Jet Substructure Tools to Identify Hadronization Timescales

Nuno Olavo Gonçalves Mendes Madureira

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisor(s): Prof. Liliana Marisa Cunha Apolinário
Prof. Raghav Kunnawalkam Elayavalli

Examination Committee

Chairperson: Prof. Maria Teresa Haderer de la Peña Stadler
Supervisor: Prof. Raghav Kunnawalkam Elayavalli
Member of the Committee: Prof. José Guilherme Teixeira de Almeida Milhano

October 2022

Para os meus amados pais,
Agostinho e Denise

Acknowledgments

First and foremost, a special thanks to my supervisors Liliana and Raghav for their guidance in my initiation into particle physics research and for the unwavering patience with which they met my numerous questions along the way. I also have to thank Instituto Superior Técnico for the 5 years of learning that lead up to this thesis and Fundação para a Ciência e a Tecnologia (FCT) for the 6 months of financial support for the EXPL/FIS-PAR/0905/2021 project.

Outside of the academy, I owe the resilience and stability I felt throughout the writing of this thesis to a wonderful family and an incredible group of friends. Firstly, I have to thank my parents for always encouraging me to pursue an education and my sister Rita for making me laugh even at moments of greater stress. I also need to thank my hometown friends: Amália, my biggest cheerleader and the person I owe the most to, Beatriz, always saying the unexpected and always available for a hug, Borlido, with whom any topic provides hours of conversation, Joana Faria, with whom I can share everything and always disagree on music and cinema, Kika, for her never-ending jokes and charming presence, Joana Lima, always ready to just be goofy with me, Patrícia, the gamemaster and adventurer, and Rafaela, for all the times she made me scream-laugh and for all the hours spent on the phone. I would be remiss if I did not thank Pedro and Inês for making my life in Lisbon actually bearable.

Resumo

Quantificar as escalas da transição entre os regimes perturbativo (pQCD) e não-perturbativo (npQCD) da Cromodinâmica Quântica é um problema de raiz da Física de Partículas. Enquanto que pQCD aplica-se aos partões (quarks e gluões), partículas finais (detetáveis) enquadram-se no regime npQCD. Estas designam-se por hadrões, partículas compostas produzidas via hadronização. Para adquirir conhecimento relativo ao processo de hadronização, esta tese propõe o uso de jets como ferramenta de análise, que são estruturas de partículas finais às quais sequências de clusterings são atribuídas como proxies da história evolutiva das partículas.

Usando jets de diferentes regiões cinemáticas e colisões de diferentes energias de centro de massa como instrumentos de seleção da espécie partónica predominante na iniciação dos jets, são realizados estudos de subestrutura que revelam que gluões produzidos em colisões energéticas tendem a iniciar jets mais largos, simétricos e céleres que os quarks. Este documento contém estudos adicionais focados na fração de momento groomed, que mede a simetria de uma emissão em termos de momento transversal, e no tempo de formação, um proxy para as escalas temporais das emissões. Estes mostram que a seleção de jets baseada em tempos de formação permite a identificação da etapa de declustering na qual as duas partículas carregas mais energéticas se desacoplam. Também revelam que este processo marca claramente na sequência de clusterings a transição de propriedades típicas de pQCD para propriedades atípicas deste regime.

Palavras-chave: Jets, Chuveiro Partónico, Hadronização, Tempo de Formação, Resolved Splitting

Abstract

Quantifying the transitory scales between the perturbative (pQCD) and non-perturbative (npQCD) regimes of Quantum Chromodynamics is a deep-rooted problem in Particle Physics. While pQCD applies to partons (quarks and gluons), final-state (detectable) particles lie within the npQCD regime. The latter are called hadrons, composite particles produced via hadronization. To get insight into the hadronization process, this thesis proposes to use jets as probing tools, which are structures of final-state particles to which clustering sequences are assigned as proxies for the particle-evolution history.

Using different jet kinematics and centre-of-mass collision energies as a means of selecting the predominant jet-initiating parton-species, jet substructure studies reveal that hard scattered gluons tend to initiate wider, more symmetric and earlier jets than quarks. This document contains additional studies focused on the groomed momentum fraction, measuring an emission's symmetry in transverse momentum space, and on formation time, a proxy for the emission timescales. They show that selecting jets on formation time allows the identification of the clustering step where the 2 leading charged particles decouple. They also reveal that this process (which resolves the 2 leading charged particles) clearly marks the transition in the clustering tree from pQCD-like to non-pQCD-like features.

Keywords: Jets, Parton Shower, Hadronization, Formation time, Resolved Splitting

Contents

List of Tables

List of Figures

Nomenclature

BNL	Brookhaven National Laboratory
CERN	European Organization for Nuclear Research
CM	Center-of-mass
DIS	Deep Inelastic Scattering
FCC	Future Circular Collider
FF	Fragmentation Function
FSR	Final-State Radiation
HERA	Hadron-Electron Ring Accelerator
HIC	Heavy-Ion Collision
ISR	Initial-State Radiation
LHC	Large Hadron Collider
LO	Leading Order
PDF	Parton Distribution Function
pp	Proton-proton
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
RHIC	Relativistic Heavy-Ion Collider
SM	Standard Model
1SD	1^{st} soft-drop
LCP	Leading Charged Particles
RSD	Resolved soft-drop

Chapter 1

Introduction

First and foremost, it is important to lay out the scientific prolificacy of high-energy collisions in the particle physics field and to state the relevance of the work performed in this thesis to current research. A brief introduction to all the major related concepts can be found in the next Sections, so to understand the primary objective and the layout of the thesis.

1.1 Motivation and Relevance

In spite of the unmatched predictive power of the Standard Model of particle physics, it is increasingly finding challenges to its applicability to extreme environments, like some of the ones produced in high-energy beam collision experiments. The most prominent of these experiments in the past decades study proton-proton (pp) and heavy-ion collisions performed at the Relativistic Heavy-Ion Collider (RHIC) at the Brookhaven National Laboratory (BNL) and Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN).

The properties of the outgoing particles detected in these colliders still cannot be fully derived from first principles alone, meaning there is much to gain from continued research and investment in how particles evolve out of hard scattering processes. This is why we see a long-lasting push from the scientific community to develop particle collider projects of even greater scales than the ones previously enumerated. In fact, as recently as 2020, news came out about the CERN Council's unanimous endorsement of a new 100-kilometre particle collider at CERN [1], to be called the Future Circular Collider (FCC).

1.2 Field Overview

The Standard Model (SM) of particle physics is a theory that describes the strong and the electroweak forces, leaving only outside of its scope the fourth fundamental interaction - Gravity. Quantum Chromodynamics (QCD) is the quantum field theory of the strong interaction, describing the color sector of the SM [2]. Quarks and gluons, collectively called partons, are the building blocks of QCD, which aims to describe the interactions of the color-charged partons and their composite color-neutral states -

hadrons and nuclei [3].

Ultra-relativistic hadron collisions can release partons from their hadronic confinement, like in the pp collisions performed at RHIC and the LHC. After deconfinement, the subsequent multi-particle states evolve along drastically different physical scales, from the high-energies of the immediate aftermath of the collision to the low momentum scales of the final hadronized particles.

At high-momentum scales, the evolution from one multi-parton state to another is governed by the splitting functions provided by QCD, which yield the probability of a parton of a given flavour and with a given energy to branch into two other partons (presented in detail in Section 2.2.2). This splitting process is therefore called parton branching. Highly energetic partons usually lead to multiple iterations of the branching process, creating a cascade of partonic decays - the parton shower.

However, partonic degrees of freedom cannot be measured directly in experiments. Only color-neutral bound states of quarks and gluons - the hadrons - can be and have been detected experimentally. One can approximately think of hadrons, according to the MIT Bag Model [4], as confined “bags” of quarks and gluons immersed in QCD vacuum. Inside these “bags”, alongside with the hadron’s valence quarks, one finds an endless well of complexity, with gluonic fields and quark-antiquark pairs constantly popping in and out of existence. The process by which independent partonic degrees of freedom get coupled, resulting in confinement and hadron production, is what is designated as hadronization [5]. Therefore, the shower continues evolving until the energy-scales become low enough for hadronization to take place, marking the transition of QCD from its perturbative regime (pQCD) to its non-perturbative one (npQCD).

Figure 1.2.1 shows a schematic representation of the different phenomena involved in a pp collision, with time flowing from the bottom to the top of the scheme. In this example, the colliding protons, represented in black circles at the bottom, interact via a hard scattering between a constituent parton from each of them, as represented in the red Feynman diagram. This hard scattering produces two outgoing highly energetic partons, birthing a parton shower through successive parton branchings, represented by the black 3-parton Feynman vertices. The event culminates with the hadronization process, in yellow, where the final-state hadrons are produced.

Hadronization is the main focus of this thesis. Since it takes place at low momentum scales, smaller than $(\Lambda_{QCD})^2 = (332\ \text{MeV})^2$ [6], the physics of hadronization cannot be derived from perturbative calculations. Perturbation theory relies on a convergence of probability calculations when taking into account processes of increasingly higher order, producing, although sometimes arduously, theoretical results of significant precision. This convergence is assured when the QCD coupling constant, α_s, is small, which happens for large momentum scales, as is the case for most of the phenomena below the yellow regions of Figure 1.2.1. However, α_s becomes large at low momenta (large distances), invalidating a perturbative expansion. As such, quarks and gluons are asymptotically free for small distances, but the strength of their interaction grows as the distance increases [3]. Therefore, for these large distances, confinement kicks in at the same time as perturbation theory can no longer be used in calculations. The invalidity of perturbation theory for the typical scales of the hadronization process contributes to the current lack of understanding of this phenomenon.

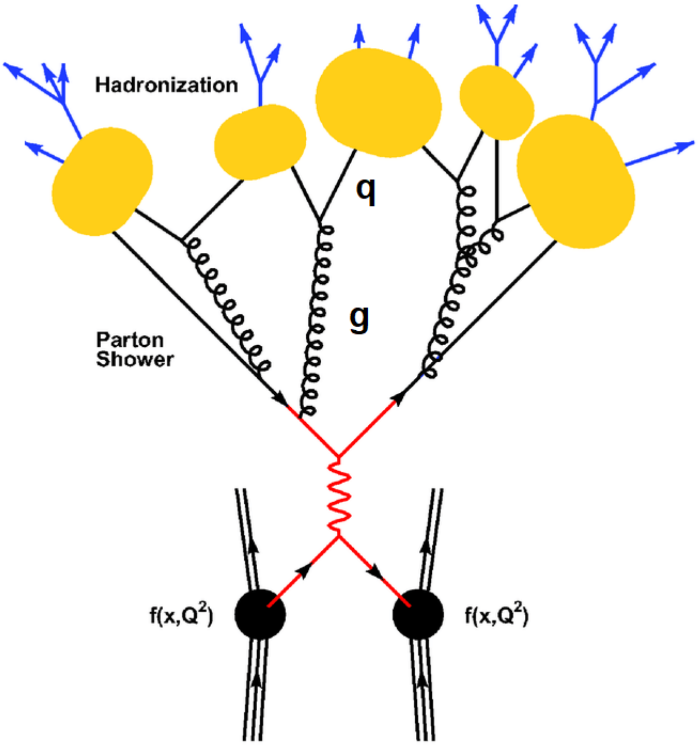


Figure 1.2.1: Schematic representation of all evolutionary stages of an ultra-relativistic proton-proton collision, from the QCD coupling of two partons, one from each proton, to the subsequent parton shower and culminating in colorless particles after the hadronization process takes place.

Current experiments span over a wide range of center of mass energies $\sqrt{s}$, covering both pp and relativistic heavy-ion collisions (HICs). Two beams of protons or heavy-ions are accelerated in opposite directions inside the collider tunnels until the combined center of mass (CM) energy of the beams reaches $\sqrt{s}$. The typical collision energies at RHIC are $\sqrt{s} = 0.2$ TeV, at which both pp and AuAu collisions are performed, while much greater ones are produced at the LHC, of around $\sqrt{s} = 5$ TeV for PbPb collisions, reaching the $\sqrt{s} = 14$ TeV CM energies for pp.

The final-state particles produced in these collisions eventually reach the detectors, where their energies are measured. The particle aftermath can then be arranged into groups called jets, according to sets of rules carefully described in Section 2.3. Jets are the objects used in this thesis to probe the evolutionary history of hard scattering events and to get insight into hadronization. Each jet contains information about a collection of final-state particles that allows the recreation of the branching steps originating each of the intermediate multi-particle states. The end result is the so-called jet clustering tree.

HICs consist of collisions between two heavy nuclei, accelerated at ultra-relativistic velocities. Such velocities contract the nuclei into highly dense disks of partonic matter due to Lorentz transformations. A HIC simulation can be seen in Figure 1.2.2, with the Lorentz-contracted ionic disks represented in red and the particle aftermath of the collision represented by the multi-colour dots between them (colour scheme established in the Figure). When the 2 disks meet, their partonic makeup experience QCD interactions, scattering partons with varying transverse momentum. However, the majority of these scatterings are, in fact, soft processes, while hard scatterings will take place very rarely, around $1{:}10^6$.

The high number of soft partons that concentrate in the middle of the receding nuclei, along with the associated color fields, constitute a highly complex and short-lived state of matter called quark-gluon plasma (QGP). This new medium significantly affects the particle aftermath of such a collision, since it will be traversed by the earliest particles produced in HICs, inevitably introducing modifications to them.

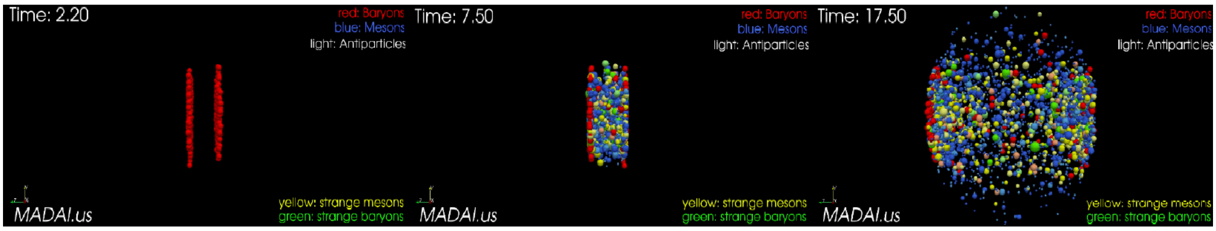


Figure 1.2.2: Timestamps of a heavy-ion collision simulation; the timestamp on the left shows the nuclei, right before the collision, as two Lorenz-contracted red disks, where the red particles represent baryons; the middle timestamp shows the immediate hadronic aftermath of the collision and the right timestamp shows the same at a later time; the blue, white, yellow and green particles represent, respectively, mesons, antiparticles, strange mesons and strange baryons.

The numerous experiments performed at RHIC and the LHC in the past few decades provide a copious data set of the particle aftermath of 2 particles colliding. In particle physics, this is called an event. In order to develop and guide new research in this field, many physicists are dedicated to develop numerical algorithms that attempt to generate simulated events. Such a framework of algorithms is called an event generator [7]. The goal is to make these simulated events as faithful to experiment as possible in order to take event generators as predictive tools for future measurements. Due to the stochastic nature of quantum phenomena, the profile of the outgoing particles produced in a given event falls on probability distributions, which the event generator has to replicate. It can extract them from a combination of theoretical and phenomenological models, tuned to previous experimental measurements.

The event generator used in this work is PYTHIA 8.306 [7]. This program is a standard tool for the generation of high-energy collision events. It applies a C++ library of parameterized physics in order to re-produce the evolution from the collision of 2 incoming high-energy protons to the set of final-state particles. PYTHIA's evolution algorithms are formulated in momentum space, leading the transverse momentum to be the most common measure of the evolution scale. This Monte Carlo event generator faithfully mimics the stochastic nature of the events by using pseudo-random numbers and extrapolating from probability distributions. It was built and expanded on the idea that collision-type events have many transversal properties, independent of beam type, fruit of the factorization between perturbative and non-perturbative physics.

1.3 Objectives and Deliverables

The goal of this project is to improve on the current estimates for the hadronization timescales, a measure of the time it takes for partons to become final-state hadrons. It focuses specifically on vacuum (pp) ultra-relativistic collisions, taking the QCD jets as probes. The clustering trees extracted from these jets reveal their spatio-temporal substructure through the calculation of a multitude of jet substructure variables. The most important ones in the context of this work are the momentum fraction (transverse momentum share between the daugther-partons), the opening angle and the formation time of selected splittings from the clustering tree. These are used to study vacuum QCD dynamics and to detect QCD-regime transitions. The observations will be drawn from a collection of jets and, therefore, an ensemble of splittings. This thesis also aims to identify the experimental feasibility of the developed analysis for RHIC and LHC energies.

1.4 Thesis Outline

Chapter 2 gives a concise introduction to the theoretical foundations underlying this work. The Chapter begins with the QCD fundamentals, crucial to establish the framework of this quantum field theory (QFT), and proceeds to lay out the physics behind hard scattering events. It concludes with a very detailed state-of-the-art review of jet techniques, many of whom employed in this thesis. Then, Chapter 3 sets out the ubiquitous simulation features imposed to the pp events through PYTHIA 8.306 and how jet analysis was computationally implemented. It follows up with the first results regarding jet substructure at tendentiously perturbative scales. Afterwards, Chapter 4 presents how the jet substructure variables compare between perturbative and manifestly non-perturbative QCD regimes, aiming to flag the point where the transition takes place. Lastly, Chapter 5 synthesises the main conclusions drawn from the results shown in Chapters 3 and 4 and reflects on future research that might follow these findings.

Chapter 2

From Perturbative to Non-Perturbative QCD

The following Sections will go through, in great detail, the theoretical bedrocks and the state of the art analysis techniques required to understand the work performed in this thesis. The reader is first presented to a compact overview of the QCD principles that govern partonic matter, focusing on the QCD Lagrangean, the scale-dependence of QCD coupling and factorization properties. Afterwards, one introduces the relevant phenomena that come into play in ultra-relativistic hard scatterings, from the break of confinement to parton shower dynamics and the modeling of hadronization. It is followed up by an algorithmic description of the fundamental object used in the analysis of high-energy particle and nuclear physics: the jet. It will go through the techniques of jet identification and how to recover the jet history from the properties of the final-state particles. The reader will find an overview of the appropriate jet substructure tools, namely the soft-grooming performed by the soft-drop algorithm, the ordering in formation time and varying clustering algorithms.

2.1 QCD Basics

As was mentioned in Chapter 1, the QFT that describes parton-dynamics is QCD. The following Sections go through the QCD Lagrangean and the implications from its symmetries, the properties of the running coupling at different energy scales and the factorization observed between QCD regimes.

2.1.1 Lagrangean

As is true for any given QFT, the fields' behaviour arises from a Lagrangian that obeys the intrinsic symmetries of the theory, each associated with a field quantum number. Fermionic quark fields in particular are invariant under transformations of the local gauge group of SU$(3)_C$, introducing a new quantum number - the color -, which translates into a charge. This enforces quarks to come in 3 different colors and their interactions to be mediated through 8 differently colored gauge bosons, designated as gluons.

Therefore, the QCD (classical) Lagrangean that encodes all of this physics between partonic fields, $\psi_{\alpha,i,f}(x)$ and $A^{\mu}_{a}(x)$ for quarks and gluons, respectively ($i \in \{1,2,3\}$ and $a \in \{1,...,8\}$ are the color indices, $f \in \{$up, down, strange, charm, bottom, top$\}$ the quark-flavour index and α, μ the Dirac spinor indices), is given by

$$\mathscr{L}_{classical} = \bar{\psi}\,(i\not{D} - M)\,\psi - \frac{1}{4} F^{a}_{\mu\upsilon} F^{\mu\upsilon}_{a} \quad , \tag{2.1}$$

where Einstein's notation applies for repeating indices and all indices are omitted for the quark and antiquark fields, ψ and $\bar{\psi}$. Also, M is the diagonal quark-mass matrix. As for the covariant derivative and the field tensor, they are explicitly given by

$$D_{\mu} = \partial_{\mu} - ig\, A^{a}_{\mu}\, \tau_{a} \quad , \tag{2.2}$$

$$F^{a}_{\mu\upsilon} = \partial_{\mu} A^{a}_{\upsilon} - \partial_{\upsilon} A^{a}_{\mu} + g\, f^{abc}\, A^{b}_{\mu}\, A^{c}_{\upsilon} \quad , \tag{2.3}$$

respectively, where g is the QCD coupling constant and f^{abc} the structure constants of the SU(3) Lie group [8]. The operators τ_a are the 8 SU(3) generators in the fundamental representation, given by

$$\tau_{a} = \lambda_{a}/2 \quad , \tag{2.4}$$

with λ_a being the Gell-Mann matrices. These generators are non-commutative

$$[\tau_{a}, \tau_{b}] = i f^{abc} \tau_{c} \qquad \text{for } \; a, b, c = 1, ..., 8 \quad , \tag{2.5}$$

meaning QCD is a non-Abelian theory [9].

Plugging the field tensor from Equation (2.3) back into the Lagrangean in Equation (2.1) leads to the emergence of gluon self-interactions in the form of cubic and quartic vertices. These are represented in the B and C vertices from Figure 2.1.1. Furthermore, in the first term of the Lagrangean, the gluon field from the covariant derivative in Equation (2.2) combines with the quark fields to form a quark-antiquark-gluon vertex [10], shown as vertex A in Figure 2.1.1. The ability of QCD's gauge boson to carry charge and interact with itself constitutes its main difference with respect to Quantum Electrodynamics (QED).

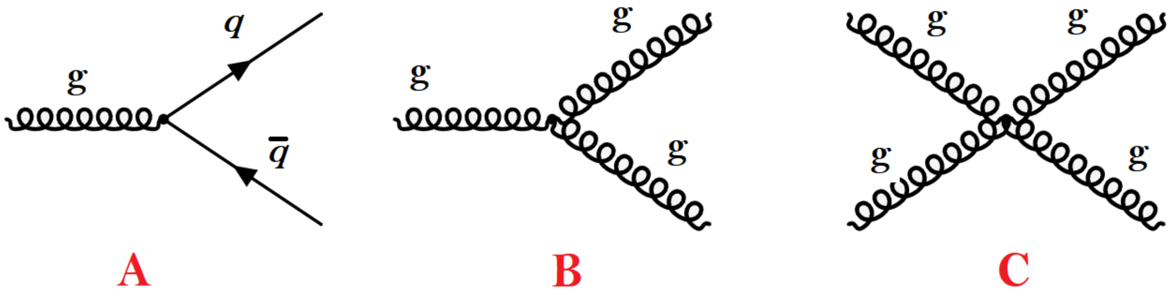


Figure 2.1.1: Tree-level QCD vertices, with A, B and C representing the quark-antiquark-gluon vertex, the cubic gluon vertex and the quartic gluon vertex, respectively.

2.1.2 Running Coupling

A coupling constant is a measure of the strength of a given interaction, at a given scale Q, and in it lies the cause for such stark differences between QED and QCD. On the one hand, the QED coupling constant $\alpha_{QED}(Q)$ is very small for a wide range of energy scales, setting electrodynamics to be a weak interaction for the vast majority of observable physics. On the other hand, α_{QED} changes very slowly with respect to Q in this wide range of energy scales, assuming a value of approximately

$$\alpha_{QED} = \frac{e^2}{4\pi}\frac{1}{\hbar\, c\, \epsilon_0} \approx \frac{1}{137} \quad , \tag{2.6}$$

with e the charge of the electron, $\hbar$ the reduced Planck's constant, c the speed of light in vacuum and ϵ_0 the vacuum permittivity.

This behaviour of the QED's coupling constant allows for perturbation theory to be used in the calculations of QED correlation functions, since the Feynman diagrams of increasingly higher loop orders become smaller and smaller contributions to the end result.

Albeit a renormalizable theory, QCD's coupling constant $\alpha_{QCD}(Q)$ is too large in a wide range of Q scales for perturbative calculations to be performed with any degree of success. These scales include the hadronic ones, which leaves the problem of confinement out of the reach of perturbation theory. Furthermore, it experiences great variations with respect to the energy scale. Just for comparison, from the scale of $Q^2 = (0.1\text{TeV})^2$ to $Q^2 = (1\text{TeV})^2$, the running of $\alpha_{QED}(Q)$ only results in a $\sim 0.4\%$ increase to the value of the coupling constant [11], while it sees the $\alpha_{QCD}(Q)$ decrease by $\sim 25\%$, as can be seen on the left panel of Figure 2.1.2. It shows a comparison between theoretical predictions and experimental measurements of the running α_{QCD}.

Even when perturbation theory applies, the QCD coupling is not as small as α_{QED}, meaning higher order calculations have to be performed to achieve the same level of precision. For instance, in the $Q^2 = (0.1\text{TeV})^2$ to $Q^2 = (1\text{TeV})^2$ energy range, the left panel from Figure 2.1.2 shows α_{QCD} to be 16 to 12 times bigger than α_{QED}, going from $\alpha_{QCD}(Q = 0.1\text{TeV}) \approx 0.12$ to $\alpha_{QCD}(Q = 1\text{TeV}) \approx 0.09$.

