td>200-300</td>
<td>19.343</td>
<td>21.504</td>
</tr>
<tr>
<td>300-400</td>
<td>18.938</td>
<td>18.923</td>
</tr>
<tr>
<td>400-500</td>
<td>20.611</td>
<td>23.750</td>
</tr>
<tr>
<td>500-600</td>
<td>19.020</td>
<td>24.184</td>
</tr>
<tr>
<td>600-700</td>
<td>98.506</td>
<td>128.483</td>
</tr>
<tr>
<td>700-800</td>
<td>43.981</td>
<td>61.904</td>
</tr>
<tr>
<td>800-1000</td>
<td>28.206</td>
<td>42.015</td>
</tr>
<tr>
<td>1000-1200</td>
<td>8.211</td>
<td>11.674</td>
</tr>
<tr>
<td>$&gt;1200$</td>
<td>4.620</td>
<td>6.438</td>
</tr>
<tr>
<td>$&gt;2000$</td>
<td>1.496</td>
<td>1.707</td>
</tr>
</tbody>
</table>

Table J.1: Bottom line test for the unfolding of the inclusive 3-jet observables. The $\chi^2$/ndf of the unfolded data compared to the PyTHIA8 GEN level (middle column), and $\chi^2$/ndf of the data at reconstructed level compared to the PyTHIA8 DET level (right column) are shown. The official MC PyTHIA8 (tune CUETP8M1) MC simulation is used.
External closure test

Figure J.3 shows the external closure test corresponding to the inclusive 3-jet distributions as a function of $\Delta\phi_{12}$. The DET level predictions from the official MC PYTHIA8 (tune CUETP8M1) MC simulation are unfolded using another, independent MC (official MADGRAPH+PYTHIA8 tune CUETP8M1). The result is then compared to the stable particle level (GEN) from the MC PYTHIA8 (tune CUETP8M1) simulation. Similar to the results for the inclusive 2-jet distributions shown in Fig. 6.29, the agreement observed in Fig. J.3 (within the unfolding systematic uncertainties) for the inclusive 3-jet observables further supports the consistency and validity of the unfolding procedure.

![Diagram showing external closure test for inclusive 3-jet distributions as a function of $\Delta\phi_{12}$]
Appendix J. Tests for the unfolding of the inclusive 3-jet $\Delta\phi_{12}$ distributions


278


282


[110] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”. 40, 123


283


284


[161] Private communication with JetMET CMS group members, 2017. 110