The perturbative running coupling of QCD, calculated at one-loop, is given by

$$\alpha_{QCD}\left(Q^2\right) = \frac{1}{\frac{\beta_0}{4\pi}\,\ln\left(Q^2/\Lambda^2_{QCD}\right)} \quad , \tag{2.7}$$

where $\beta_0 = 11 - \frac{2}{3}N_f$, with N_f the number of quark flavours active at the scale Q^2 and Λ_{QCD} the Landau pole of QCD, at which the coupling constant diverges [12]. For $N_f = 3$, the average value from the Particle Data Group for the Landau pole is $\Lambda_{QCD} = (332 \pm 17)$ MeV [6].

The expression for α_{QCD} reveals that for large momenta or, equivalently, for small distances, the running coupling is small and, therefore, partons are asymptotically free particles (pQCD). However, for small momenta (large distances), partons become tightly bound in highly non-perturbative processes (npQCD). This is especially important for the work performed in this thesis, since one is looking at a variable-scale problem. In a hard scattering, the partons evolve from large momentum scales to low momenta final-state particles, through the process of hadronization. Perturbative calculations can therefore be applied in the earlier stages of the parton shower, but they stop being valid at a later point that is yet to be precisely determined, where calculations require non-perturbative methods such as lattice QCD and functional methos.

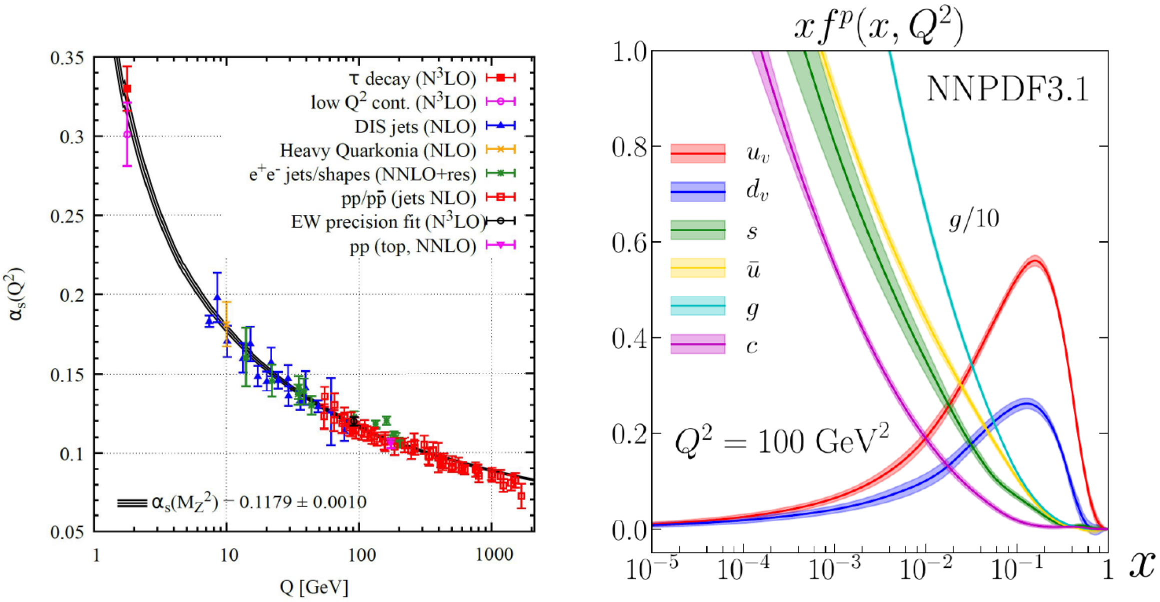


Figure 2.1.2: **Left**: aggregate of measurements of QCD's running coupling, here denoted by α_s, from a wide range of experiments, alongside the theoretical function constrained by $\alpha_s\left(M_Z^2\right) = 0.1179 \pm 0.0010$ at the scale of the Z boson (taken from [10]); **Right**: parton distribution functions (PDFs) for the proton taken from the NNPDF3.1 (NNLO) analyses, where the uncertainty bands correspond to Monte Carlo 68% confidence levels (taken from [13]).

2.1.3 Factorization

Luckily, ultra-relativistic hadron collisions have their high-momenta regime, where perturbation theory can be applied to describe the partonic interactions, factorized from their low-momenta regime, encoding all the complexity of hadronic physics. Therefore, instead of a convoluted problem, one has a neatly separated one. For a pp collision which produces at least one final-state hadron, the cross section is given by

$$d\sigma_{p+p'\to h+Y} = \sum_{a,b,c} \int dx_a \int dx_b \; f_{a/p}\left(x_a, Q^2\right) f_{b/p'}\left(x_b, Q^2\right) \; d\sigma_{a+b\to c}\left(x_a, x_b, Q^2\right) \; D_{h/c}\left(x_h, Q^2\right) \quad , \tag{2.8}$$

where the indices a, b, c run through the partonic species. The parton $a(b)$ has a momentum fraction $x_{a(b)}$ of the incoming proton $p(p')$. The hadron h has a momentum fraction x_h of the hard scattered parton c.

The pQCD evolution is, therefore, inscribed in the parton-level cross section $d\sigma_{a+b\to c}$, which computes the probability of partons a and b interacting to produce parton c (amongst other byproducts). It encodes the so-called Final-State Radiation (FSR), describing the evolution from each multi-parton state to next one. Meanwhile, npQCD evolution is encoded, on the one hand, in the so-called Fragmentation Function (FF) $D_{h/c}$, which computes through hadronization models the probability of parton c evolving to an inclusive state which includes, amongst other particles, the specific hadron h. On the other hand, npQCD is also incoded by the Parton Distribution Functions (PDFs) $f_{a(b)/p(p')}\left(x_{a(b)},\mu^2\right)$, which describe the so-called Initial State Radiation (ISR) for μ^2 scales up to the energy scale of the collision Q^2. PDFs give the probability distributions of quark and gluon momenta inside a given hadron. In the example from Equation (2.8), they compute the probability of finding parton flavour $a(b)$ inside proton $p(p')$ with a fraction $x_{a(b)}$ of the proton's total 4-momentum at scale of μ^2. The proton PDFs as functions of the momentum fraction x for the different partonic species, at the fixed scale of $Q^2 = (100\text{GeV})^2$ and at next-to-next-to-leading order (NNLO), are shown on the right panel of Figure 2.1.2.

While FFs encode the npQCD physics of hadronization, PDFs encode pre-collision npQCD evolution of the proton partonic makeup. Factorization allows the separation of both of them from the pQCD physics in $d\sigma_{a+b\to c}$. Therefore, the PDFs evolution up to the hard scattering scale Q^2 and the FFs evolution from Q^2 up to the hadronization scale Λ^2 is completely separated from the partonic treatment of the collision. Although both the FFs and the PDFs are non-perturbative objects, their evolution with μ^2 can be described by pQCD. However, the full objects are dependent on external parameters extracted from experimental measurements, such as the ones performed for high-precision deep inelastic scatterings (DIS) in the Hadron-Electron Ring Accelerator (HERA) [14], amongst others.

2.2 Collision Physics

Hadron and heavy-ion collisions are used to gain insight at the partonic level of matter. They typically have high enough energies to eject so-called hard or highly-penetrating partons from hadronic confinement, usually characterized by large momentum transfers Q^2 (measure of the 4-momentum transfer squared) or large transverse momenta p_T (with respect to the beamline) [15]. The concept of "hardness" is used as a qualitative scale of particle energy. In order to dive into the complexities of the collision dynamics, the problem has first to be partitioned into the different scales it spans over, allowed by QCD factorization. Therefore, the hard scattering process responsible for the eventual "deconfinment" of partons is divided into the mechanics of the parton shower birthed by these partons and into the hadronization process, when energy scales decrease to a point where partons can no longer benefit from asymptotical freedom.

2.2.1 Hard-Scattering

Let's consider an elementary $a+b \to f_n$ process, where a, b are the 2 incoming partons that originate, through a hard scattering, the n-particle final-state f_n. The elementary differential cross section of this process in the high-energy limit approximation is given by

$$\frac{d\hat{\sigma}}{d\mathbf{\Phi}_n} \approx \frac{\overline{|\mathcal{M}|^2}}{2\hat{s}} \quad , \tag{2.9}$$

where $\hat{s} = (p_a + p_b)^2$ is the (squared) invariant CM energy of the $a-b$ elementary partonic system, not to be confused with the (squared) CM energy of the pp collision s. The QCD amplitude $\mathcal{M}$ is a matrix element that holds the dynamics of the evolution from the incoming partons to the outgoing final-state hadrons calculated at leading-order (LO). Measurements are not sensitive to spin or color configurations of a and b. Therefore, the actual physics of the collision is encoded in the mean square amplitude $\overline{|\mathcal{M}|^2}$, where the product of the amplitude by its complex conjugate is averaged/summed over all spins and colors of the incoming/outgoing particles, respectively. The high-energy limit approximation ($m_{parton} << \sqrt{\hat{s}}$) assumes vanishing parton masses vanish, valid for RHIC and LHC high-energy collisions.

As for the Lorenz-invariant phase-space density $d\mathbf{\Phi}_n$ of the multi-particle final-state f_n in Equation 2.9, it is expressed by

$$d\mathbf{\Phi}_n\,(P;p_1,p_2,...,p_n) = (2\pi)^4\;\delta^{(4)}\left(P-\sum_{i=1}^{n}p_i\right)\prod_{i=1}^{n}\frac{d^3p_i}{(2\pi)^3\,2p_i^0}\quad, \tag{2.10}$$

where P is the total 4-momentum (usually considered in the rest frame and only then boosted into the frame of interest) and p_i the individual 4-momenta of the $i = 1,...,n$ outgoing particles, with the δ-function guaranteeing energy-momentum conservation.

However, due to the impossibility of isolating partons a and b, the colliding beam particles are actually protons. Calling the 2 protons A and B ($s = (p_A + p_B)^2$), one needs to factor in the proton PDFs $f_{a/A}$ and $f_{b/B}$ at the hard scattering scale Q^2 to account for the probability of finding a and b with a x_a and x_b fraction of the proton momentum, as can be seen in Equation (2.8). The collision can only be dealt with as a $a-b$ partonic scattering due to the factorization properties of QCD, explored in Section 2.1.3, effectively separating PDF evolution from the cross section calculations.

2.2.2 Parton Shower

Instead of explicitly including higher order corrections to the LO matrix elements $\mathcal{M}$ (to the detriment of computational efficiency), the probabilistic picture derived in this Section from quark and gluon radiation is used to build a parton shower, which mimics (although not as accurately) the higher order terms that the $\mathcal{M}$ should naturally contain.

Therefore, the parton shower is an attempt to codify the higher order physics of the immediate aftermath of a hard process by creating a partonic-emission path across the full range of the pQCD regime. This succession of emissions is built recursively, starting from the large scale of the hard process Q^2 and evolving towards increasingly softer and collinear resolution scales until it reaches a lower cutoff at around the hadronization scale Λ^2. Parton shower evolution requires a metric for the ordering of parton emissions - an ordering variable -, typically set to be physical kinematic observables such as the transverse momentum, the virtuality ("mass") of the emitting particle or the emission angle.

As explained in Section 2.1.3, QCD factorization allows for the isolated treatment of the parton shower with respect to hadronization and also the factorization of the parton shower between ISR and FSR contributions. Each branching step increases the total number of partons by one and the heuristic nature of each emission leads to a large variability of final-states. This means the multi-parton final-state is the result of a succession of $1 \rightarrow 2$ splittings, obeying the QCD Feynman rules.

Looking first at the FSR dynamics, the differential probability of a parton a to undergo a splitting at scale μ^2 can be expressed by

$$d\mathcal{P}_a\left(z,\mu^2\right) = \frac{d\mu^2}{\mu^2}\,\frac{\alpha_s\left(\mu^2\right)}{2\pi}\sum_{b,c} P_{a\to bc}(z)\;dz \quad , \tag{2.11}$$

where b and c are the daughter-partons (which iterate over all parton-species) and z the fraction of the 4-momentum from the mother-parton carried by its softest daughter.

Also, $P_{a\to bc}$ are the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting kernels [16]. The possible $1 \to 2$ splittings in the parton shower come from the QCD Lagrangian in Equation (2.1). They are the splitting of a quark into a quark and a gluon ($q \to q + g$) and the gluon splittings, which can produce a gluon pair ($g \to g + g$) or a quark-antiquark pair ($g \to q + \bar{q}$), represented in Figure 2.1.1. The DGLAP (un-regularized) splitting functions calculated at LO are given by

$$P_{q\to qg} = \frac{4}{3}\frac{1+z^2}{1-z} \quad , \quad P_{g\to q\bar{q}} = \frac{1}{2}\left(z^2 + (1-z)^2\right) \quad \text{and} \quad P_{g\to gg} = 3\frac{(1-z(1-z))^2}{z(1-z)} \quad . \tag{2.12}$$

Since gluons are massless, 3-body processes like $g \to g + g$ and $g \to q + \bar{q}$ cannot conserve energy-momentum in isolation. The same happens for the quark splitting $q \to q + g$ in the high-energy limit. Therefore, shower radiation is considered to be produced by a colour dipole instead of by a point-like particle. In this splitting framework, the branching parton, called the emitter, has a unique anticolour partner - the recoiler - guaranteeing 4-momentum conservation. Parton branchings are assumed to occur isotropically in azimuthal angle ϕ in the rest frame of the respective dipole.

Integrating Equation (2.11) over z, one gets the differential branching probability of a in a $d\mu^2$ infinitesimal step

$$d\mathcal{P}_a\left(\mu^2\right) = \frac{d\mu^2}{\mu^2}\,\frac{\alpha_s\left(\mu^2\right)}{2\pi}\sum_{b,c}\int_{z_{min}(\mu^2)}^{z_{max}(\mu^2)} P_{a\to bc}(z)\;dz \quad . \tag{2.13}$$

Focusing now on ISR, the dynamics are derived from the PDF evolution equations

$$df_{a/h}\left(x_a,\mu^2\right) = \frac{d\mu^2}{\mu^2}\frac{\alpha_s\left(\mu^2\right)}{2\pi}\sum_b\int\frac{dz}{z}\,f_{b/h}\left(\frac{x_a}{z},\mu^2\right)\,P_{a/b}(z) \quad , \tag{2.14}$$

where $f_{i/h}$ is the PDF of parton i for hadron h, x_i the fraction i carries of the total momentum of h and $z = x_a/x_b$. Also, the $P_{a/b}$ functions give the probability of finding parton species a in parton species b, relating to the DGLAP splitting kernels like $P_{a/b}(z) = P_{a\to bc}(z)$, except for $P_{g/g} = 2P_{g\to gg}(z)$.

In contrast with FSR, which is formulated in terms of the decay rate of $a \to bc$, the ISR is formulated in terms of the production rate of a inside hadron h. The PDFs of the colliding hadrons evolve from a low scale Q_0^2 (when they are apart) to the scale of the hard process Q^2, the ones actually contributing to the cross section calculations. However, the colliding partons may not actually coincide with the 2 incoming showers. This problem is solved by using the evolved PDFs to select the hard process of interest and then construct the incoming showers backwards in time, from the high Q^2 scale to the low Q_0^2.

Therefore, the differential probability of parton a being produced in hadron confinement with x_a and for the infinitesimal evolution step $d\mu^2$ comes from Equation (2.14), such that

$$d\mathcal{P}_a\left(x_a,\mu^2\right) = \frac{df_{a/h}\left(x_a,\mu^2\right)}{f_{a/h}\left(x_a,\mu^2\right)} = \frac{d\mu^2}{\mu^2}\frac{\alpha_s\left(\mu^2\right)}{2\pi}\sum_b \int_{z_{min}(\mu^2)}^{z_{max}(\mu^2)} dz\, \frac{\frac{x_a}{z}\, f_{b/h}\left(\frac{x_a}{z},\mu^2\right)}{x_a\, f_{a/h}\left(x_a,\mu^2\right)}\, P_{a/b}(z) \quad . \tag{2.15}$$

The simple shower is, historically, the first parton shower model implemented in PYTHIA 8 and is still the default algorithm in PYTHIA 8.306. The shower evolution is ordered in terms of transverse momenta, computing the FSR and ISR splitting probabilities by setting $\mu^2 = p^2_{\text{T, evol}}$ in Eqs. (2.13) and (2.15), respectively. This choice of ordering variable stems from its proven capability of replicating key coherence features observed in experiment. Let's consider a partonic process $a+b \to c+d$, where a and b are extracted from incoming beams A and B. The ordering in transverse momenta is first guaranteed by starting from a maximal scale $p_{\text{T,max}}$. Then, each of the 4 partons a, b, c and d are sampled with a given branching p_{T} scale and the hardest parton is the one subjected to a splitting, creating a new 5-parton state. The selected p_{T} is taken as the new maximal scale from which to evolve the remaining partons further down in p_{T}. This algorithm is iterative and only stops when the branching p_{T} decreases below a minimal cutoff scale $p_{\text{T,min}}$, which has to be provided both for the ISR and FSR showers. For the latter, $p_{\text{T,min}}$ is closely related to the point at which hadronization kicks in.

2.2.3 Hadronization

The treatment of hadronization has to reflect the non-perturbative conversion of the final outgoing, coloured partons into colourless hadrons through a combination of analytical results and different QCD-based phenomenological models. In PYTHIA, hadronization is implemented following the Lund String model [17, 18]. For other hadronization mechanisms, one should look at the HERWIG multi-purpose particle physics event generator, which implements the cluster hadronization model [19].

The Lund String model is based on the assumption that the confining force field between a colour and an anticolour charge, such as the one in the simple example of a color-singlet $q\bar{q}$, can be viewed as a flux tube with potential energy increasing linearly with the inter-charge distance. This linear increase can be seen on the plot of Figure 2.2.1 a, where the QCD potential (in black) acquires linear behaviour for sufficiently large inter-particle distances r. This initial assumption, supported by lattice QCD results, likens hadronic confinement to the behaviour of a string, where an increase in the separation between the partons leads to the transfer of energy from the partons at the end of the string to the string itself (if the short-distance Coulomb term, represented in blue in Figure 2.2.1 a, is neglected). The energy transfer was found to occur at a rate of $k \approx 1$ GeV/fm in the linear confinement framework.

When looking at a $q\bar{q}$ dipole, as the q and $\bar{q}$ move apart form their common production vertex, a stretching occurs of the color flux tube established between them. This stretching can be modeled by a massless relativistic string with no transverse degrees of freedom, allowing us to treat the confinement problem as one-dimensional. Depending on the energies of the original pair, it may evolve into a stable so-called “yoyo” mode, for smaller production energies, or into a sequence of string fragmentations, for higher energies.

A yoyo mode is identified as a meson [7], whose flavour is determined by its quark content. Considering such a mode, aligned at the x-axis, that at timestamp $t = 0$ has all its energy stored in the string

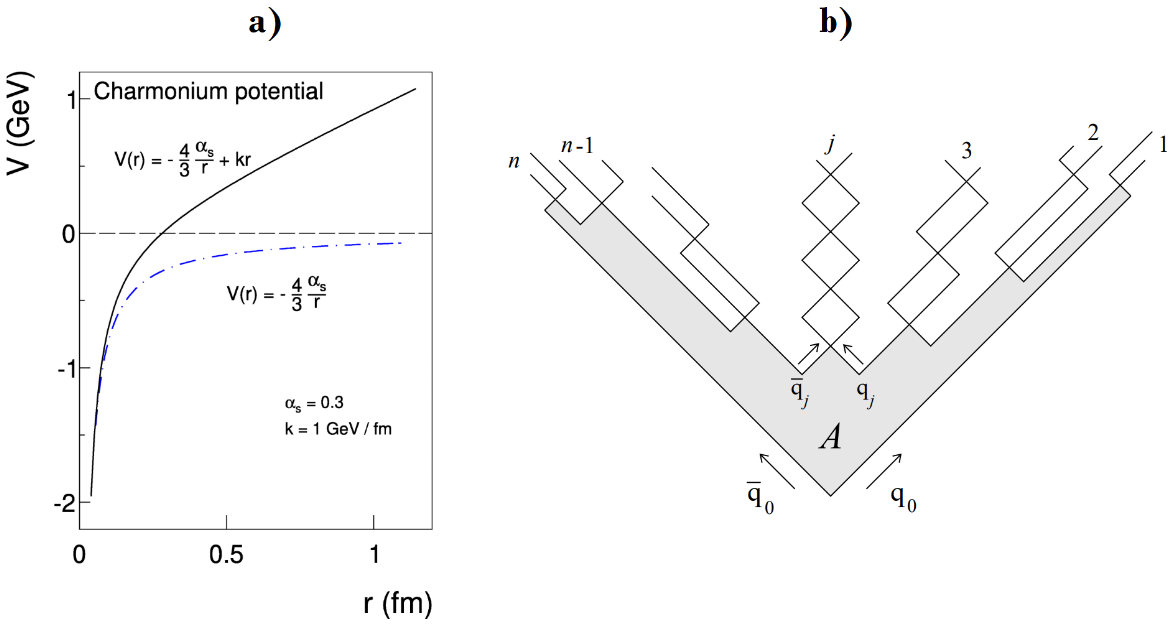


Figure 2.2.1: a) Typical QCD potential calculated at LO for a charmonium system (bound state of charm, c, and anti-charm, $\bar{c}$, quarks via strong interaction) taken from [20]; b) Quark-antiquark string breaking into hadrons in the reference frame of the original pair (Lund String model) taken from [21].

ends and none in the actual string, the kinematics of the dipole are $(E, p_x)_{q\bar{q}} = \frac{1}{2}\left(\sqrt{s}, \pm\sqrt{s}\right)$ and $E_{\text{string}} = 0$. The string then reaches its maximal extension at time $t = \sqrt{s}/2k$, at which one has $(E, p_x)_{q\bar{q}} = (0, 0)$ and $E_{\text{string}} = \sqrt{s}$. At timestamp $t = \sqrt{s}/k$, the string ends return to the starting point, but with symmetric momenta. At $t = 2\sqrt{s}/k$, the string completes a full cycle.

As for the case of a very high-energy $q\bar{q}$ dipole, created for instance in ultra-relativistic hadron collisions, the potential energy of the string increases with the increasing distance between the partons until it is high enough for a new parton/antiparton pair to be formed [22]. This leads the string to be broken up by the formation of new $q\bar{q}$ colour-singlet systems. After successive fragmentations, the energy scales reduce to a point where on-mass-shell hadrons ($p^2_{\text{hadron}} = m^2_{\text{hadron}}$) are produced. In the Lund String model, this is the point at which the fragmentation process is assumed to end. A schematic representation of string fragmentation for a highly energetic $q\bar{q}$ dipole is shown in Figure 2.2.1 b.

2.3 Jets

Present day experiments performed at the RHIC and LHC colliders show that jets, being extended objects, can provide detailed information about the QCD dynamics that governs the parton shower evolution [23] and some insight into the hadronization process. Since the partons that undergo branching in the aforementioned collisions are very highly energetic and the showers are angular ordered, their paths are highly collinear to the one of the original parton. The highly-collimated (collider reference frame) cone of energetic final-state particles that are subsequently produced is what is called a jet [5]. Jets circumvent the fact that the original partons out of the hard scattering cannot be measured experimentally. Therefore, one takes jets to be first order proxies for the individual incoming partons.

Jets display a fractal substructure rooted in the perturbative production of increasingly nearby partons. This jets-inside-jets-inside-jets substructure is naturally capped by the non-perturbative hadronization process, which takes the final-state partons and produces hadrons. The study performed in this work will revolve around jet substructure.

2.3.1 Clustering Algorithms

A jet definition is formed by a jet algorithm and a recombination scheme. To provide a consistent framework of jet definitions in both theory and experiment, multiple algorithms were developed to group collections of final-state particles into structures called jets and calculate the total jet momentum [5]. A jet algorithm is a set of rules that dictate the order by which one clusters the particles, pairing them up according to their kinematic information. Since each particle-pair-grouping is called a clustering step, these algorithms are named clustering algorithms. Clustering algorithms are associated with 1 or more parameters setting the angular and/or energy properties of jets. One of the most common parameters is the jet radius R, which defines the maximum *angular* reach of the jet algorithm, limiting how close 2 particles must be for them to belong to the same jet. This limitation leaves, in general, part of the radiation from the hard scattering parton outside of the jet's reach, which is why jets are only proxies for their initiators and not fully-faithful structures.

The steps taken by the clustering algorithm provide, nevertheless, an opportunity to probe the emission phase-space within jets via its substructure, where one searches for the effects of the hadronization procedure. In particular, they have been shown very recently in [24] to allow for the association of a timescale to the jet splittings that build up the fragmentation pattern. On the other hand, the recombination scheme outlines the way one assigns the total energy-momentum to the combination of 2 particles. For this thesis' work, the picked recombination scheme calculates total 4-momentum by simple addition, component by component, of the 4-momenta of the particles [25].

The very first jet algorithm was conceived for e^+e^- collisions in the late 1970's [26], giving birth to a multitude of jet algorithms that look for stable cone-shaped jets - the cone algorithms. However, the most commonly used jet algorithms (and the ones used in this thesis) are the sequential recombination jet algorithms. This happens because sequential recombination algorithms go beyond just finding jets and implicitly assign a clustering sequence to them [5], mimicking the heuristic parton branching pictures explored in Section 2.2.2. They treat jet constituents at the pair-level by quantifying the probability of 2 particles having the same origin using a certain distance measure d_{ij} (i, j indices identify the particles). This usually yields, however, jets of irregular geometry, unlike the ones found using cone algorithms.

The most widely used sequential recombination algorithms today belong to the so-called generalised k_t family of algorithms. They are, namely, the k_t, the Cambridge-Aachen (C/A) and the anti-k_t algorithms. Take $p_{\mathrm{T},i}$ as the transverse momentum of particle i and p as an algorithm-defining parameter that is 1 for the k_t algorithm, 0 for the C/A and -1 for the anti-k_t. Then, the d_{ij} distance measure used by the generalised k_t algorithms is defined as

$$d_{ij} \overset{\text{def}}{=} \min\left(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2} \quad , \tag{2.16}$$

where ΔR_{ij}, a measure of the *angular* distance between particles i and j, is given by

$$\Delta R_{ij} \overset{\text{def}}{=} \sqrt{\left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2} \quad , \tag{2.17}$$

with $y_{i(j)}$ the rapidity of the particle $i(j)$ and $\phi_{i(j)}$ its azimuthal angle. For a particle with energy E and longitudinal momentum p_L with respect to the beam direction, rapidity is a variable defined as

$$y \overset{\text{def}}{=} \frac{1}{2} \ln\left(\frac{E + p_L}{E - p_L}\right) \quad . \tag{2.18}$$

In the high-energy limit approximation (quarks assumed massless), rapidity becomes a new variable called pseudorapidity, defined as

$$\eta \overset{\text{def}}{=} - \ln\left(tan\left(\frac{\theta}{2}\right)\right) \quad , \tag{2.19}$$

where θ is the angle of the particle's 3-momentum with respect to the beamline. Small θ angles lead to very high η, while $\theta = 90^\circ$ is equivalent to $\eta = 0$.

Generalised k_t algorithms have a second distance measure known as the *particle-beam* distance, defined as

$$d_{B,i} \overset{\text{def}}{=} p_{\mathrm{T},i}^{2p} \quad . \tag{2.20}$$

These jet algorithms go over the following steps:

1. from the list of all final-state particles, calculate the distance measures $d_{B,i}$ and d_{ij} for each particle and every particle-pair;
2. for each particle i, find the minimum distance measure and
    - if the smallest distance measure is $d_{B,i}$, remove i from the list of particles and recombine it with the beam, adding it to the so-called "beam-jet";
    - if the smallest distance measure is one of the d_{ij} and $d_{ij} < R$, cluster i and j into a new pseudo-particle k, remove i and j from the list of particles and add k;
    - if the smallest distance measure is one of the d_{ij} but it verifies $d_{ij} \geq R$, i is a jet; repeat step 2 for another particle;
3. repeat the algorithm for the new list of particles until all particles have $d_{ij} \geq R$ has their smallest distance measure; at that point, all jets are found.

The first time these algorithms were adapted to hadron-beam collisions happened for the k_t algorithm in 1992 [27]. Its inter-particle distance measure, obtained by setting $p = 1$ in Equation (2.16), is minimized by clusterings with at least 1 low p_{T} particle, favouring the construction of parton showers with an abundance of soft collinear branchings, consistent with QCD predictions. However, this renders the k_t algorithm inadequate to the treatment of pp collisions, due to the high jet multiplicities of such events.

The Cambridge algorithm was firstly developed for e^+e^- experiments, but its relevant extension to hadron colliders is the Cambridge/Aachen algorithm ($p = 0$) [28, 29]. This was designed to be a purely geometrical algorithm, with its distance measure $\Delta R_{ij}^2/R^2$ only dependent on geometric distance scales. It returns jets with decreasing *angular*-ordered branching schemes, consistent with pQCD-prescribed angular ordering, that make it possible to consistently view them across various angular and momentum scales.

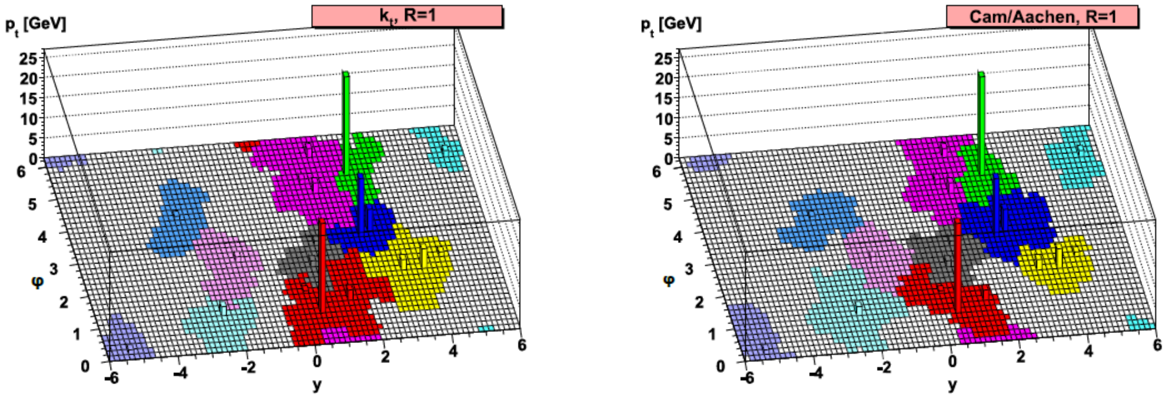
More recently, the anti-k_t algorithm ($p = -1$) [30] was discovered to favour clusterings that involve hard particles, to the detriment of clusterings that involve soft particles (k_t algorithm) or energy-independent clusterings (C/A algorithm). Calling the hardest particle i^* and the particle closest to it in the y-ϕ plane j^*, they are immediately clustered together as a new k^* particle (provided $d_{i^* j^*}$ survives the $min(d_{i^* j^*}, d_{B,i^*})$ imposition). Then, k^* is the new hardest particle of the list and it repeats this collinear clustering logic until the geometrical distance in y-ϕ becomes larger than R, at which point a jet is fully found. This increases the likelihood of identifying actual products of the hard scattering event, produced with high transverse momenta, making the anti-k_t algorithm a regular analysis tool used in the LHC and RHIC high-energy colliders. It also grants the anti-k_t algorithm a greater protection against the predominantly soft emissions of background radiation, which the other 2 algorithms do not possess. Furthermore, the anti-k_t algorithm usually yields circular-shaped jets in the *angular* coordinates plane. The tagging of soft-resilient circular jets combines all the features of an “ideal” cone algorithm, effectively rendering them obsolete. However, anti-k_t clustering trees lack the physical interpretation of each clustering step as a branching. The anti-k_t algorithm utilizes a nonphysical ordering variable, dependent on $1/p_T^2$, which requires all clusterings to include at least 1 hard particle, leading to clustering trees with a single hard prong from which softer particles are emitted with no further branchings to follow. This jet substructure is manifestly non-QCD-like, with pQCD calculations strongly contradicting it.

The effects of algorithm-choice on jet-identification for an example data set can be seen in Figure 2.3.1 (clusterings with k_t, C/A and anti-k_t).

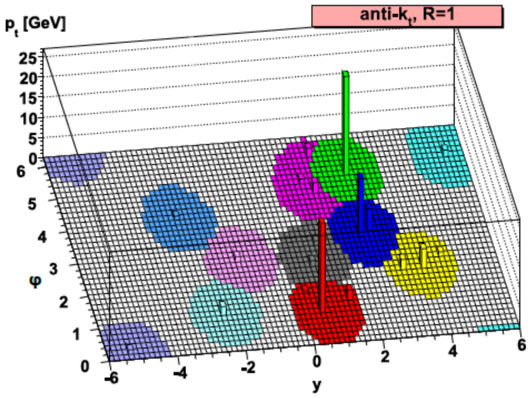
Other algorithms can arise from different values of the parameter p, namely the τ algorithm for $p = 0.5$ [24], which is explored in Section 2.3.4. These algorithms are used in combination in this thesis to balance out the strengths and weaknesses they have in the processes of jet-identification and jet-reconstruction, as is described in detail in Section 3.2.

2.3.2 Soft-drop Algorithm

One of the most recent techniques applied to jet analysis is concerned with removing soft wide-angle radiation from the identified jets, in a process designated by jet grooming. Jet grooming usually results in the removal of particles from the periphery of jet-cones, slimming them down to their harder collinear cores. Jet grooming algorithms allow for better comparisons between experimental data and pQCD calculations (like the DGLAP splitting functions in Equations (2.12)), since they clean the jets



(a) The k_T algorithm. **(b)** The C/A algorithm.



(c) The anti-k_T algorithm.

Figure 2.3.1: Jet identification using the k_t (a), C/A (b) and anti-k_t (c) jet clustering algorithms, shown in the y-ϕ plane as their "active" catchment areas in bright colors over the grey grid, for a given sample of a parton-level event (taken from [5]).

from soft contamination (mainly npQCD effects) clustered into the branching sequences by certain jet algorithms, such as the k_t and C/A. Many grooming algorithms have been conceptualized. However, the most commonly used one and the one chosen to incorporate in this work is the soft-drop (SD) algorithm [23].

As is explored in greater detail in Section 3.2, this thesis focuses on pp collisions whose jets are initially identified using the anti-k_t algorithm. Then, the identified jets are re-clustered with the C/A algorithm to obtain angular ordered clustering trees, guaranteeing early emissions to be more wide-angled. The SD algorithm takes the C/A re-clustered jets and de-clusters them recursively, following the main branch (branch of highest transverse momentum). During this process, the soft-drop criterion, expressed in Equation (2.21), is applied to check, at each de-clustering step, if soft emissions occur, flagging them to be dropped if they do [23]. The SD criterion is, therefore, expressed by

$$\frac{\min\left(p_{\mathrm{T}1}, p_{\mathrm{T}2}\right)}{p_{\mathrm{T}1} + p_{\mathrm{T}2}} > z_{cut}\left(\frac{\Delta R_{12}}{R}\right)^{\beta} \quad , \tag{2.21}$$

with ΔR_{12} the *angular* distance defined in Equation (2.17) and $p_{\mathrm{T}1}$ and $p_{\mathrm{T}2}$ the transverse momenta of the J_1 and J_2 branching subjets, each one a proxy for 1 of the daughter-partons that arise from the splitting.

The fixed parameters z_{cut}, which is the soft threshold, and β, which is the angular exponent, regulate the severeness of the SD condition and, therefore, the effectiveness of the jet grooming. For this thesis, the angular exponent is set to $\beta = 0$, which selects the so-called modified mass drop tagger (mMDT) algorithm from the SD algorithm family. However, from this time forward, it will be referred simply as the SD algorithm. Also, the soft threshold is set to $z_{cut} = 0.1$, meaning a branching is only SD approved if the partition of the mother's initial p_{T} is more equitable than a 10% / 90% share between the softest and hardest particle, respectively.

The actual SD algorithm starts from the fully (re-)clustered jet J and looks for the 1st soft-drop emission, corresponding to the first de-clustering step that satisfies the SD criterion. It does that by

1. de-clustering J into 2 subjets, J_1 and J_2, by undoing the last stage of C/A clustering;
2. subjecting J_1 and J_2 to the SD criterion;
3. if the SD criterion
    - is passed, the algorithm stops and the 1st SD emission coincides with the 1st de-clustering step of the tree;
    - fails, the jet is redefined as the subjet with highest jet p_{T} (main branch), while the softer subjet is groomed away;
4. repeating this steps until a de-clustering step that passes the SD criterion is found; otherwise, the entire jet is discarded, since it contains no SD emission.

The features of the groomed jet that arises (if any does) from this picture of the SD algorithm are fundamentally portrayed by 2 new variables: the groomed jet momentum fraction z_g and the groomed jet radius R_g (not Lorentz invariant). They are both evaluated at the 1st SD emission and are given by

$$z_g \overset{\text{def}}{=} \frac{\min(p_{\mathrm{T}1}, p_{\mathrm{T}2})}{p_{\mathrm{T}1} + p_{\mathrm{T}2}} \qquad \text{and} \qquad R_g \overset{\text{def}}{=} \Delta R_{12} \tag{2.22}$$

with 1 and 2 now referring to the 2 earliest surviving branches of the groomed jet and ΔR_{ij} explicitly provided by Equation (2.17). The groomed momentum fraction $z_g > z_{cut}$ is the fraction of the total transverse momentum of the source object that is carried out by the softest daughter of the 1st SD emission. The groomed radius R_g is the *angular* distance between the 2 subjets and constitutes the maximally allowed geometric opening in the groomed clustering tree, given the *angular* ordering of the C/A algorithm.

However, after the 1st SD emission is found (if any is found), the algorithm can be repeated along the main branch by following along the hardest prong at each de-clustering step. The result of this iterative procedure will yield the 2nd, 3rd, ... SD emissions, as will be necessary for some of the work performed in this thesis.

State of the art experimental [31–33] and theoretical [34–37] research, more recently in the HICs field, has increased our understanding of soft-drop groomed jet observables, making it an increasingly popular jet analysis tool. It was specifically shown in [24] to be an outstanding tool to increase the correlation between the jet clustering history and the actual parton shower when taking a look at splitting timescales. This study is performed with the substructure scale of formation time, which is a proxy for the time it takes for a parton to split (presented in full detail in Section 2.3.3). Using the SD algorithm with $z_{cut} = 0.1$ and $\beta = 0$, a significant decrease was observed in the dispersion between the formation time of the first de-clustering step and the first parton shower emission, as is shown in Figure 2.3.2.

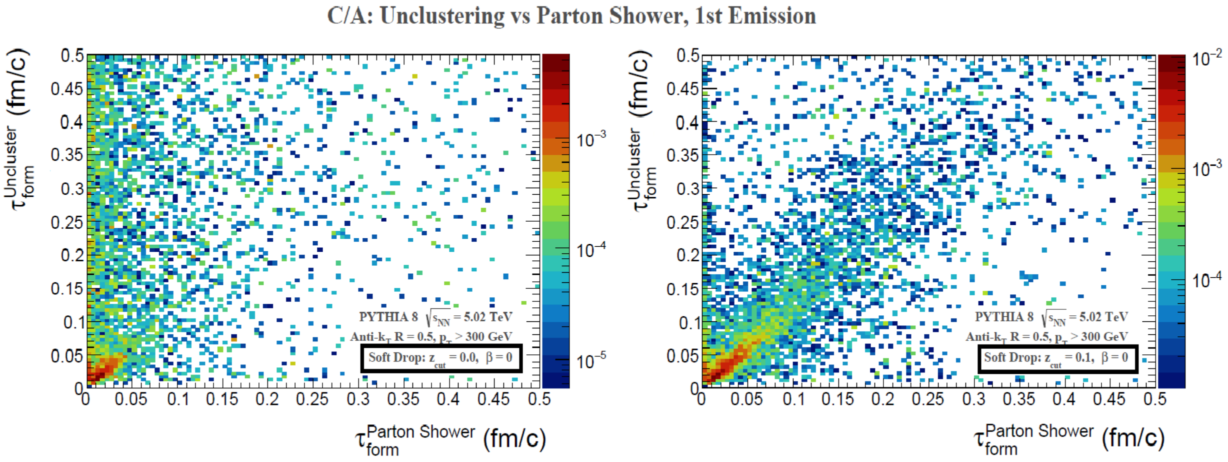


Figure 2.3.2: **Left**: correlation of the formation time τ_{form} between the first parton shower emission and the first de-clustering step of a C/A clustered jet with no jet grooming; **Right**: same correlated variables after SD-grooming the C/A clustering trees with $z_{cut} = 0.1$ and $\beta = 0$ (taken from [24]).

Though the SD algorithm was initially conceived for C/A re-clustering, this thesis will extend SD-grooming to jets re-clustered with the τ algorithm presented in Section 2.3.4, to the likes of what is performed in [24].

2.3.3 Formation Time

The concept of formation time was firstly developed as a jet analysis tool in the paper [24] and was recently measured at RHIC by the STAR collaboration [38].

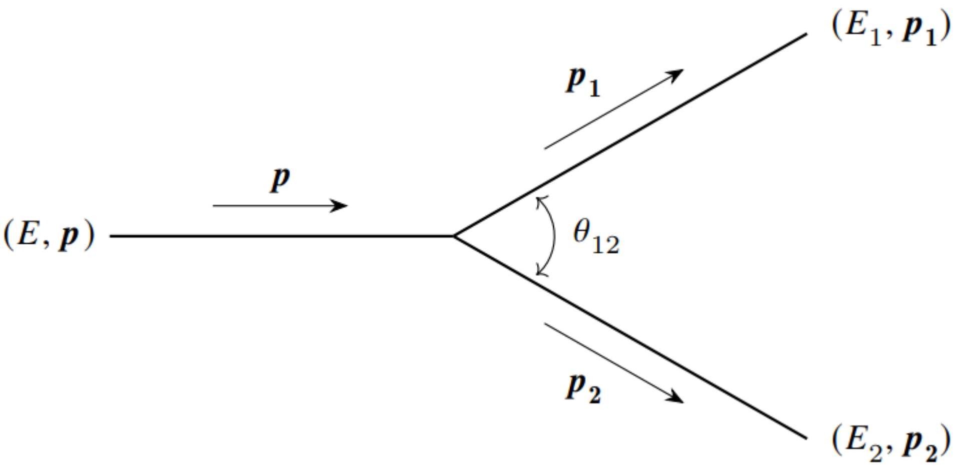


Figure 2.3.3: Scheme of a source with momentum p splitting into 2 daughters of momentum p_1 and p_2 and opening angle θ_{12};

The formation time τ_{form} is the time that a quantum state, such as a parton, takes to behave as 2 independent sources of additional radiation [39], like the 2 daugther-partons from a splitting. A given $1 \to 2$ splitting involves a source, with 4-momentum $p^{\mu} = \left(E, p_x, p_y, p_z\right)$ and mass $m^2 = p^{\mu} p_{\mu}$, and 2 daughters, with 4-momenta p_1 and p_2 and masses m_1 and m_2, respectively (see splitting configuration in Figure 2.3.3). The simplest estimation of the formation time of such a splitting is given by dimensional analysis (in natural units) [39] and is formulated, in the rest frame of the source, as

$$\tau_{\text{form, CM}} \propto \frac{1}{m} \quad . \tag{2.23}$$

By Lorenz boosting the splitting into an arbitrary new frame (like the collider's reference frame), the time estimate transforms to

$$\begin{aligned} \tau_{\text{form}} \sim \frac{E}{m} \cdot \frac{1}{m} &= \frac{E}{m_1^2 + m_2^2 + 2 p_1 \cdot p_2} \\ &\approx \frac{E}{2 E_1 E_2 \left(1 - \cos\theta_{12}\right)} = \frac{1}{2\, E\, z\, (1 - z)\, \left(1 - \cos\theta_{12}\right)} \quad , \end{aligned} \tag{2.24}$$

where $E = E_1 + E_2$ is the total energy, E_1 and E_2 are daughters' energies, θ_{12} the angle between the trajectories of the daughters and z is the energy fraction, defined as

$$z = \frac{\min\left(E_1, E_2\right)}{E_1 + E_2} \quad . \tag{2.25}$$

The last expression in Equation (2.24) is the one used in this thesis to calculate the formation time of selected splittings [24]. It was derived from the expression in the line above it through the high-energy limit approximation $m_i << |\boldsymbol{p_i}|$, with $\boldsymbol{p_i}$ the euclidean 3-momentum of particle i.

From the contribution of the total energy to the denominator in Equation (2.24), it comes that the more energetic a parton is, the shorter is the time it takes to split into 2 other partons (smaller τ_{form}). Indeed, the soft-drop emission with the smallest formation time is usually the first one, since it takes place at the highest energy scales of the branching sequence. Then, since parton showers are p_{T} ordered (and p_{T} serves as a proxy for energy), the splitting formation time has great potential to serve as a shower-ordering variable. This is precisely what is explored in Section 2.3.4. Equation (2.24) also reveals that splittings happen earlier when they are energetically symmetric and have wider opening angles.

2.3.4 τ algorithm

Applying the soft ($z << 1$) and collinear ($\theta_{12} << 1$) limits to a parton splitting like the one portrayed in Figure 2.3.3, the formation time estimate in Equation (2.24) can be simplified into

$$\tau_{\text{form}} \sim \frac{1}{E\, z\, \theta_{12}^2} \quad . \tag{2.26}$$

Recalling the main distance measure d_{12}^{p} of the generalized k_t algorithm of parameter p, expressed in Equation (2.16), and using p_{T} as a proxy for E and ΔR_{12} as a proxy for θ_{12}, the simplified formation time becomes

$$(\tau_{\text{form}})^{-1} \sim \min\left(p_{\mathrm{T}1}, p_{\mathrm{T}2}\right)(\Delta R_{12})^2 = R^2\, d_{12}^{p=0.5} \quad . \tag{2.27}$$

Therefore, the main distance measure $d_{12}^{p=0.5}$ of the $p = 0.5$ generalized k_t algorithm - the τ algorithm - becomes a proxy for the formation time of the clustered particles, when in the high-energy, soft and collinear limits. This means the ordering variable of the τ algorithm can be generally taken to be the splitting formation time of successive clustering steps. Clusterings that maximize the formation time are favoured over others, which will be very useful in the search for the hadronization timescales. However, it is important to note that, since the high-energy soft collinear approximation is not universally valid, the τ clustering trees may not, in fact, be strictly τ_{form}-ordered. The time-ordering is, nevertheless, reinforced by soft-drop-grooming of τ trees, which removes wider-angled emissions that fall outside these approximations.

A comparative analysis of the generalised k_t algorithms (described in Section 2.3.1) used to perform the re-clustering of anti-k_t jets also yields interesting results in terms of the dispersion between the timescales of the clustering history and of the parton shower branching sequence. That study was performed in [24] and it found that τ re-clustering optimizes the correlation between the two. It also showed the τ algorithm to perform the best when trying to get unbiased estimates of splitting formation times.

One starts by evaluating the difference $\Delta\tau_{\text{form}}$ between the formation time of a given de-clustering step and the formation time of the corresponding parton shower emission for jets re-clustered with different generalised k_t algorithms. These different algorithms arise from setting the parameter p in Equation (2.16) to a set of different values: $p \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. Then, for a given jet population and for re-clustering performed with each one of these generalised k_t algorithms, the median Q_2 and the quartiles Q_1 and Q_3 were calculated in [24] for the distributions obtained with respect to $\Delta\tau_{\text{form}}$. The results are shown in Figure 2.3.4 with the values for the first splitting represented in orange, for the second splitting in green and for the entirety of the main branch in purple markers. The jet populations used in the study were all groomed with the SD algorithm using $z_{cut} = 0.1$ and $\beta = 0$.

From Figure 2.3.4, one can observe (on the right panel) that Q_2 is very close to 0 for all p, but it gets the closest to 0 for $p \sim 0.5$. Another observation (from the left panel) is the fact that the $\Delta\tau_{\text{form}}$ distributions are generally narrower and more symmetric the closer p is to 0.5, with more equitable Q_1 and Q_3 quartiles. The generalised k_t algorithm with $p = 0.5$ is shown here to optimize the correlation of the recombination sequence with the parton shower. Also, the symmetric nature of the distributions for the τ re-clustering algorithm is particularly important to obtain unbiased estimates of formation time.

Therefore, the τ algorithm will be used in parallel with the C/A algorithm to perform a robust comparative study of the relevant jet substructure observables and to evaluate the algorithm-dependence of

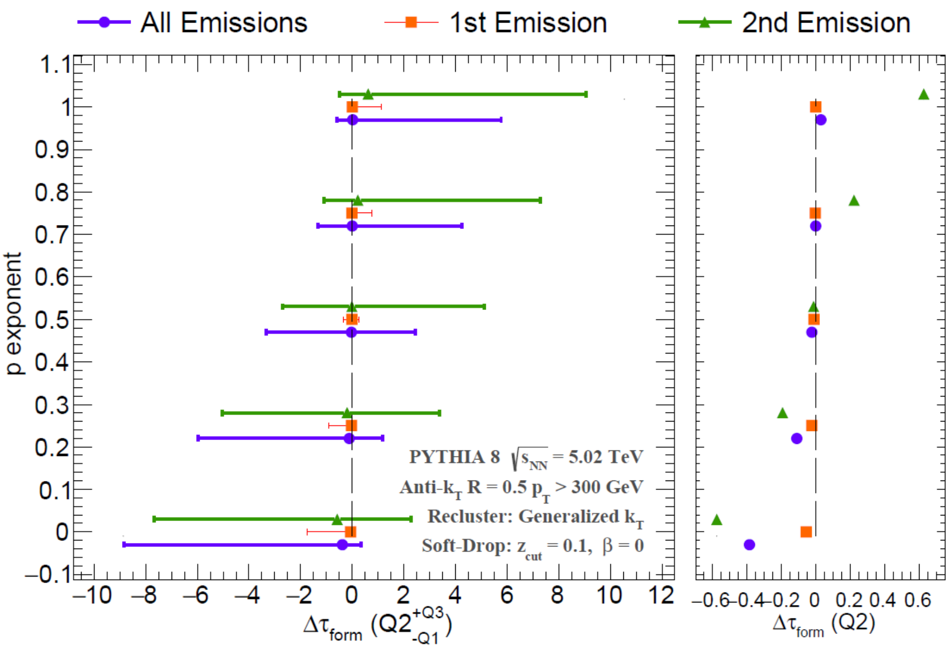


Figure 2.3.4: **Left**: the median Q_2 value of the $\Delta\tau_{\mathrm{form}}$ distribution evaluated for different values of the parameter p of the generalised k_t algorithms, shown for the first, second and all SD emissions along the main branch in orange, green and purple markers, respectively (the median values are accompanied by asymmetric error bars, corresponding to the $+Q_3$ and $-Q_1$ quartiles); **Right**: a close-up of the Q_2 values (for each p exponent value, the orange, green and purple dots are slightly offset vertically from the central value to improve readability); Figure taken from [24].

the estimates at the re-clustering level. When re-clustering with the τ algorithm, the first SD approved de-clustering step is usually the one with the smallest τ_{form}, since the clusterings are generally time-ordered. For the C/A re-clustering, time-ordering is even less strict. However, the probability of the first de-clustering step from C/A SD-groomed trees not being the one with shortest formation time was found to be of a few percentage points [24].

2.4 Jet-quenching

A lot of new research is being developed for collisions of heavy ions at relativistic energies (HICs). At these energies, the ions are Lorentz contracted longitudinally, which allows for very high densities to be reached in these collisions, replicating the conditions of the early Universe (although at very small scales). Current telescopes are limited in their ability to see beyond the cosmic microwave background and methods that look past it, like the gravitational wave astronomy performed at LIGO, are still at their infancy [38]. Therefore, HICs are currently the most promising attempt of the scientific community to study the early Universe.

Immediately after a HIC, an emergent state of matter appears - the quark-gluon plasma (QGP). The QGP is a “soup” of deconfined quarks and gluons that lasts for a very short period of time. The QGP experiences hydrodynamic expansion soon after its creation, accompanied by a cooling that eventually leads the deconfined partons to go through a QCD phase transition into color neutral hadrons [40]. However, the emergence of the QGP introduces changes to the jets produced in this type of events. For instance, QGP induces energy losses in the ejected partons that ultimately change the average profile of a quenched jet with respect to a vacuum pp jet. Jet quenching is, precisely, “a collection of medium-induced modifications of the jets’ internal structure that occur through their development in dense QCD matter” [24].

Although this thesis focuses only on vacuum pp collisions, the results obtained can potentially translate into useful tools to the heavy-ion field of research, namely to study the space-time evolution of QGP.

Chapter 3

Jet Substructure: Perturbative Regime

With the theoretical foundations laid in great detail in Chapter 2, it is now the place for displaying and discussing the results, both graphical and numerical, obtained from jet analysis of simulated pp collisions. This Chapter focuses on how jet substructure translates into the study of QCD physics at perturbative scales, namely at the early parton branching level. Appropriate substructure variables are studied - the groomed opening angle and groomed momentum fraction, defined in Equation (2.22), and the groomed formation time, defined in Equation (2.24) -, in addition to a comparative analysis based on varying collision energies and phase-space cuts in jet transverse momentum. This allows for an in-depth study of the influence of jet-initiating parton kinematics on substructure dynamics.

This Chapter starts by looking at the first soft-drop emission and is afterwards expanded to later soft-drop splittings, but still from the earliest stages of the clustering tree to evaluate how substructure changes alongside the partonic branching history. This Chapter also specifies the simulation setups that are implemented using the multi-purpose Monte Carlo event generator PYTHIA, in its latest 8.306 version (as of the writing of this document) and the algorithmic framework employed for jet analysis.

3.1 Simulation Setup

When simulating events with PYTHIA, the user has to set a number of minimal initialization parameters. One of them is the CM energy of the pp collision one wants to simulate, which can be tunned via the *Beams:eCM* input variable. As is motivated in Chapter 1, this thesis aims to reproduce RHIC and LHC collisions. As such, PYTHIA is alternately set in the CM frame with $\sqrt{s} = 0.2$ TeV (RHIC), $\sqrt{s} = 5$ or 14 TeV (LHC).

Furthermore, PYTHIA requires the user to input what are the type of processes they want to simulate. In order to enable QCD $2 \to 2$ quark/gluon scattering processes (excluding heavy-quarks starting from charm [7]), the $HardQCD : all$ flag is turned on, which internally weights all the processes, such as $gg \to gg$ and $gg \to q\bar{q}$, by their respective cross sections.

Another input parameter is a minimum transverse momentum requirement $\hat{p_T}^{min}$ that the user imposes on the hard scattering partons via the *PhaseSpace:pTHatMin* variable. This is required to only generate events where partons are produced with $p_{\mathrm{T}} \gtrsim \hat{p_T}^{min}$. This is useful for a narrow study on a given $p_{\mathrm{T,jet}}$ range, since it improves the sampling efficiency on that phase-space cut. In fact, the p_{T} spectrum for ultra-relativistic hadron collisions has a sharp (power law) decreasing behaviour with respect to transverse momenta (shown ahead in Figure 3.2.1), creating a necessity for such cuts when investigating higher transverse momentum jets. However, the hard cut in transverse momentum imposed by $\hat{p_T}^{min}$ causes a distortion to the $p_{\mathrm{T,\,jet}}$ spectrum around that cutoff value. Therefore, the $\hat{p_T}^{min}$ threshold has to be set at values sufficiently lower than the lower limit of the $p_{\mathrm{T,jet}}$ range of selection to preserve the jet spectra shapes for those ranges. The $\hat{p_T}^{min}$ cutoff value set for the several $p_{\mathrm{T,jet}}$ intervals studied in this thesis can be seen in Table 3.1.

$p_{\mathrm{T,jet}}$ Range [GeV/c]	$\hat{p_T}^{min}$ [GeV/c]
$15 < p_{\mathrm{T,jet}} < 20$	10
$20 < p_{\mathrm{T,jet}} < 40$	16
$30 < p_{\mathrm{T,jet}} < 40$	25
$200 < p_{\mathrm{T,jet}} < 300$	170

Table 3.1: Cutoff values of $\hat{p_T}^{min}$ imposed via PYTHIA for the different ranges of jet transverse momentum $p_{\mathrm{T,jet}}$ relevant to this thesis.

When it comes to other event features, such as the parton shower model or the hadronization details that are implemented, the events use the default shower model (simple shower) and the default Lund String model of PYTHIA 8.306.

Finally, one has to set the total number of events one wants to simulate. For each different set of PYTHIA input parameters, a sample of $2 \cdot 10^6$ pp collisions is simulated to assure relatively small statistical uncertainties. For each event, PYTHIA returns a list of the particles produced in the aftermath of the collision, from which one retrieves the final-state particles alongside their kinematics, given by their 4-momenta $p^{\mu} = \left(E, p_x, p_y, p_z\right)$, and their electric charges. This information is then transferred to the jet finder package *FastJet* [25] and used to create a *FastJet* object called a *PseudoJet* for each final-state particle.

3.2 Jet Analysis Details

FastJet provides extensively tested implementations of jet definitions, namely of the clustering algorithms described in Section 2.3. The initial jet-finding clustering is performed with the anti-k_t algorithm and requires only 1 parameter: the jet radius, set to $R = 0.6$ for all simulation settings. As is carefully explained in Section 2.3.1, anti-k_t performs the best at triggering on a “hard-like” structure and identify it as a jet. The jets found through this initial clustering procedure are then subjected to a selection process that discards the ones with pseudorapidities (defined in Equation 2.19) outside the $-1 < \eta_{\mathrm{jet}} < 1$ interval. The left panel from Figure 3.2.1 shows the p_{T} spectrum of the inclusive particles produced in $\sqrt{s} = 0.2, 5, 14$ TeV pp collisions, represented in red circle, blue square and black diamond markers, respectively. The right panel from Figure 3.2.1 shows, for the same CM energies and with the same marker scheme, the $p_{\mathrm{T,jet}}$ spectra for the inclusive jets found with $|\eta_{\mathrm{jet}}| < 1$ and with anti-k_t clustering.

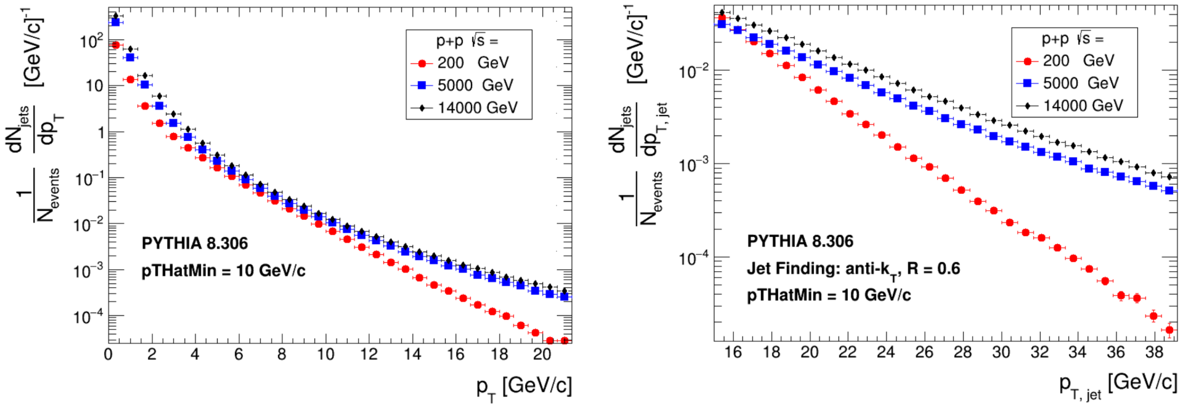


Figure 3.2.1: **Left**: transverse momentum (p_{T}) spectra for the inclusive particles produced in pp collisions with center-of-mass (CM) energies $\sqrt{s} = 0.2, 5, 14$ TeV, represented in red circle, blue square and black diamond markers, respectively; **Right**: jet transverse momentum ($p_{\mathrm{T,jet}}$) spectra found with anti-k_t clustering, $R = 0.6$ and $|\eta_{\mathrm{jet}}| < 1$, for the same CM energies and shown with the same marker scheme; simulations are set with $pTHatMin = 10$ GeV/c and all distributions are normalized by the total number of simulated pp collisions, $2 \cdot 10^6$.

Figure 3.2.1 reveals sharp drops in the probability of finding both particles and jets with higher transverse momenta. It also shows that greater $\sqrt{s}$ vacuum collisions produce harder final-state particles and, therefore, jets with higher transverse momenta. However, the particle p_{T} spectra are significantly shifted towards smaller values than the $p_{\mathrm{T,jet}}$ spectra since jets are generally composed of a multitude of particles, whose transverse momenta sum up to the jet transverse momentum.

It is known from the discussion in Section 2.3.1 that the identified jets, with their anti-k_t clustering trees, are not appropriate for jet substructure studies as they are, having a single hard branch from which all particles are emitted. Therefore, each jet is further de-clustered through the extraction of its constituents. The jet constituents are then submitted to 2 distinct re-clustering processes, one using the C/A algorithm (Section 2.3.1) and the other using the τ algorithm (Section 2.3.4), both with jet radius set to $R_{\mathrm{new}} = 1.0$. The anti-k_t clustering factors particle transverse momentum explicitly in its distance measure through a $1/p_T^2$ factor (see Equation (2.16) with $p = -1$), meaning particles can have *angular* distances greater than 0.6 and still find themselves being clustered together and included in the same jet if one of their p_T is high enough. Since the chosen re-clustering algorithms calculate, in general, higher distance measures than anti-k_t (p_T^0 dependence for C/A and p_T^1 dependence for τ), the re-clustered jets may surpass the previous $R = 0.6$ limit. This will happen mainly for particles close to the jet frontiers. Therefore, the choice of $R_{\mathrm{new}} > R$ ensures that each initial jet is replaced by 1 and only 1 re-clustered jet, with the exact same constituents as the first one, for each re-clustering algorithm.

In the end, each anti-k_t jet is replaced by both a C/A and a τ version of itself. The C/A algorithm, with its purely geometric distance measure, produces clustering trees which resemble QCD angular ordering, making it the most suited to evaluate QCD-like features, such as the groomed momentum fraction and opening angles like the groomed jet radius. However, the τ algorithm performs better at identifying correctly the formation time, namely for the 1^{st} soft-drop emission (as is shown in Section 2.3.4), since it is built around τ_{form} approximately being its ordering variable. Since this thesis focuses extensively on both z_g and τ_{form}, the analysis is performed using both re-clustering algorithms for a comparative study of the results' dependency on re-clustering procedures.

The re-clustered jets are then groomed by the soft-drop algorithm (Section 2.3.2), which is implemented externally to *FastJet* using *C++* code with $\beta = 0$ and $z_{cut} = 0.1$.

For 2 of the kinematic setups imposed in this thesis - $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions (RHIC) and $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV ones (LHC) -, the total number of soft-drop approved splittings along the main branch N_{splits} found for each jet is, on average, of around $3-4$ and $4-5$, respectively, for both re-clustering algorithms, as can be deduced from the distributions in Figure 3.2.2. The main branch of a jet is the one that follows the hardest prong from successive branchings. The RHIC distributions are represented on the left panel in red square and black diamond markers for C/A and τ re-clustering, respectively, while the LHC ones are shown on the right panel in blue square and black diamond markers. The N_{splits} is, on average, greater for the LHC kinematic settings because greater $\sqrt{s}$ collisions tend to require larger numbers of branchings to reduce energy scales to hadronization ones.

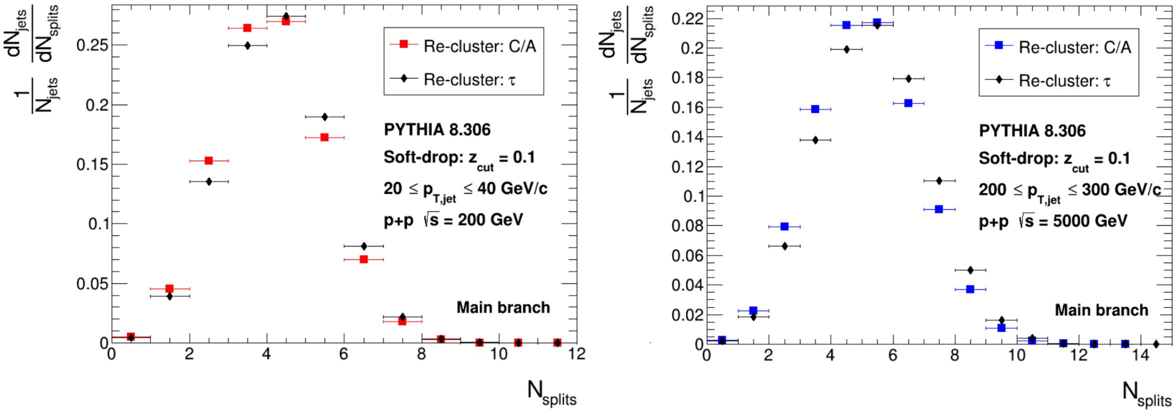


Figure 3.2.2: **Left**: distributions with respect to the per jet total number of soft-drop (SD) emissions along the main branch for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions, represented in red square and black diamond markers for C/A and τ re-clustering, respectively; **Right**: same distributions found for $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV collisions, represented in blue square and black diamond markers for C/A and τ re-clustering, respectively; the SD criterion is applied with $\beta = 0$ and $z_{cut} = 0.1$ to each emission and all distributions are self-normalized.

All the graphical results are plotted using the ROOT [41] framework. The error bar for each histogram bin is calculated using ROOT's *SetDefaultSumw2()* method, which calculates and stores the statistical error ϵ associated with the content of bin i by applying the formula

$$\epsilon_i = \sqrt{\sum_{j=1}^{n} \omega_{i,j}^2} \quad , \tag{3.1}$$

where $\omega_{i,j}$ is the weight of the j-th bin entry from bin i, out of a total of n entries. All histograms presented in this Chapter are self-normalized and each histogram's bin content is divided by the width of the respective bin.

3.3 Parton-Species Studies

This Section investigates jet substructure features - the groomed jet radius R_g, the groomed jet momentum fraction z_g and the groomed jet formation time τ_{form} - in the various CM energies and jet transverse momentum ranges mentioned in Section 3.1.

In Section 2.1.3, the proton PDFs are introduced as the probability density functions of finding a parton of a given species carrying a x fraction of the proton's total energy-momentum, also known as the Bjorken scale. Proton PDFs are shown, for a fixed $Q^2 = 100$ GeV^2 energy scale, in Figure 2.1.2. Even though the pp collisions of interest to this thesis have significantly higher hard scattering scales than $Q^2 = 100$ GeV^2, the proton PDFs are qualitatively the same, with gluon probability somewhat increasing for higher Bjorken x. With x acting as an estimation of the resolution scale at which the proton is probed in a pp collision, Figure 2.1.2 indicates that the proton as a predominance of quark-matter for small resolutions, while greater resolutions quickly present an overwhelming abundance of gluons. The Bjorken x can be estimated, to a first order approximation [42], by

$$x \sim \frac{p_{\text{T}}}{\sqrt{s}/2} \quad . \tag{3.2}$$

This is what motivates an analysis in different $p_{\text{T, jet}}$ ranges, as well as $\sqrt{s}$ values, as a way of varying the scales at which the proton is resolved and consequently the heuristic quark-gluon fraction of the jet-initiating partons scattered in pp collisions. The choice of kinematic settings acts as a selection on the parton-species at the hard parton-production level. Lower transverse momentum (softer) jets from higher CM energy hard scattering events probe the proton at smaller Bjorken x (higher resolution scales) and vice-versa. The Bjorken x estimates for all kinematic settings used in this thesis can be found in Table 3.2. They show that all selected jets from $\sqrt{s} = 5000$ GeV and $\sqrt{s} = 14000$ GeV collisions are predominantly initiated by gluons, while jets from $\sqrt{s} = 200$ GeV pp collisions are more dominated by quarks for higher $p_{\text{T, jet}}$, shifting towards gluons for lower $p_{\text{T, jet}}$. Therefore, the study performed in this Section deepens the understanding on how the nature of the jet-initiating partons conditions the substructure of the jets produced in vacuum collisions.

To get a fuller picture when it comes to parton-species studies, one also has to account for the effects of the parton-species on the jet-reconstruction level. Since gluons radiate more than quarks (by approximately a factor of 2), jet-reconstruction recovers smaller fractions of the total radiation emitted by a hard gluon. This inevitably leads to gluon-initiated jets having a significantly smaller $p_{\text{T,jet}}$ than the p_{T} of the original scattered gluon. This loss of radiation also happens for quark-initiated jets, but to a lesser degree. Therefore, gluons emitted with p_{T} between a given $p_{\text{T,jet}}$ range of interest will often translate, after jet-reconstruction, into jets with $p_{\text{T,jet}}$ below that range. It is true that gluons with p_{T} above that given $p_{\text{T,jet}}$ range of interest may translate into jets that fall inside it, but the power law decay of the particle spectrum makes it much more unlikely than the previous scenario. Therefore, cuts in $p_{\text{T,jet}}$ will bias jet-reconstruction towards light-quarks, underestimating gluon-initiation - the survival bias. Since higher $\sqrt{s}$ and lower $p_{\text{T,jet}}$ favours the selection on hard gluons, these are also the kinematics for whom survival bias is the most relevant.

$\sqrt{s}$ [GeV]	$p_{\mathrm{T,jet}}$ [GeV/c]	Bjorken x
200	$15 < p_{\mathrm{T,jet}} < 20$	$0.150 - 0.200$
	$30 < p_{\mathrm{T,jet}} < 40$	$0.300 - 0.400$
	$20 < p_{\mathrm{T,jet}} < 40$	$0.200 - 0.400$
5000	$15 < p_{\mathrm{T,jet}} < 20$	$0.006 - 0.008$
	$30 < p_{\mathrm{T,jet}} < 40$	$0.012 - 0.016$
	$200 < p_{\mathrm{T,jet}} < 300$	$0.080 - 0.120$
14000	$15 < p_{\mathrm{T,jet}} < 20$	$0.002 - 0.003$
	$30 < p_{\mathrm{T,jet}} < 40$	$0.004 - 0.006$
	$200 < p_{\mathrm{T,jet}} < 300$	$0.029 - 0.043$

Table 3.2: Bjorken x values probed for all kinematic settings imposed to the pp collisions studied in this thesis, with varying jet transverse momenta $p_{\mathrm{T,\,jet}}$ and CM energies $\sqrt{s}$.

To restrict this Section's substructure studies to the perturbative regime of QCD, the substructure variables are calculated at the 1$^{\mathrm{st}}$ soft-drop emission (1SD), already introduced in Section 2.3.2, corresponding to the first de-clustering step of the tree that satisfies the SD criterion. It physically translates into the first hard pQCD-like emission within that jet (hardness scale set by $z_{\mathrm{cut}} = 0.1$). The 1SD emission for the hypothetical branching scheme of a jet produced in a hard scattering event is shown in Figure 3.3.1, where the parton branchings are represented by the black Feynman QCD vertices, except for the 1SD emission, which appears as the highlighted magenta splitting (assumed to satisfy the SD criterion).

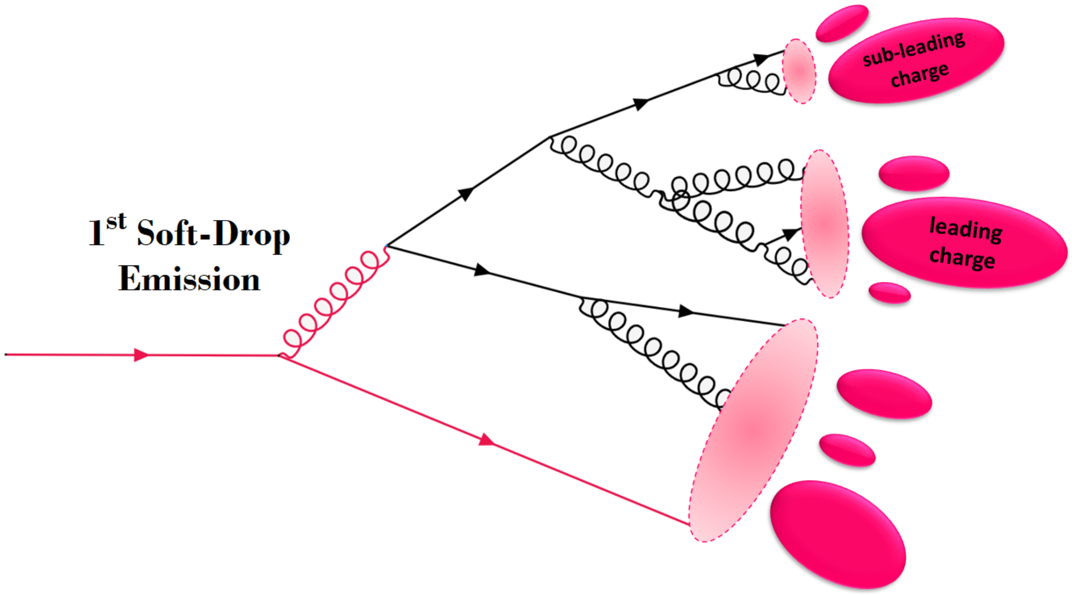


Figure 3.3.1: First soft-drop emission (1SD), represented as the highlighted magenta parton splitting in the beginning of the hypothetical branching scheme of a jet produced in a hard scattering event; arrowed lines represent quarks, spiral lines gluons, translucent magenta ellipses the hadronization process and opaque magenta ellipses the outgoing hadrons.

3.3.1 Groomed Jet Radius

One of the relevant substructure variables is the groomed jet radius R_g, introduced in Equation (2.22). It is defined as the angular distance in the $y - \phi$ plane between the 2 subjets found after grooming the jet with the SD algorithm. Therefore, R_g consists of the opening angle of 1SD emission, represented in Figure 3.3.1, and it quantifies a jet's degree of collimation. In the *FastJet* framework, R_g is evaluated using the *delta_R()* method between the 2 daughter-prongs of the 1SD emission and is from heron out referred to as the jet opening angle.

Figure 3.3.2 shows, on the left panel, the R_g distributions for C/A re-clustered jets with transverse momentum falling on the $15 < p_{\mathrm{T,jet}} < 20$ GeV/c interval, produced in pp collisions with CM energies of $\sqrt{s} = 0.2, 5, 14$ TeV, represented in red circle, blue square and black diamond markers, respectively. On the right panel, the same distributions are shown for jets re-clustered using the τ algorithm.

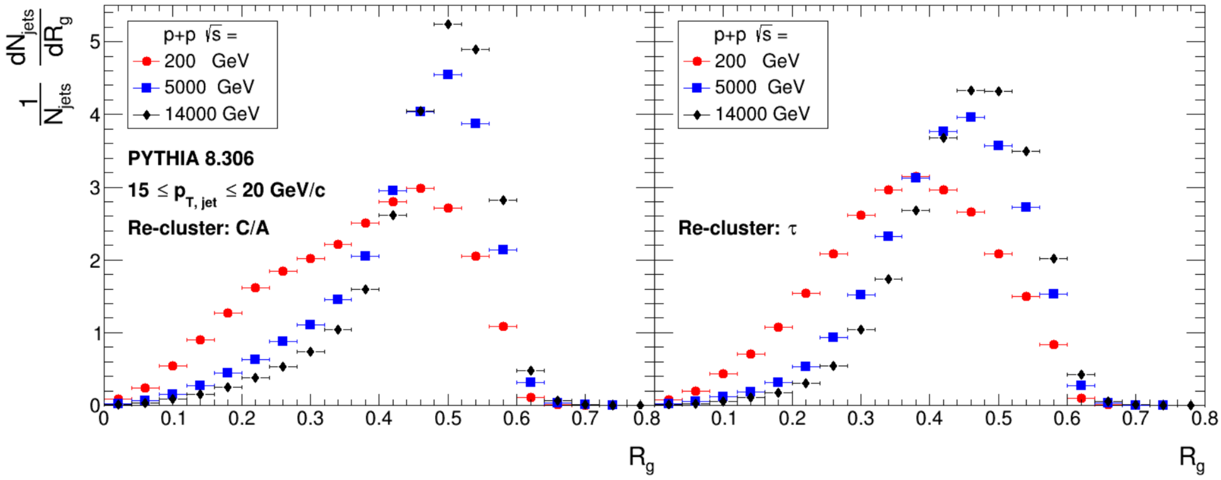


Figure 3.3.2: Groomed jet radius distributions for jets with $15 < p_{\mathrm{T,jet}} < 20$ GeV/c, produced in pp collisions with $\sqrt{s} = 0.2, 5, 14$ TeV, represented in red circle, blue square and black diamond markers, respectively; the jets are re-clustered using the C/A algorithm for the left panel distributions and the τ algorithm for the right panel ones.

With the jet population used for plotting the results in Fig. 3.3.2 restricted in transverse momentum, the distribution variations can be understood in light of the different collision energies. They show for both re-clustering algorithms that lower CM energy collisions produce, on average, more highly-collimated jets, with overall smaller jet openings. Higher CM energies, on the other hand, tend produce thicker jets. These observations, combined with the quark-gluon fraction discussion presented in the beginning of Section 3.3, reveal that gluon-initiated jet populations, more abundant in pp collisions of higher CM energy, have, on average, larger jet radii than the quark-initiated ones. The overwhelmingly gluon-dominated blue and black R_g distributions are even shown to peak extremely close to the jet radius $R = 0.6$ imposed in the anti-k_t clustering. This shows that gluons radiate in such wider angles that a significant part of gluonic radiation falls outside the boundaries of the identified jets. However, since R_g distributions still peak for $R_g < R$, most of the radiation is nevertheless being included in the jet.

The shape differences between C/A and τ distributions of identical $\sqrt{s}$ come from the different clustering trees that these 2 algorithms produce. C/A and τ trees cluster the particles differently between the 2 subjets resolved by the 1SD emission. This leads, in general, to different groups of particles being included in each one of these subjets and, consequently, to 1SD emissions with different kinematics for each re-clustering procedure. As was discussed in Section 3.2, C/A re-clustering is the most geometrically sensitive one, purposefully constructed for substructure studies such as R_g analysis.

To further investigate the observed thickening of the jets due to the increase of the CM energy, the same plots displayed in Figure 3.3.2 are now presented in Figure 3.3.3 with the same marker scheme. However, the new R_g distributions sample jets from higher transverse momenta, in the $30 < p_{\mathrm{T,jet}} < 40$ GeV/c range.

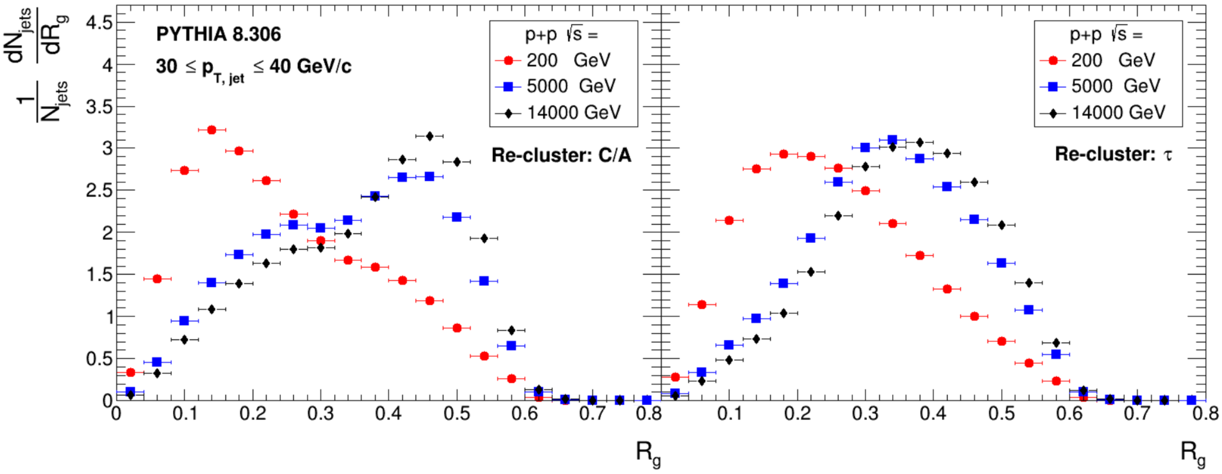


Figure 3.3.3: Groomed jet radius distributions for jets with $30 < p_{\mathrm{T,jet}} < 40$ GeV/c, produced in pp collisions with $\sqrt{s} = 0.2, 5, 14$ TeV, represented in red circle, blue square and black diamond markers, respectively; the jets are re-clustered using the C/A algorithm for the left panel distributions and the τ algorithm for the right panel ones.

The same observations made for Figure 3.3.2 about the left-shift in R_g when decreasing the collision CM energy can also be observed in Figure 3.3.3. However, the latest plots show even greater differences between the R_g distributions for different $\sqrt{s}$. Since the jet populations sampled in these distributions are chosen from a higher $p_{\mathrm{T,jet}}$ range, the quark-gluon fraction is much more sensitive to the CM energy (check Table 3.2). This is why the $\sqrt{s} = 200$ GeV collisions have a predominance of quark-initiated jets and, therefore, produce significantly narrower jets than the other 2 $\sqrt{s}$ settings for this new $p_{\mathrm{T,jet}}$ region. However, the $\sqrt{s} = 5$ TeV and $\sqrt{s} = 14$ TeV distributions show notable similarities between each other for both $p_{\mathrm{T,jet}}$ ranges due to an overwhelming predominance of gluons in both instances.

However, many of the C/A distributions show bumps that suggest a superposition of the contributions from quark- and gluon-initiated jet populations, weighted by their fraction. For instance, the red circle distribution on the left panel of Figure 3.3.2 peaks for large R_g, but shows a bump for smaller jet openings that indicates a greater parity between quark- and gluon-initiated jets (tilting in favour of the gluon-initiated ones), when compared to its blue square and black diamond counterparts. However, for the corresponding red circle distribution on the left panel of Figure 3.3.3, the quark/gluon fraction flips to quark predominance, with the peak occurring for small R_g and the bump produced by gluon-initiated jets appearing for larger openings. The $\sqrt{s} = 5$ TeV and $\sqrt{s} = 14$ TeV distributions on the left panel of Figure

3.3.3, according to the x estimates on Table 3.2, should be overwhelmingly gluon-dominated. However, they show bumps before the peaks, tending towards quark-initiated features, signalling that a shift to light quarks is perhaps being caused by survival bias.

To dig deeper into the $p_{\mathrm{T,jet}}$ influence on jet opening angles, Figure 3.3.4 shows the R_g distributions for $\sqrt{s} = 5$ TeV pp collisions, using jets with transverse momentum falling on the $15 < p_{\mathrm{T,jet}} < 20$ GeV/c, $30 < p_{\mathrm{T,jet}} < 40$ GeV/c and $200 < p_{\mathrm{T,jet}} < 300$ GeV/c ranges, represented in red circle, blue square and black diamond markers, respectively.

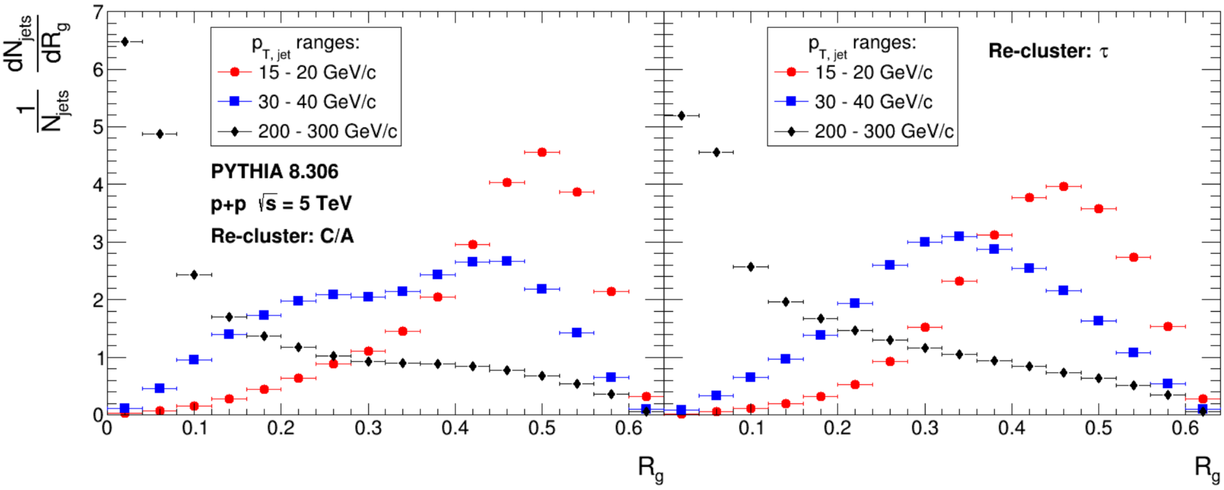


Figure 3.3.4: Groomed jet radius distributions for jets with $15 < p_{\mathrm{T,jet}} < 20$ GeV/c, $30 < p_{\mathrm{T,jet}} < 40$ GeV/c and $200 < p_{\mathrm{T,jet}} < 300$ GeV/c, produced in pp collisions with $\sqrt{s} = 5$ TeV, represented in red circle, blue square and black diamond markers, respectively; the jets are re-clustered using the C/A algorithm for the left panel distributions and the τ algorithm for the right panel ones.

Figure 3.3.4 reveals that low transverse momentum jets from $\sqrt{s} = 5$ TeV pp collisions are wide, with the distributions for $15 < p_{\mathrm{T,jet}} < 20$ GeV/c jets peaking for high R_g, while high transverse momenta jets are narrow, with the distributions for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets peaking for small R_g. For the jets in the intermediate transverse momentum region of $30 < p_{\mathrm{T,jet}} < 40$ GeV/c, C/A re-clustering reveals a relatively uniform distribution for a wide range of jet openings. These observations further corroborate that gluon-initiated jets are wide and quark-initiated ones are narrow.

The $R = 0.6$ imposition on the anti-k_t clustering allows emissions to have angular openings up to $R_g \lesssim 2R$. However, this is shown to happen very rarely. Approximately 0.5%, 1% and 2% of the $15 < p_{\mathrm{T,jet}} < 20$ GeV/c jets from $\sqrt{s} = 0.2, 5, 14$ TeV collisions, respectively, have (for both re-clustering algorithms) groomed jet radii R_g larger than $R = 0.6$, but confined to the $0.6 \leq R_g < 0.8$ range (shown in Figure 3.3.2). The same happens for approximately 0.2%, 0.5% and 1% of the $30 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 0.2, 5, 14$ TeV collisions, respectively (shown in Figure 3.3.3).

3.3.2 Groomed Jet Momentum Fraction

Another important variable regarding perturbative substructure features is the groomed jet momentum fraction z_g, also defined in Equation (2.22), evaluated at the 1SD emission. It corresponds to the fraction of the total transverse momentum that is carried by the softest daughter-subjet from a $1 \rightarrow 2$

splitting. The SD criterion presented in Equation (2.21) limits the momentum fraction of any SD emission from below, such that $z_g > z_{\text{cut}} = 0.1$. It is also maximally limited by the equal-sharing scenario, meaning $0.1 < z_g < 0.5$.

Figure 3.3.5 shows, on the left top panel, the z_g distributions for C/A re-clustered jets with $15 < p_{\text{T,jet}} < 20$ GeV/c and produced in $\sqrt{s} = 0.2, 5, 14$ TeV pp collisions, represented in red circle, blue square and black diamond markers, respectively. The right top panel shows the same z_g distributions for jets re-clustered using the τ algorithm and the bottom panels for $30 < p_{\text{T,jet}} < 40$ GeV/c jets. The y–range is set to be identical for the 4 panels, so to facilitate the comparisons between $p_{\text{T,jet}}$ intervals and re-clustering algorithms.

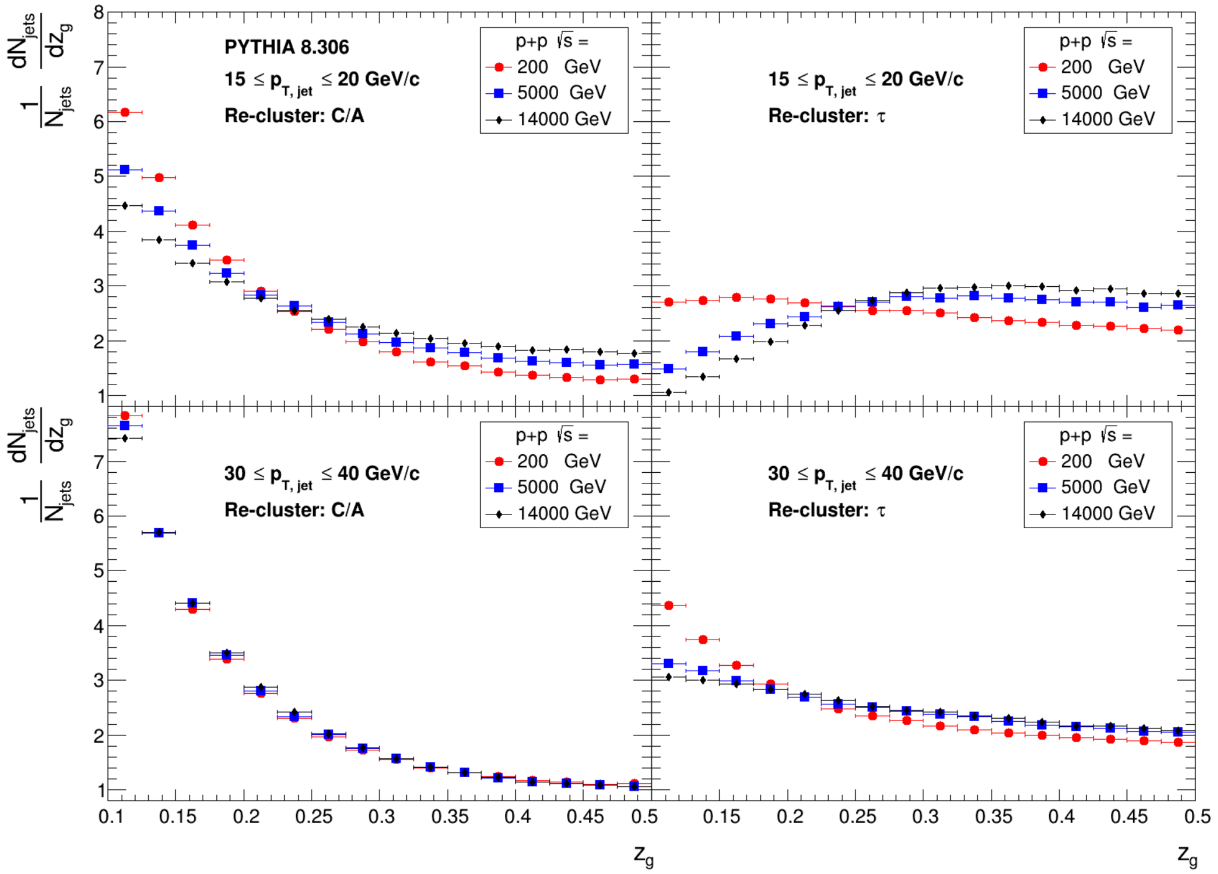


Figure 3.3.5: Soft-drop (SD) groomed jet momentum fraction distributions (at the 1$^{\text{st}}$ SD emission) for jets with $15 < p_{\text{T,jet}} < 20$ GeV/c, on the top panels, and for jets with $30 < p_{\text{T,jet}} < 40$ GeV/c, on the bottom panels, produced in pp collisions with $\sqrt{s} = 0.2, 5, 14$ TeV, represented in red circle, blue square and black diamond markers, respectively; the left plots used jets submitted to C/A re-clustering, while the ones on the right side used τ re-clustered ones.

For the top panels of Figure 3.3.5, sampling jets of lower $p_{\text{T,jet}}$, there are clear differences as the $\sqrt{s}$ is varied. They indicate that increasing the CM energy produces more symmetric energy shares between the jet's main and secondary branches, since they become successively less peaked at small z_g with the increase of $\sqrt{s}$, leading to flatter distributions. For the bottom panels, with the distributions for higher $p_{\text{T,jet}}$, the differences are less stark but still follow the same trends found for lower $p_{\text{T,jet}}$. Furthermore, for fixed $\sqrt{s}$, higher $p_{\text{T,jet}}$ jets are shown to have significantly more asymmetrical 1SD emissions than the

the ones found for lower $p_{\mathrm{T,jet}}$ since the latter z_g distributions are significantly less shifted to small z_g than the first ones.

In light of what is discussed in the beginning of Section 3.3 with regard to the different scales at which the proton is resolved in these pp collisions, one can interpret these results as indicating that gluon-initiated jets are, on average, more symmetric in transverse momentum space than its quark-initiated counterparts, in agreement with insight provided by the DGLAP splitting functions in Equations 2.12. The quark splitting function, as well as the $g \to gg$ one, favour asymmetric emissions (with smaller z_g). However, the other term for the gluon-splitting, introduced by the intrinsically more symmetric $g \to q\bar{q}$ channel, is consequently flatter than the previous two, making the gluon splitting functions overall flatter than the quark ones, as is clear from the results in Figure 3.3.5.

The flattening of the z_g distributions with the increase of $\sqrt{s}$ occurs for both re-clustering algorithms. However, the τ re-clustered jets lead to z_g distributions that vary significantly from the shapes of the theoretical DGLAP splitting functions. This thesis is, in fact, the first time τ re-clustering is being used in z_g analysis for low $p_{\mathrm{T,jet}}$. The z_g, shown in Chapter 4 to be highly sensitive to the transition from the pQCD to the npQCD regimes, hints (when evaluated using τ re-clustering) that low $p_{\mathrm{T,jet}}$ at RHIC means the 1SD emission is already taking place at the hadronic level.

However, these conclusions are not as straightforward as they are for the groomed jet radius. In fact, the $30 < p_{\mathrm{T,jet}} < 40$ GeV/c range brought the biggest differences for the R_g distributions, presented in Figure 3.3.3, but it also seems to bring the biggest similarities for the z_g ones. The influence of $p_{\mathrm{T,jet}}$ on the z_g distributions is studied more carefully in the results from Figure 3.3.6, which shows the z_g distributions for $\sqrt{s} = 5$ TeV pp collisions, sampling jets with $15 < p_{\mathrm{T,jet}} < 20$ GeV/c, $30 < p_{\mathrm{T,jet}} < 40$ GeV/c and $200 < p_{\mathrm{T,jet}} < 300$ GeV/c, represented in red circle, blue square and black diamond markers, respectively.

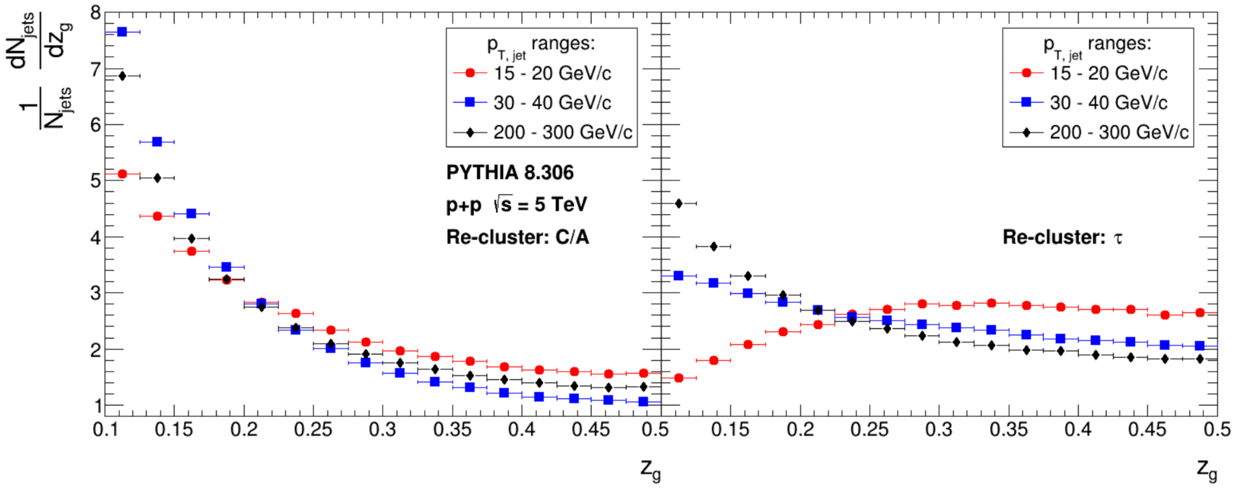


Figure 3.3.6: Groomed jet momentum fraction distributions for jets with $15 < p_{\mathrm{T,jet}} < 20$ GeV/c, $30 < p_{\mathrm{T,jet}} < 40$ GeV/c and $200 < p_{\mathrm{T,jet}} < 300$ GeV/c, produced in pp collisions with $\sqrt{s} = 5$ TeV, represented in red circle, blue square and black diamond markers, respectively; the jets are re-clustered using the C/A algorithm for the left panel distributions and the τ algorithm for the right panel ones.

Figure 3.3.6 reinforces, for both re-clustering algorithms, that the z_g distributions are tendentiously flatter (1SD emission is more symmetric) for kinematic settings that favour gluon jet-initiation (lower $p_{\mathrm{T,jet}}$). The only case where this is not happening is for the C/A distributions of $30 < p_{\mathrm{T,jet}} < 40$ GeV/c and of $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets. The latter should be, according to the x estimates on Table 3.2, more peaked for low z_g than the previous one. However, the R_g analysis on Section 3.3.1 already showed survival bias to sway jet-reconstruction towards light quarks and, therefore, more asymmetric 1SD emissions in the $30 < p_{\mathrm{T,jet}} < 40$ GeV/c range, which could be at the root of this discrepancy.

3.3.3 Groomed Jet Formation Time

Finally, the groomed jet formation time τ_{form} is computed for the 1SD emission using Equation (2.24), which factors in the PYTHIA's kinematic information of the two daughter-subjets, J_1 assumed to be the hardest and J_2 the softest. In fact, the total energy E in Equation (2.24) is calculated through $E = E_1 + E_2$ and the energy fraction from Equation (2.25) becomes

$$z = \frac{E_2}{E_1 + E_2} \quad . \tag{3.3}$$

Furthermore, the $\cos\theta_{12}$ variable is calculated through the canonical inner product between the 3-momenta from J_1 and J_2, such that

$$\cos\theta_{12} = \frac{p_{1,x}p_{2,x} + p_{1,y}p_{2,y} + p_{1,z}p_{2,z}}{\sqrt{p_{1,x}^2 + p_{1,y}^2 + p_{1,z}^2}\sqrt{p_{2,x}^2 + p_{2,y}^2 + p_{2,z}^2}} \quad . \tag{3.4}$$

The τ_{form} distributions, in units of femtometre divided by the vacuum speed of light (fm/c), are shown in Figure 3.3.7. They sample jets with transverse momentum in the $15 < p_{\mathrm{T,jet}} < 20$ GeV/c range, produced in pp collisions with CM energies of $\sqrt{s} = 0.2, 5, 14$ TeV. These are represented in red circle, blue square and black diamond markers, respectively. The left panel has the distributions for C/A re-clustered jets and the right panel for τ re-clustered ones. The x- and y-axis are shown in logarithmic scale to improve readability across multiple orders of magnitude. The choice of differently sized bins in the histogram's construction is accounted for by the division of the bin contents by the respective bin widths.

The observation of Figure 3.3.7 immediately leads to the conclusion that an increase in the CM energy of the pp collision leads to smaller 1SD formation times, on average. This is evident from the left shift that takes place when one goes from the red to the blue distribution, and then from the blue to the black one. To put it into quantitative terms, the average $\tau_{\mathrm{form,\,1SD}}$ values (μ) for each one of these distributions are presented in Table 3.3 (first column of values). They clearly show that the 1SD emission happens earlier, on average, for higher $\sqrt{s}$, meaning the jet-initiating partons become more eager to branch. The decrease of the $\tau_{\mathrm{form,\,1SD}}$ with respect to the increase in $\sqrt{s}$ is verified, both graphically in Figure 3.3.7 and numerically in Table 3.3, for both re-clustering algorithms.

Therefore, similarly to the parton-species analysis performed in Sections 3.3.1 and 3.3.2, one concludes from these new observations that gluons radiate more than quarks, generally branching into 2 subjets earlier than quarks.

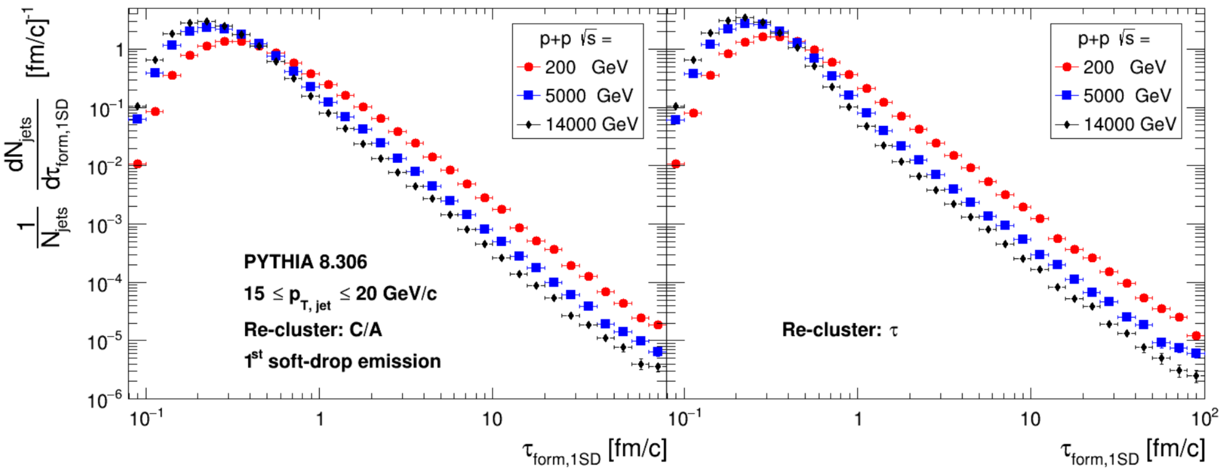


Figure 3.3.7: Formation time distributions for the 1st soft-drop emission of jets with $15 < p_{\text{T,jet}} < 20$ GeV/c, produced in pp collisions with $\sqrt{s} = 0.2, 5, 14$ TeV, represented in red circle, blue square and black diamond markers, respectively; the jets are re-clustered using the C/A algorithm for the left panel and the τ algorithm for the right panel.

Table 3.3 also shows the average $\tau_{\text{form, 1SD}}$ for jets sampled with $30 < p_{\text{T,jet}} < 40$ GeV/c. Since the $30 < p_{\text{T,jet}} < 40$ GeV/c range is higher, the jet population is more quark-initiated, leading one to expect higher average 1SD formation times when compared to $15 < p_{\text{T,jet}} < 20$ GeV/c jets. However, the comparison between the first and second columns from Table 3.3 reveals this expected increase to be happening only for $\sqrt{s} = 200$ GeV/c pp collisions and when C/A re-clustering is performed. For all other $\sqrt{s}$ and re-clustering procedures, the opposite is taking place. To further test the $p_{\text{T,jet}}$ dependency of these observations, Figure 3.3.8 shows the $\tau_{\text{form, 1SD}}$ distributions for $\sqrt{s} = 5$ TeV pp collisions, sampling jets with $15 < p_{\text{T,jet}} < 20$ GeV/c, $30 < p_{\text{T,jet}} < 40$ GeV/c and $200 < p_{\text{T,jet}} < 300$ GeV/c, represented in red circle, blue square and black diamond markers, respectively.

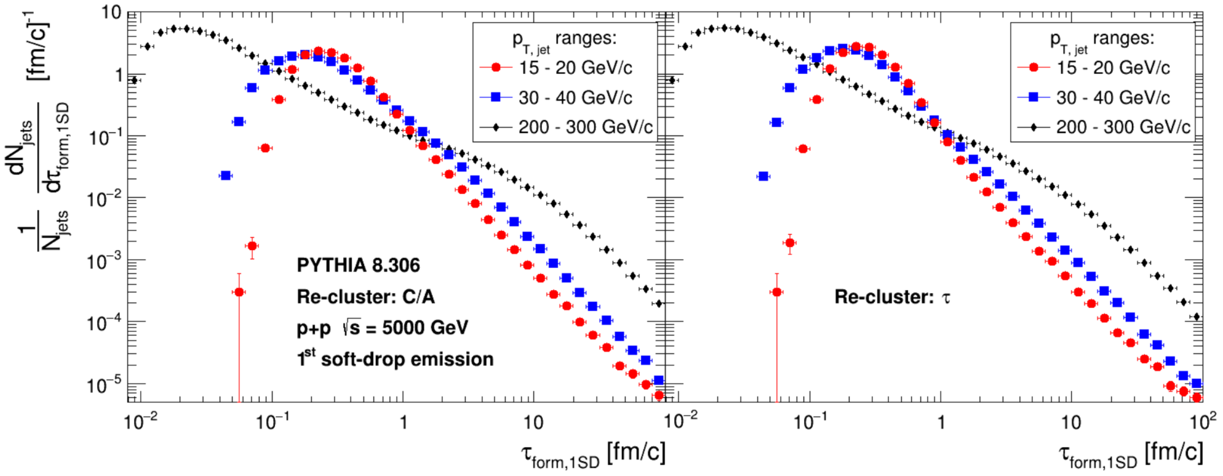


Figure 3.3.8: Formation time distributions for the 1st soft-drop emission of jets with $15 < p_{\text{T,jet}} < 20$ GeV/c, $30 < p_{\text{T,jet}} < 40$ GeV/c and $200 < p_{\text{T,jet}} < 300$ GeV/c, produced in pp collisions with $\sqrt{s} = 5$ TeV, represented in red circle, blue square and black diamond markers, respectively; the jets are re-clustered using the C/A algorithm for the left panel distributions and the τ algorithm for the right panel ones.

Figure 3.3.8 reinforces the previous observation by showing, for $\sqrt{s} = 5$ TeV pp collisions, that jets with higher $p_{\text{T,jet}}$ have earlier 1SD emissions, against expectations based solely on proton x resolution scales. This is confirmed by the average $\tau_{\text{form, 1SD}}$ presented in Table 3.3 for both of its $\sqrt{s} = 5$ TeV rows. Therefore, comparisons of the formation time between different $p_{\text{T,jet}}$ ranges cannot be simply done in terms of the jet-initiating parton-species.

Re-cluster	$\sqrt{s}$ [GeV]	μ [fm/c] for $p_{\text{T,jet}}$ ranges [GeV/c]		
		$15 < p_{\text{T,jet}} < 20$	$30 < p_{\text{T,jet}} < 40$	$200 < p_{\text{T,jet}} < 300$
C/A	200	0.500 ± 0.002	0.521 ± 0.002	-
τ	200	0.454 ± 0.001	0.442 ± 0.001	-
C/A	5000	0.3377 ± 0.0008	0.301 ± 0.001	0.101 ± 0.002
τ	5000	0.3131 ± 0.0006	0.2630 ± 0.0007	0.093 ± 0.001
C/A	14000	0.2912 ± 0.0005	0.2608 ± 0.0008	0.090 ± 0.001
τ	14000	0.2729 ± 0.0004	0.2309 ± 0.0006	0.083 ± 0.001

Table 3.3: Average values μ of the 1^{st} soft-drop emission formation time distributions for jets with $15 < p_{\text{T,jet}} < 20$ GeV/c and $30 < p_{\text{T,jet}} < 40$ GeV/c produced in pp collisions of $\sqrt{s} = 0.2, 14$ TeV center-of-mass energies and for jets with $15 < p_{\text{T,jet}} < 20$ GeV/c, $30 < p_{\text{T,jet}} < 40$ GeV/c and $200 < p_{\text{T,jet}} < 300$ GeV/c from $\sqrt{s} = 5, 14$ TeV events; for each $\sqrt{s}$ value, the mean values are presented for re-clustering performed with both the C/A and the τ algorithms.

3.4 Evolution along the Clustering Tree

Alongside the substructure features at the early jet development, it is also relevant to investigate how substructure evolves along the clustering tree. To do that, the study is restricted to only 2 kinematic setups: $\sqrt{s} = 200$ GeV pp collisions, selecting jets in the $20 < p_{\text{T,jet}} < 40$ GeV/c range and $\sqrt{s} = 5$ TeV pp collisions, for jets in the $200 < p_{\text{T,jet}} < 300$ GeV/c interval. These settings are chosen in order to enable future comparisons to experimental data from the RHIC and LHC colliders, respectively. The N_{splits} distributions shown in Figure 3.2.2 reveal that both of these kinematic restraints produce branching sequences with an average $N_{\text{splits}} \sim 4$ (post-SD-grooming), sufficing to follow the first 3 soft-drop emissions along the main jet branch to get a reasonable picture of the substructure evolution along the clustering tree: the 1^{st} soft-drop emission (1SD), the 2^{nd} soft-drop emission (2SD) and the 3^{rd} soft-drop emission (3SD). For each successive branching, the energy per parton decreases. This allows to check the substructure dependency on energy scales, namely on how it evolves when approaching hadronization.

Jet substructure is probed in this Section using the opening angle ΔR, the momentum fraction z_g and the formation time τ_{form} of these 3 emissions. The 1SD distributions sample all jets with at least 1 soft-drop emission, the 2SD ones all jets with at least 2 soft-drop emissions and the 3SD ones all jets with at least 3 soft-drop emissions.

Figure 3.4.1 shows the ΔR distributions (see Equation (2.17)) for the first 3 soft-drop emissions along the main jet branch, represented in red circle, blue square and black diamond markers for the 1SD, 2SD and 3SD emissions, respectively. The top panels show the ΔR distributions for jets with transverse momentum falling on the $20 < p_{\mathrm{T,jet}} < 40$ GeV/c range from $\sqrt{s} = 200$ GeV pp collisions and the bottom panels the distributions for jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c from $\sqrt{s} = 5$ TeV pp collisions. The distributions on the left panels sample C/A re-clustered jets and the right panels the τ re-clustered ones.

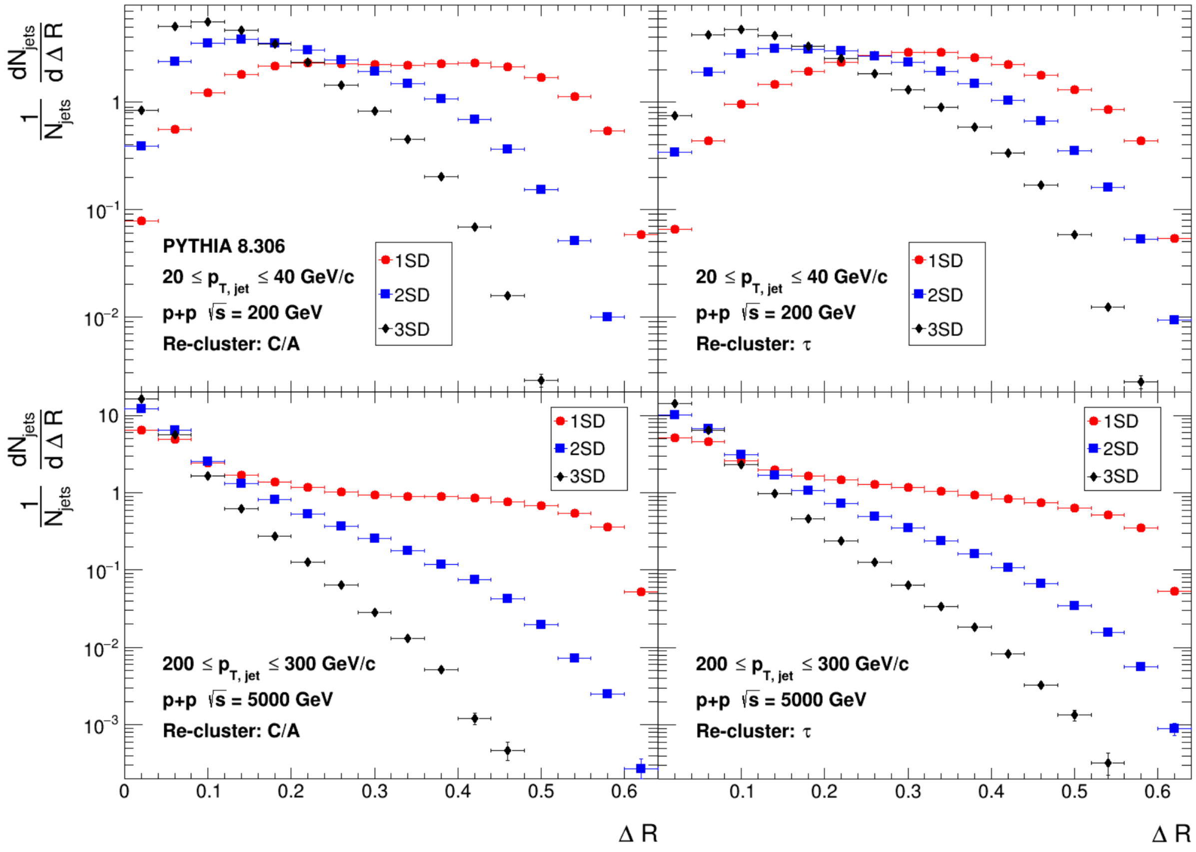


Figure 3.4.1: Opening angle distributions for the first three soft-drop emissions along the main branch of $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from 200 GeV pp collisions, on the top panels, and of $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from 5 TeV pp collisions, on the bottom panels; the 1SD, 2SD and 3SD ΔR distributions are represented in red circle, blue square and black diamond markers, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

All panels from Figure 3.4.1 reveal a narrowing of the SD emissions along the clustering tree. In fact, from the 1SD emission to the 2SD emission and then from the 2SD to the 3SD one, the ΔR distributions are shifted to significantly smaller opening angles. This happens for both kinematic settings imposed on the sampled jets, with the $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from 5 TeV pp collisions being significantly narrower for each of the first 3 soft-drop emissions. However, it shows that even though low $p_{\mathrm{T,jet}}$ jets start with wide angle splittings, angular ordering dictates following SD emissions. Results are consistent across both re-clustering algorithms.

When it comes to the evolution of splitting functions along the clustering tree, Figure 3.4.2 clearly shows that soft-drop groomed, C/A re-clustered jets have progressively flatter z_g distributions along the clustering tree, with each soft-drop emission following the main jet branch getting more symmetrical than the one before. For the re-clustering performed with the τ algorithm, conceived around the emission formation time, the evolution of splitting functions throughout the tree is less sequential, with the 1SD

z_g distribution contradicting this trend. However, the flattening of the z_g distributions along the clustering tree is still observed for SD emissions posterior to the 1SD for the τ re-clustered jets. Therefore, lower energy scales lead, on average, to more symmetrical emissions.

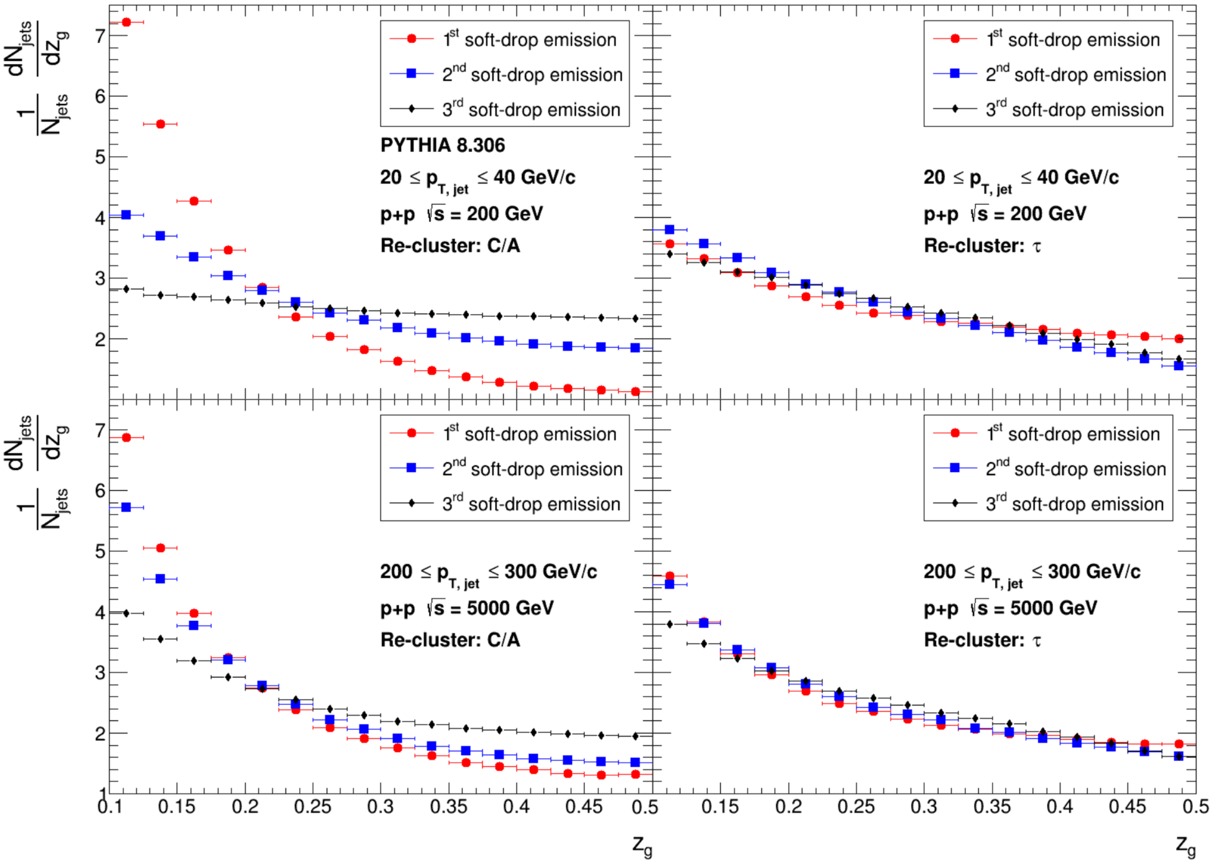


Figure 3.4.2: Momentum fraction distributions for the first three soft-drop emissions along the main branch of $20 < p_{\text{T,jet}} < 40$ GeV/c jets from 200 GeV pp collisions, on the top panels, and of $200 < p_{\text{T,jet}} < 300$ GeV/c jets from 5 TeV pp collisions, on the bottom panels; the 1SD, 2SD and 3SD z_g distributions are represented in red circle, blue square and black diamond markers, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

Finally, it is important to check how the emission formation time evolves when energy scales are reduced towards the hadronizaiton ones. Therefore, Figure 3.4.3 shows the τ_{form} distributions for the first 3 soft-drop emissions along the main jet branch, represented in red circle, blue square and black diamond markers for the 1SD, 2SD and 3SD emissions, respectively, obtained by sampling $20 < p_{\text{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions. The same τ_{form} distributions are shown in Figure 3.4.4 for jets produced in 5 TeV pp collisions and with $200 < p_{\text{T,jet}} < 300$ GeV/c.

Both Figures 3.4.3 and 3.4.4 reveal that SD emissions take more time to happen along the clustering tree, with 3SD τ_{form} being significantly higher than 2SD ones and both of these higher than 1SD τ_{form}. Therefore, the smaller energy scales at which later SD emissions take place determine, in general, higher formation time for those splittings, meaning the partons become less eager to emit along the clustering trees. This is verified for both re-clustering algorithms and for both sets of kinematic conditions.

When it comes to the choice of the jet algorithm used for the re-clustering of the anti-k_t jets, Section

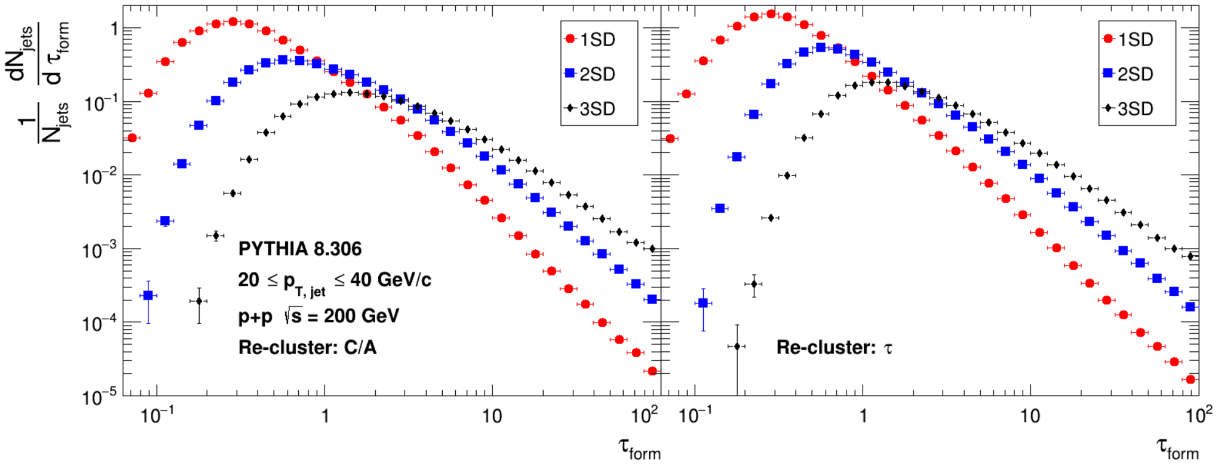


Figure 3.4.3: Formation time distributions for the first three soft-drop emissions along the main branch of $20 < p_{\text{T,jet}} < 40$ GeV/c jets from 200 GeV pp collisions, with the 1SD, 2SD and 3SD τ_{form} distributions represented in red circle, blue square and black diamond markers, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

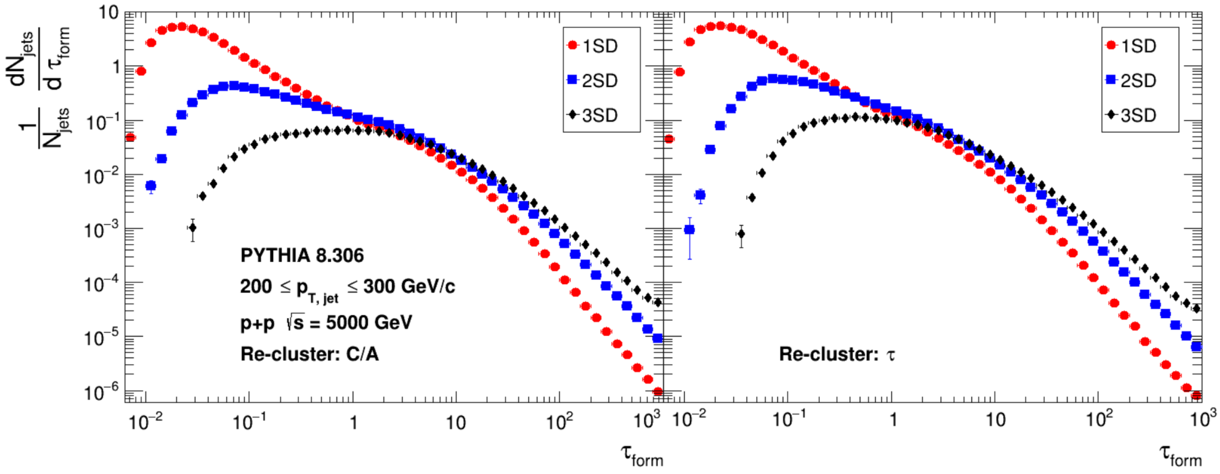


Figure 3.4.4: Formation time distributions for the first three soft-drop emissions along the main branch of $200 < p_{\text{T,jet}} < 300$ GeV/c jets from 5 TeV pp collisions, with the 1SD, 2SD and 3SD τ_{form} distributions represented in red circle, blue square and black diamond markers, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

3.2 establishes the (not strictly) time-ordered τ algorithm as the most reliable one to provide a formation time picture of the jet splittings. The logarithmic scales of the y–axis somewhat hide the differences between C/A and τ formation time distributions, in the name of improving readability for a wide range of τ_{form}. However, Table 3.3 already reveals the significance of these differences. It shows, for instance, that the C/A algorithm overestimates the average $\tau_{\text{form, 1SD}}$ of $15 < p_{\text{T,jet}} < 30$ GeV/c jets by approximately 10.1%, 7.9% and 6.7% for pp collisions of $\sqrt{s} = 0.2, 5, 14$ TeV, respectively.

Therefore, for a more in-depth study of the algorithm-influence on the results, it makes sense to look at the distributions of the bin-by-bin ratio between the τ and the C/A τ_{form} distributions for each SD emission. These ratios are shown in Figure 3.4.5 for the τ_{form} distribution obtained with τ re-clustering with respect to the τ_{form} distribution obtained with C/A re-clustering and for each of the first 3 soft-drop emissions. The top panels show these ratios in red circle markers for the 1SD emission, the middle panels in blue square markers for the 2SD and the bottom ones in black diamond markers for the 3SD. On the left panels, the ratios are represented for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV/c pp collisions, while the right panels show them for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions. A dashed line is drawn for the unitary ratio.

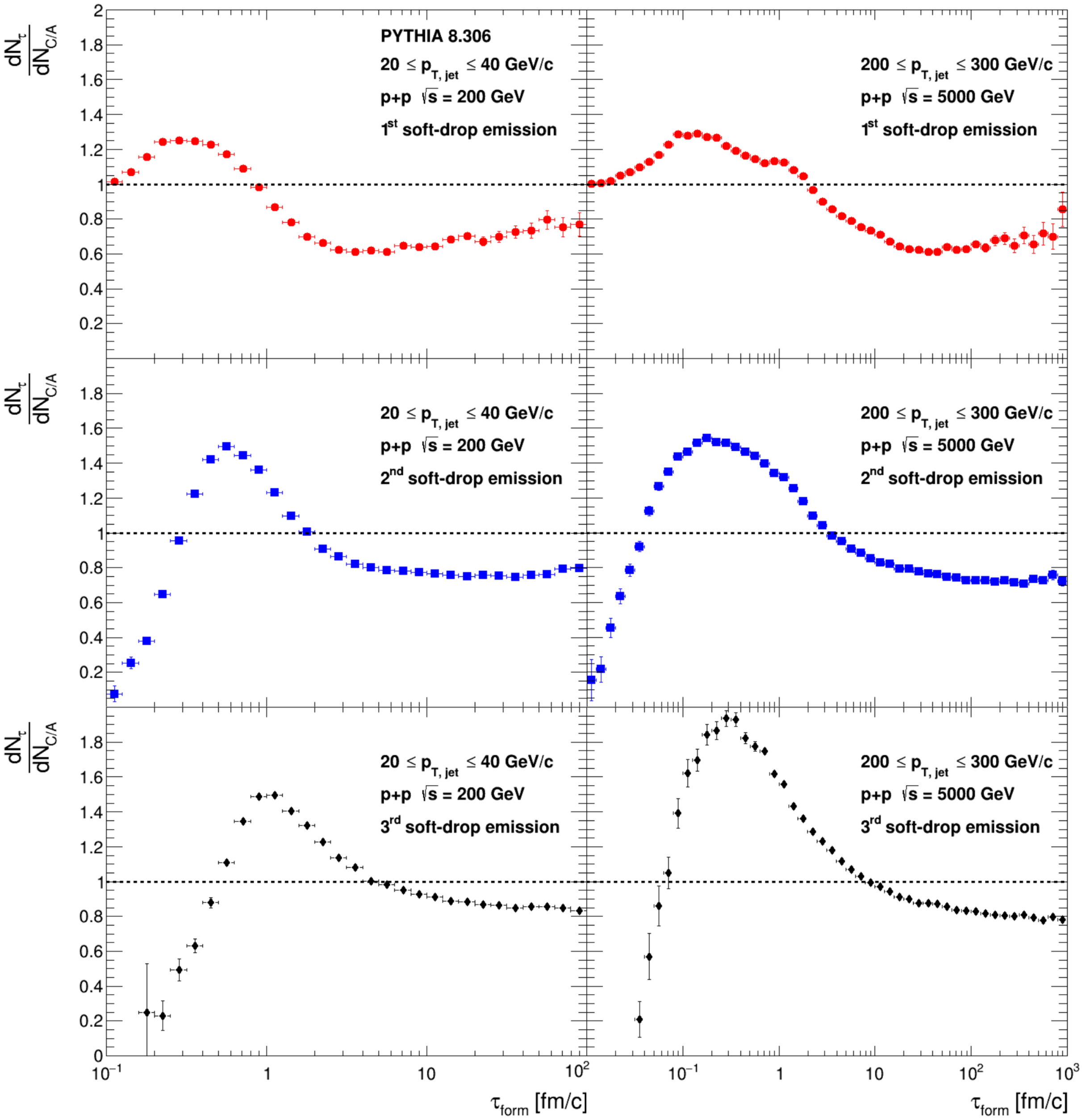


Figure 3.4.5: Bin-by-bin ratios of the formation time distributions obtained for τ re-clustered jets with respect to the ones obtained from C/A re-clustered jets, evaluated at the 1st, 2nd and 1rd soft-drop emissions for top, middle and bottom plots, respectively (in red circle, blue square and black diamond markers, respectively too); the left panels show the distributions for jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c from 200 GeV pp collisions and the right panels for jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c produced in 5 TeV pp collisions.

Figure 3.4.5 shows that the re-clustering performed using the τ algorithm produces clustering trees with not only earlier 1SD emissions, as is shown in Table 3.3, but also the subsequent 2SD and 3SD emissions along the main jet branch. This is extrapolated from the fact that $N_\tau/N_{C/A}$ ratios are generally larger than unity for small τ_{form}, becoming smaller than unity for large τ_{form}. Comparisons on the basis of re-clustering algorithm are allowed by the fact that the same jets are are sampled in both cases, varying only the clustering trees.

Therefore, τ re-clustered trees tendentiously branch earlier than C/A re-clustered ones. Furthermore, the C/A overestimation of the formation time values becomes larger along the branching history. Variations for 1SD distributions between the 2 algorithms go up to approximately $+25\%$ for smaller τ_{form} and -40% for larger τ_{form}, with the following SD emissions verifying even greater variations between the algorithms. This means C/A becomes progressively less sensitive to τ_{form} along the clustering trees.

Chapter 4

Jet Substructure: Towards Non-Perturbative Physics

This Chapter advances the study of jet substructure towards non-perturbative hadronization scales. In order to do that, 2 new splittings of interest are added as objects of analysis to the extensively studied SD emissions from Chapter 3.

One of these new splittings is a construct assuming the 2 final-state charged particles with the highest transverse momentum from a jet, called the leading and sub-leading charged particles, are produced in a hadronic splitting. This is called the leading charged particles splitting (LCP). The LCP splitting takes the 2 leading charged particles, instead of 2 leading particles (regardless of electric charge), because of its experimental potential (in the RHIC and LHC colliders, only charged particles interact with the tracker detectors, leaving bubble trajectories behind). A sketch of the LCP splitting for a jet produced in a hard-scattering event is shown on the top region of Figure 4.0.1, where the branching tree is represented by the black Feynman QCD vertices, the highest p_{T} particles are identified as the leading charge and the sub-leading charge and the LCP splitting itself is represented by the dashed magenta lines.

The other new splitting of interest is the soft-drop emission that resolves the 2 leading charged particles, called the resolved soft-drop splitting (RSD). Therefore, the RSD splitting corresponds to the de-clustering step where the sources of the leading and sub-leading charged particles get separated into 2 different prongs. This emission is shown on the bottom region of Figure 4.0.1 (assumed to satisfy the SD criterion), alongside a representation of the 2 RSD prongs as magenta cones, each containing one of the leading charged hadrons.

Therefore, the following Sections provide a comparative study of jet substructure between 3 different splittings: the 1SD, RSD and LCP splittings. The LCP splitting will serve as a baseline for effects taking place at hadron level (manifestly non-perturbative). On the other hand, the 1SD emission is mostly dominated by partonic branchings, as shown in Chapter 3. Finally, the RSD splitting will give information on where the 2 leading charged particles will behave as independent sources, which can happen at widely different scales. This comparison holds the potential of revealing how to use jet substructure observables to flag the transition from partonic to non-perturbative hadronic scales.

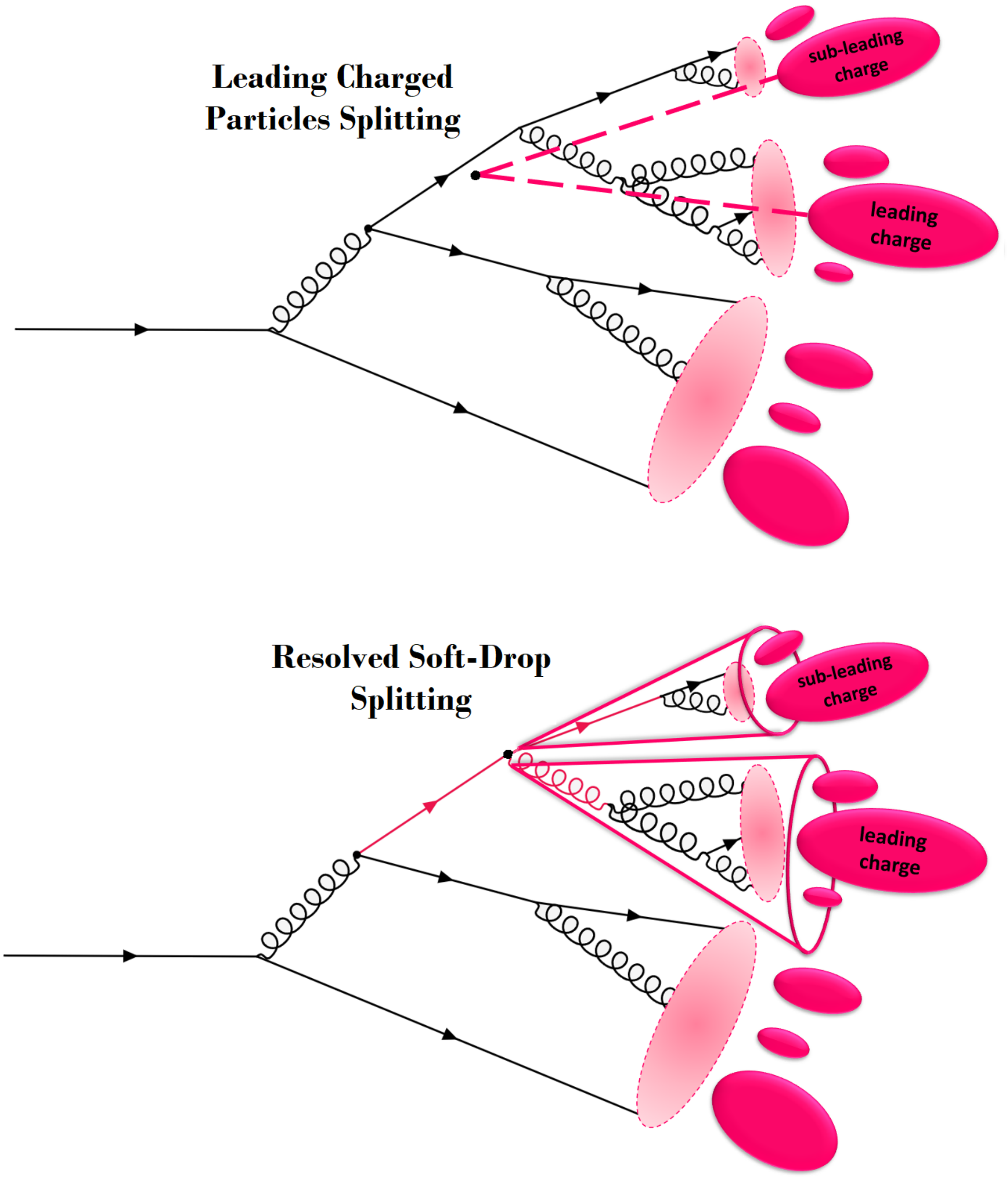


Figure 4.0.1: Leading charged particles splitting (LCP), represented by the dashed magenta lines in the branching scheme shown on the top half of the Figure and the resolved soft-drop splitting (RSD), represented by the highlighted magenta emission in the identical branching scheme shown on the bottom half of the Figure, alongside the magenta cones representing the 2 RSD prongs; arrowed lines represent quarks, spiral lines gluons, translucent magenta ellipses the hadronization process and opaque magenta ellipses the outgoing hadrons.

The study performed in this Chapter is focused on $\sqrt{s} = 200$ GeV pp collisions, for jets in the $20 < p_{\mathrm{T,jet}} < 40$ GeV/c range, and on $\sqrt{s} = 5$ TeV pp collisions, for jets in the $200 < p_{\mathrm{T,jet}} < 300$ GeV/c range. These kinematic settings are motivated by their relevancy in experimental applications at the RHIC and LHC colliders, respectively. The PYTHIA 8.306 event-simulation details are layed out in Section 3.1. The jet analysis performed using *FastJet* submits final-state particles to an anti-k_t clustering procedure with $R = 0.6$ and the identified jets are then submitted to a $|\eta_{\mathrm{jet}}| < 1$ selection and a re-clustering procedure using, alternately, the C/A and τ algorithms (for more details, check Section 3.2). The only difference with respect to the jet analysis performed in Chapter 3 arises from the fact that any anti-k_t jet without at least 2 electrically charged final-state particles is discarded, in order to ensure the 1SD, RSD and LCP distributions always sample the same jet populations and are, therefore, comparable. All distributions are self-normalized and have their bin contents divided by the respective bin widths unless explicitly stated otherwise.

4.1 Jet Substructure across the 3 Splittings

The RSD splitting is, by definition, either coincident with the 1SD emission or posterior to it (along the main jet branch). If it takes place in any of the first 3 soft-drop emissions, RSD distributions should resemble the ones presented in Section 3.4 from the previous Chapter. However, the RSD can also take place later in the clustering tree and it may in fact coincide with the LCP splitting, in which case both their distributions should be the same for any given jet observable in study.

To understand how many SD emissions occur after the RSD splitting ($\mathrm{N_{post\text{-}RSD}}$), one investigates the jet distributions with respect to this variable. $\mathrm{N_{post\text{-}RSD}}$ provides a measure of how far the RSD splitting is from the end of the clustering tree. This number can be traced along both the main and the secondary RSD prongs. The $\mathrm{N_{post\text{-}RSD}}$ for RHIC and LHC kinematics are shown in Figures 4.1.1 and 4.1.2, respectively. Left panels show the counting performed along the main RSD prong and the right panels the counting along the secondary one. Distributions represented with square markers use C/A re-clustering and the ones with diamond markers use τ.

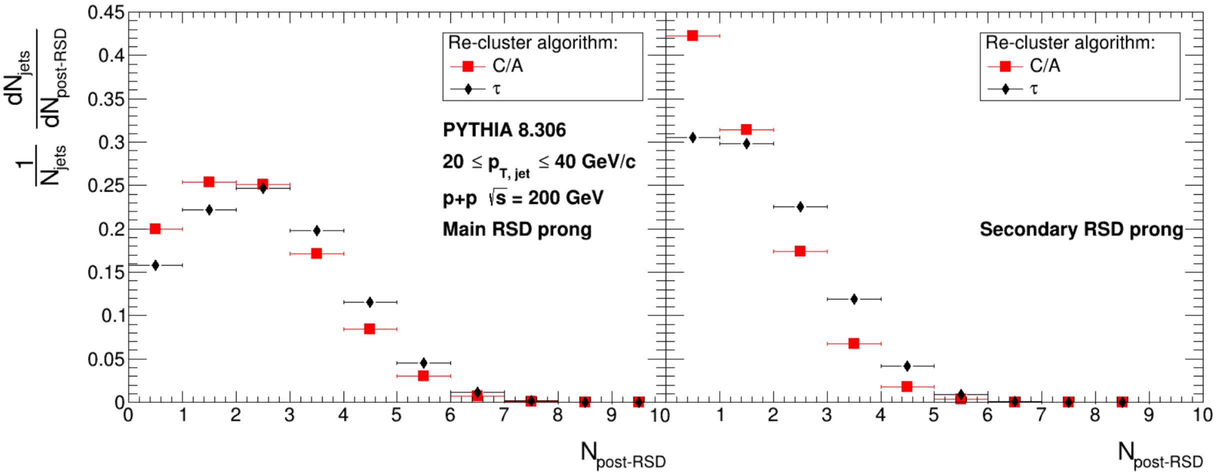


Figure 4.1.1: Jet distributions for the number of post-RSD splittings ($\mathrm{N_{post\text{-}RSD}}$) for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions, represented by red square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show these distributions for SD emission-counting along the main and secondary RSD prongs, respectively.

First and foremost, the similarity between the shapes of RHIC and LHC $\mathrm{N_{post\text{-}RSD}}$ distributions is noteworthy, with the LHC ones being somewhat more shifted to larger numbers. Furthermore, all $\mathrm{N_{post\text{-}RSD}}$ distributions show that the secondary RSD prong, being the softest, has smaller number of SD emissions post-RSD than the main RSD prong. They show that, in spite of the main RSD prong having, on

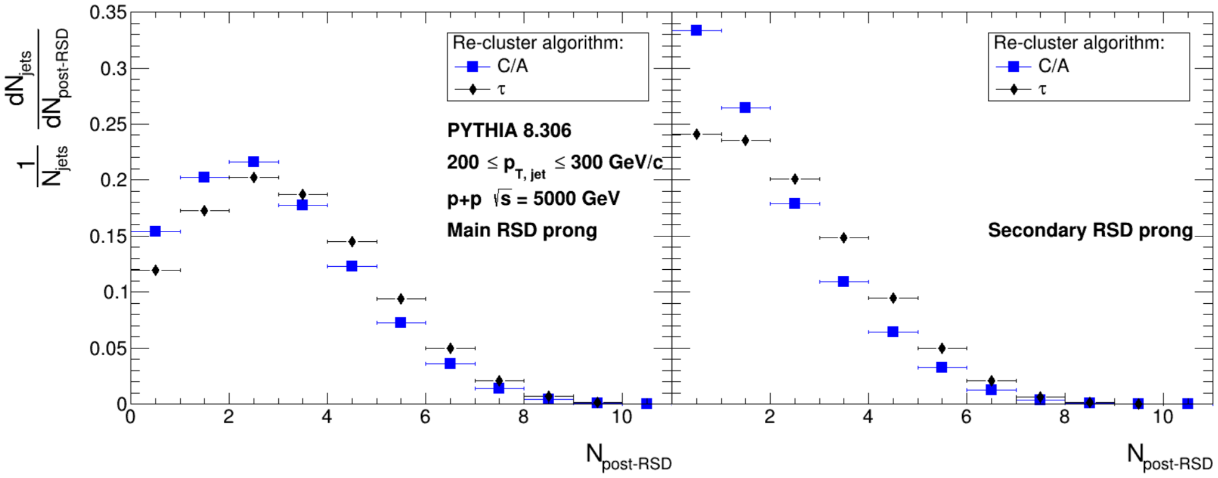


Figure 4.1.2: Jet distributions for the number of post-RSD splittings ($N_{\text{post-RSD}}$) for $200 < p_{T,\text{jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions, represented by blue square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show these distributions for SD emission-counting along the main and secondary RSD prongs, respectively.

average, a few SD emissions left, the secondary prong often has no further SD emissions, effectively setting the outermost edge of the clustering tree. In fact, a sizable minority of the jet populations will have no further SD emissions on either RSD prong, in which case the RSD splitting naturally coincides with the LCP. However, a more detailed study regarding the RSD placement in the clustering tree (performed in Appendix A.1) also reveals that the RSD splitting is placed at the 1SD emission for a relative majority of jets (from both kinematic settings).

With these expectations in mind, the rest of this Section follows with a substructure study on multiple jet observables. Starting with the groomed momentum fraction z_g (introduced in Eq. (2.22)), Figure 4.1.3 reinforces the observation made in Section 3.4 regarding the 1SD emission being highly asymmetrical, as are the pQCD parton splitting functions in Equations (2.12). This translates into a significantly uneven momentum share between the main and secondary jet branches. However, Figure 4.1.3 also reveals the LCP splitting to have the exact opposite behaviour, with a much higher tendency to equitably share the transverse momentum between the 2 leading charged particles. As for the RSD splitting, its behaviour is somewhere in between the latter 2, being more symmetrical than the 1SD and more asymmetrical than the LCP. While the 1SD distributions are highly peaked for small z_g and LCP ones for large z_g, the RSD distributions are shown to be significantly flatter. Nevertheless, the RSD splitting is shown to be tendentiously symmetrical, more to the likes of the LCP splitting and diverging from the DGLAP splitting functions expected for pQCD. Overall, the behaviour of the 1SD, RSD and LCP momentum fraction distributions is shown to be very similar for RHIC and LHC jets.

Figure 3.4.2 (z_g distributions for the first 3 soft-drop emissions found along the main jet branch) already established that SD emissions become progressively more symmetrical in transverse momentum along the clustering tree. However, the RSD z_g distributions from Figure 4.1.3 have significant shape differences with respect to the z_g distributions for the 1SD, 2SD and 3SD emissions. RSD is generally more symmetrical than all 3 of them, in spite of being situated in the 1SD, 2SD or 3SD emission the vast majority of times (study performed in A.1). This happens because the RSD splitting resolves, by

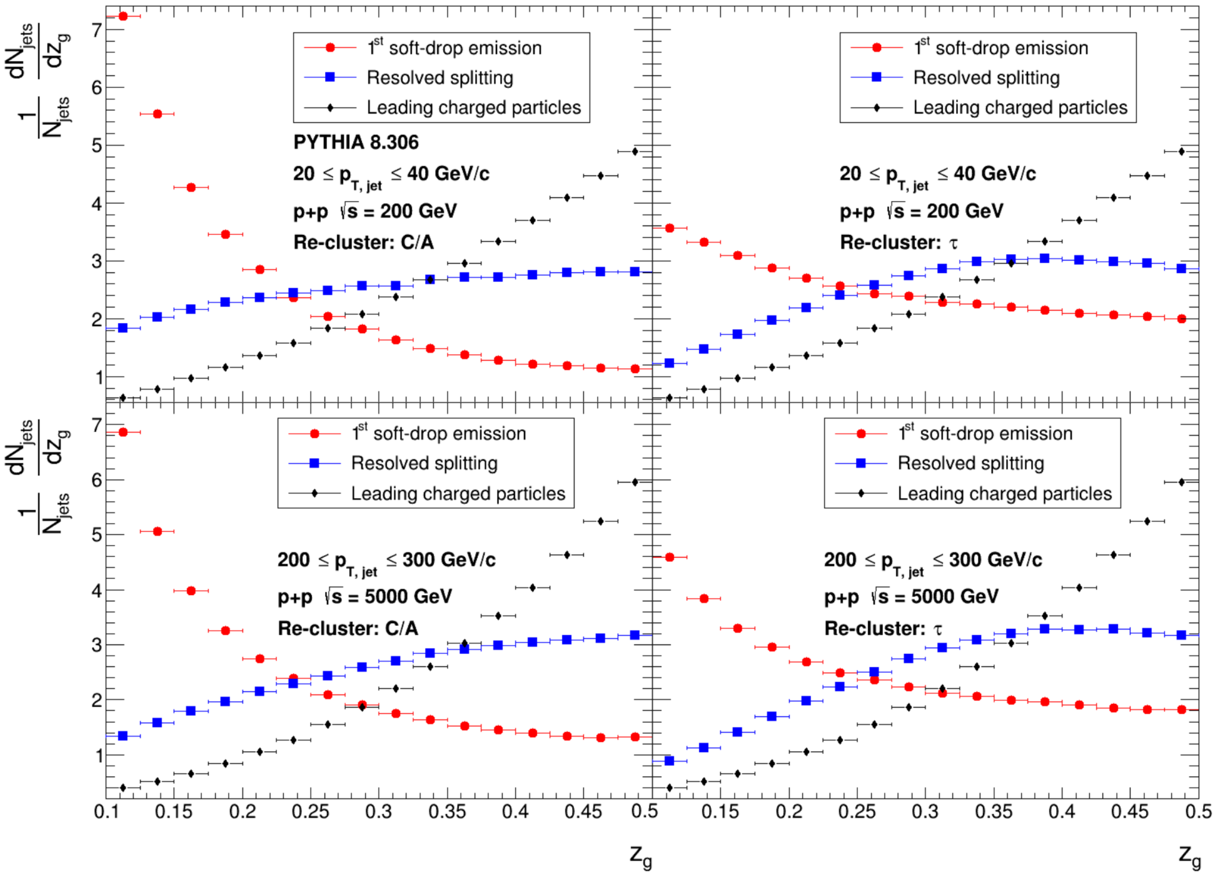


Figure 4.1.3: Momentum fraction (z_g) distributions for splittings from jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c produced in pp collisions with energies of 200 GeV, in the top panels, and from jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c produced in pp collisions with energies of 5 TeV, in the bottom ones; z_g distributions are represented in red circle, blue square and black diamond markers for the 1st soft-drop emission, the resolved soft-drop splitting and the leading charged particles splitting, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

definition, the 2 most energetic charged particles of the jet, producing typically 2 hard RSD prongs. Even if the RSD is known to be located at the nth SD emission, its z_g will be higher than the z_g of the generic nth SD emission.

The same study is performed with the splitting opening angle ΔR in Appendix B. It reinforces that the RSD splitting moves away from typical pQCD behaviour, with the RSD ΔR distribution being significantly more similar to the LCP than to the 1SD ones.

Finally, one focuses on the formation time τ_{form} at each of these 3 splittings. The τ_{form} is calculated using Equations (2.24), (3.3) and (3.4). The results in Figure 4.1.4 show a shift to higher τ_{form} caused by the 1SD → RSD → LCP transition for RHIC kinematic settings. RSD distributions follow the 1SD ones more closely for small τ_{form} and then the LCP ones for large τ_{form}. In fact, the $\tau_{\mathrm{form,\,RSD}}$ distributions are significantly flatter than the $\tau_{\mathrm{form,\,1SD}}$ ones. The greater variability in $\tau_{\mathrm{form,\,RSD}}$ stems from the fact that the leading charged particles need not be resolved neither at the very first soft-drop splitting nor at the very last emission of the tree, leading to a more uniform RSD distribution.

To make a more rigorous analysis of the relative behaviour of τ_{form} distributions for the 3 relevant splittings, it is interesting to look at the plots with the ratios between these distributions. The ratios are

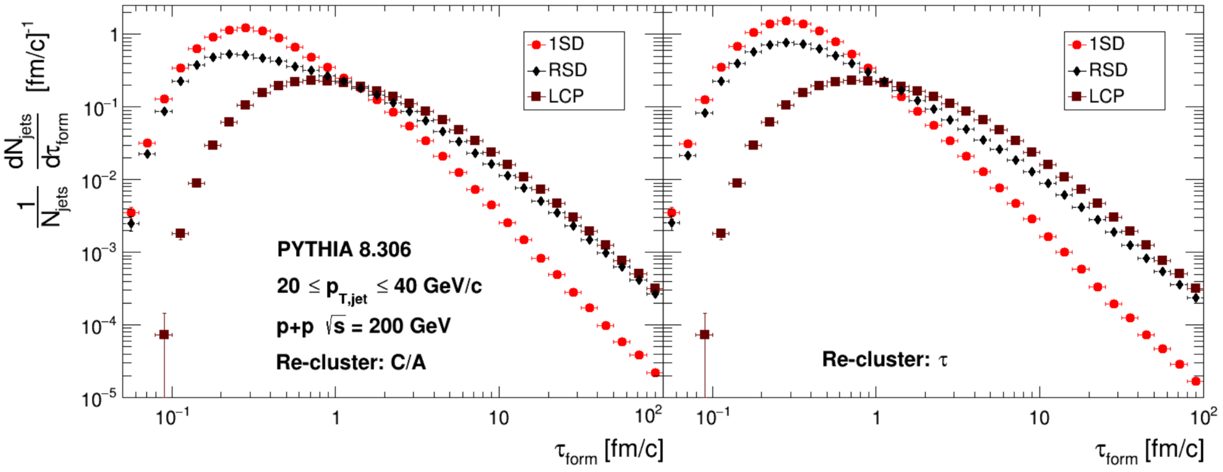


Figure 4.1.4: Formation time distributions for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in pp collisions with $\sqrt{s} = 200$ GeV, corresponding to the 1st soft-drop emission (1SD), in red circles, the resolved soft-drop splitting (RSD), in black diamonds, and the leading charged particles splitting (LCP), in brown squares; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

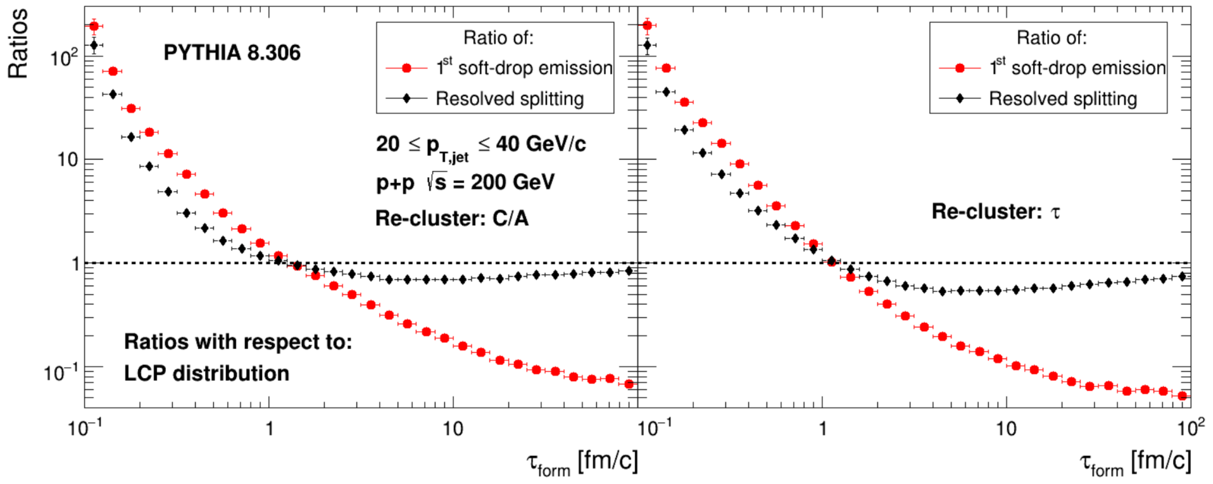


Figure 4.1.5: Ratios of the formation time distributions for the 1st soft-drop emission and for the resolved splitting with respect to the formation time distribution for the leading charged particles splitting (LCP), represented in red circle and black diamond markers, respectively, for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in pp collisions with energies of 200 GeV; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

chosen to be with respect to the LCP distribution, so to establish comparisons in relation to the only splitting that is independent of the re-clustering algorithm. Figure 4.1.5 shows the shapes of $\tau_{\mathrm{form,\,1SD}}$ and $\tau_{\mathrm{form,\,LCP}}$ distributions are never the same. Small τ_{form} is dominated by 1SD, while large τ_{form} is dominated, naturally, by the LCP, leading the 1SD/LCP ratio to sharply decrease with τ_{form} (for both re-clustering algorithms).

However, when it comes to the RSD/LCP ratio, Figure 4.1.5 reveals that the RSD splitting dominates at small τ_{form}, but it quickly decreases and stagnates at ratios below but very close to unity. For that range, RSD and LCP seem to very closely match when it comes to the shape of their τ_{form} distributions. In order to identify the τ_{form} range for which the RSD/LCP ratio behaves as a “plateau” near unity, the τ_{form} value where the RSD/LCP ratio becomes smaller than unity is taken as reference. Using this method, the RHIC “plateau” region of the RSD/LCP ratio is identified for both re-clustering algorithms as

$$\text{Plateau:} \quad \tau_{\text{form}} > 1.3 \pm 0.3 \text{ fm/c} \quad . \tag{4.1}$$

This hints that cuts in τ_{form} will select jets with differently-placed RSD splittings. If the cut is made on $\tau_{\text{form,RSD}}$ for the range in Equation (4.1), the RSD of the selected jets seems to have the hadron-like time-structure of the LCP splitting.

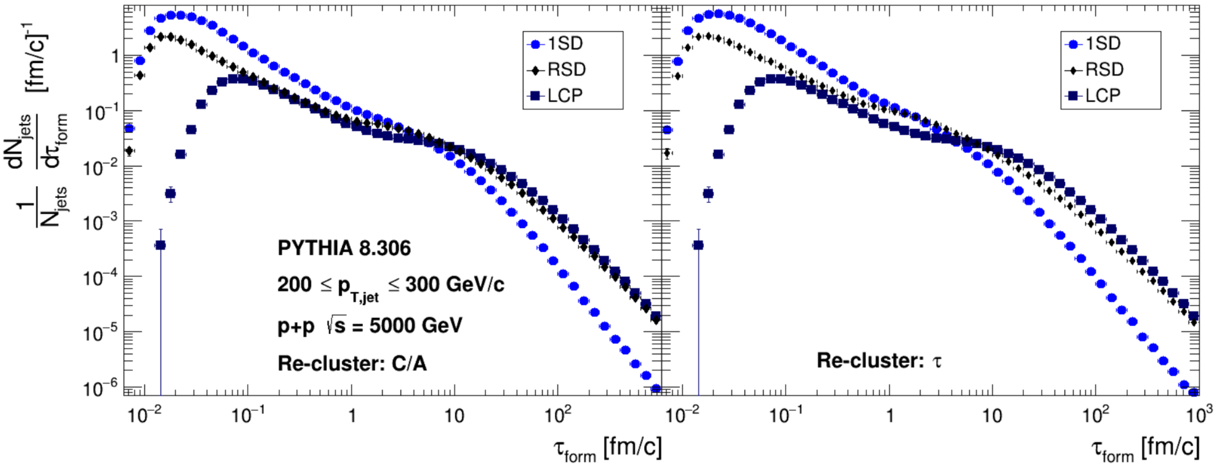


Figure 4.1.6: Formation time distributions for jets with $200 < p_{\text{T,jet}} < 300$ GeV/c produced in pp collisions with energies of 5 TeV, corresponding to the 1st soft-drop emission distribution, in light blue circles, the resolved soft-drop splitting, in black diamonds, and the leading charged particles splitting, in dark blue squares; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

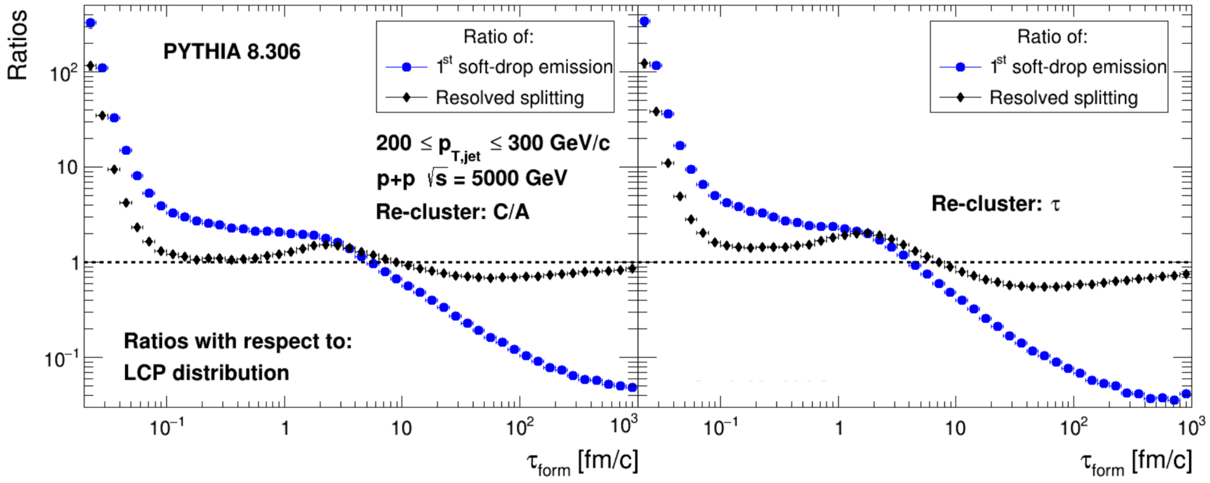


Figure 4.1.7: Ratios of the formation time distributions for the 1st soft-drop emission and for the resolved splitting with respect to the formation time distribution for the leading charged particles splitting (LCP), represented in red circle and black diamond markers, respectively, for $200 < p_{\text{T,jet}} < 300$ GeV/c jets produced in pp collisions with energies of 5 TeV; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

The τ_{form} distributions for LHC kinematic settings, shown in Figure 4.1.6, reveal the same τ_{form}-ordering for the different splittings, with τ_{form} increasing significantly for 1SD → RSD → LCP. It also shows the shape of the RSD distributions to resemble the 1SD ones for small τ_{form} and then the LCP ones for large τ_{form}.

In Figure 4.1.7, the 1SD/LCP ratio distribution shows once again that there are substantial shape differences between these 2 distributions, while the RSD/LCP ratio quickly drops until it eventually stabilizes around unity for large τ_{form}. However, the RSD/LCP ratio drop happens much earlier for LHC jets than it does for RHIC settings and it also behaves very differently around the unitary ratio. The ratio first acquires values close but above unity and then transitions to values close but below unity. This allows the identification of a double-"plateau" structure in this LHC ratio plot. The 2^{nd} plateau is identified using the τ_{form} for which RSD/LCP becomes smaller than unity. However, since the 1^{st} "plateau" does not cross the unity line, it is identified as the τ_{form} range for which the RSD/LCP ratio is above unity and within a 0.2 variation interval. Then, the "plateaus" are identified as

$$\begin{aligned} &1^{\text{st}}\text{ Plateau:} \quad 0.10 \pm 0.03 < \tau_{\text{form}} < 1.0 \pm 0.3 \text{ fm/c} \\ &2^{\text{nd}}\text{ Plateau:} \quad \tau_{\text{form}} > 8 \pm 2 \text{ fm/c} \end{aligned} \tag{4.2}$$

for C/A re-clustering and

$$\begin{aligned} &1^{\text{st}}\text{ Plateau:} \quad 0.10 \pm 0.03 < \tau_{\text{form}} < 0.6 \pm 0.2 \text{ fm/c} \\ &2^{\text{nd}}\text{ Plateau:} \quad \tau_{\text{form}} > 6 \pm 2 \text{ fm/c} \end{aligned} \tag{4.3}$$

for τ re-clustering.

Once again, these cuts in $\tau_{\text{form, RSD}}$ will select jets with more LCP-like RSD splittings, indicating they might be taking place at hadronic scales.

The effect on τ_{form} introduced by identifying the 2 leading charged particles instead of making no charge imposition on them, as well as some other charge-related studies, are presented in Appendix C.

4.2 Jet Selection via Formation Time

Given the "plateau" structure of RSD/LCP ratios found at RHIC and LHC energies, one proceeds to separate the jet samples according to the RSD formation time cuts presented in Equations (4.1), (4.2) and (4.3). The $\tau_{\text{form,RSD}}$ ranges for the sharply falling ratios (drop regions) are considered to be the ones immediately prior to the RHIC "plateau" ($\tau_{\text{form,RSD}} < 1.3$ fm/c) and to the LHC 1^{st} "plateau"($\tau_{\text{form,RSD}} < 0.10$ fm/c). With this jet selection on $\tau_{\text{form,RSD}}$, one looks for time-specific substructure properties that may help identify if the RSD does in fact flag the transition between pQCD and npQCD for larger timescales.

4.2.1 Number of SD Splittings Post-RSD

The $N_{\text{post-RSD}}$ distributions are presented in Figure 4.2.1 for RHIC jets with a $\tau_{\text{form, RSD}}$ cut on the drop region of the RSD/LCP ratio and in Figure 4.2.2 for RHIC jets with a $\tau_{\text{form, RSD}}$ cut on the "plateau" region.

The distributions are represented in red square and black diamond markers for C/A and τ re-clustered jets, respectively. The left panel shows SD-counting along the main RSD prong, while the right panel does it for the secondary one. The same plots are shown for LHC jets in the drop (Figure 4.2.3), 1^{st} "plateau" (Figure 4.2.4) and 2^{nd} "plateau" (Figure 4.2.5) regions of the RSD/LCP ratio, now with C/A distributions represented by blue square markers.

Figures 4.2.1 and 4.2.2 show that RHIC jets with $\tau_{\text{form, RSD}}$ in the drop region (small $\tau_{\text{form, RSD}}$) generally have still a few SD emissions taking place after the RSD along both branches. However, jets that fall on the "plateau" range (large $\tau_{\text{form, RSD}}$), in spite of still having a few SD emissions along the main RSD prong, are shown to have practically none along the secondary branch. To get a more rigorous notion of these numbers, Table 4.1 shows the averages of the RHIC distributions. One concludes that the RSD splitting effectively marks the end of the clustering tree along the secondary RSD prong for RHIC jets with $\tau_{\text{form, RSD}}$ in the "plateau" region of the RSD/LCP ratio. In this time interval, the $\text{N}_{\text{post-RSD}}$ distributions previously presented in Figure 4.1.1 are significantly shifted to smaller numbers. This hints that larger $\tau_{\text{form, RSD}}$ selects jets whose RSD splittings are in the vicinity of the hadronization scales.

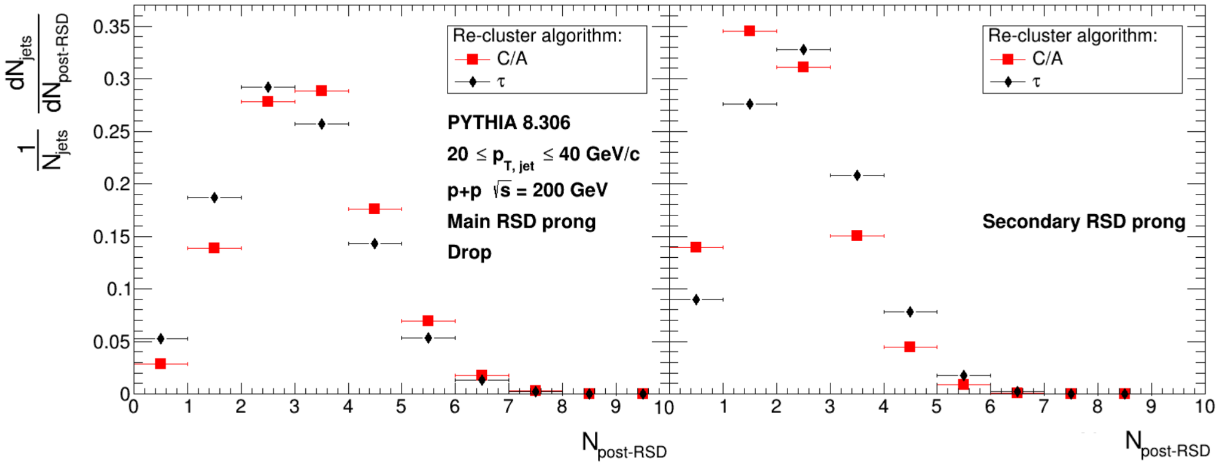


Figure 4.2.1: Jet distributions for the number of post-RSD splittings ($\text{N}_{\text{post-RSD}}$) for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions and for whom the RSD splitting has formation time in the drop range of the RSD/LCP ratio; they are represented by red square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show them for SD-splitting counting along the main and secondary RSD prongs, respectively.

The LHC results in Figures 4.2.3, 4.2.4 and 4.2.5 also show that jets sampled from smaller RSD time cuts still have, on average, multiple SD splittings taking place after the RSD along both RSD prongs. In fact, the $\text{N}_{\text{post-RSD}}$ distributions found for the drop and 1^{st} "plateau" regions of the RSD/LCP ratio are very similar with each other. This suggests that the flat shape of the 1^{st} "plateau" may not have any physical meaning when it comes to tagging the transition from parton to hadron levels (this matter is investigated further in the next Section). However, for the higher $\tau_{\text{form, RSD}}$ of the 2^{nd} "plateau", the clustering tree keeps evolving for most jets along the main RSD prong, but it effectively stops along the secondary prong. For quantitative observations, Table 4.2 shows the average $\text{N}_{\text{post-RSD}}$ for the LHC distributions. This averages clearly show that the drop and 1^{st} "plateau" cuts are indeed very similar to each other, while the 2^{nd} "plateau" shows a sharp decrease in $\text{N}_{\text{post-RSD}}$ along both prongs.

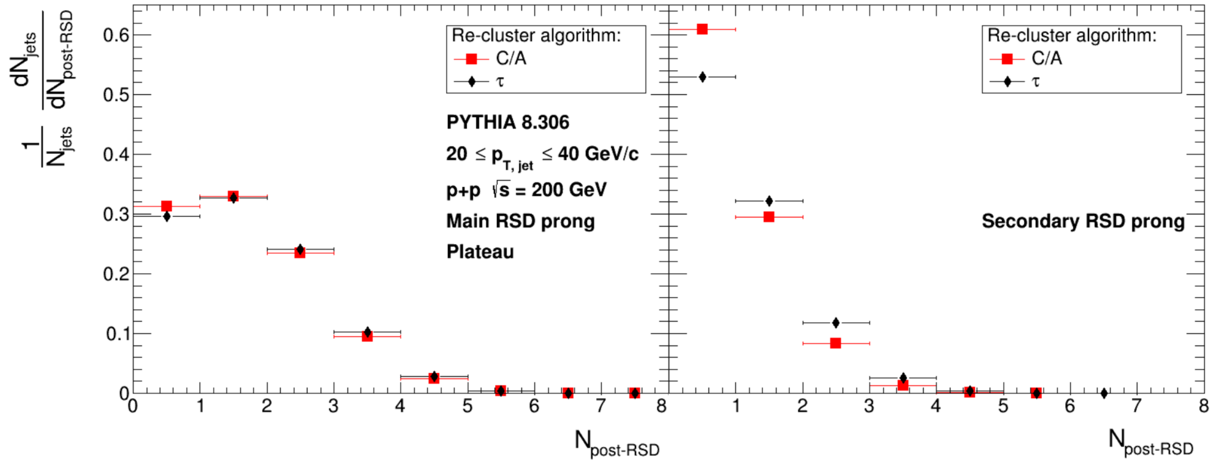


Figure 4.2.2: Jet distributions for the number of post-RSD splittings ($N_{\text{post-RSD}}$) for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions and for whom the RSD splitting has formation time in the plateau range of the RSD/LCP ratio; they are represented by red square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show them for SD-splitting counting along the main and secondary RSD prongs, respectively.

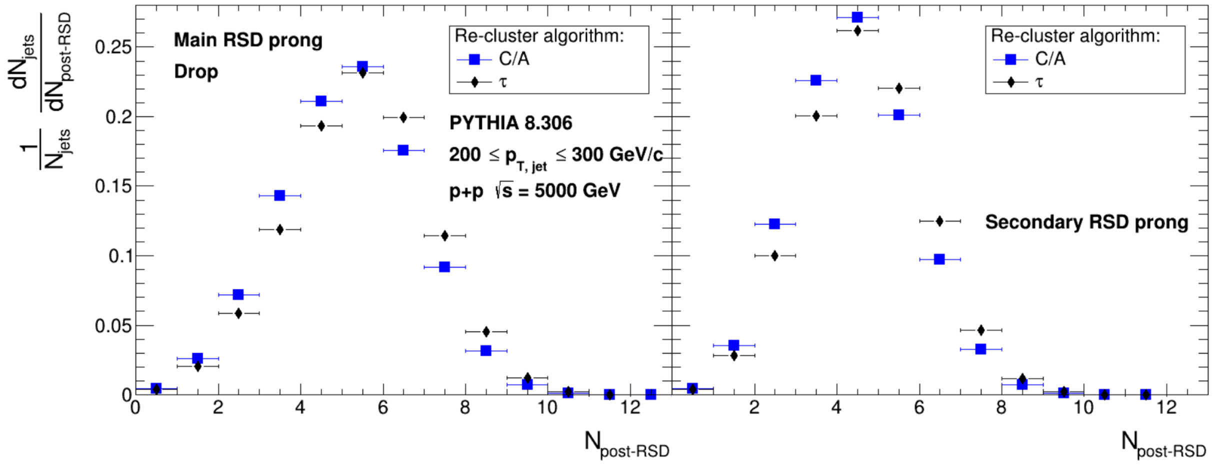


Figure 4.2.3: Jet distributions for the number of post-RSD splittings ($N_{\text{post-RSD}}$) for $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions and for whom the RSD splitting has formation time in the drop range of the RSD/LCP ratio; they are represented by red square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show them for SD-splitting counting along the main and secondary RSD prongs, respectively.

The similarity between the plots in Figure 4.2.2 (“plateau” of the RHIC ratio) and in Figure 4.2.5 (2^{nd} “plateau” of the LHC ratio) is noteworthy. That resemblance can also be seen by comparing the average $N_{\text{post-RSD}}$ in the last row of Tables 4.1 and 4.2. This reinforces the conclusion that large $\tau_{\text{form, RSD}}$ selects jets whose RSD splitting comes very close to the ending stages of the clustering tree.

The study performed in Appendix A.2 shows in more detail the relative placement of the RSD splitting, both with respect to the beginning and to the ending stages of the clustering tree, for the different $\tau_{\text{form, RSD}}$ cuts. It compares the total number of SD emissions counted along the main jet branch (N_{splits}) with the number of SD emissions left on the main RSD prong ($N_{\text{post-RSD}}$). It also reveals that cuts for higher $\tau_{\text{form, RSD}}$ select jets whose RSD usually takes place after the 1SD, while the rest of the jets

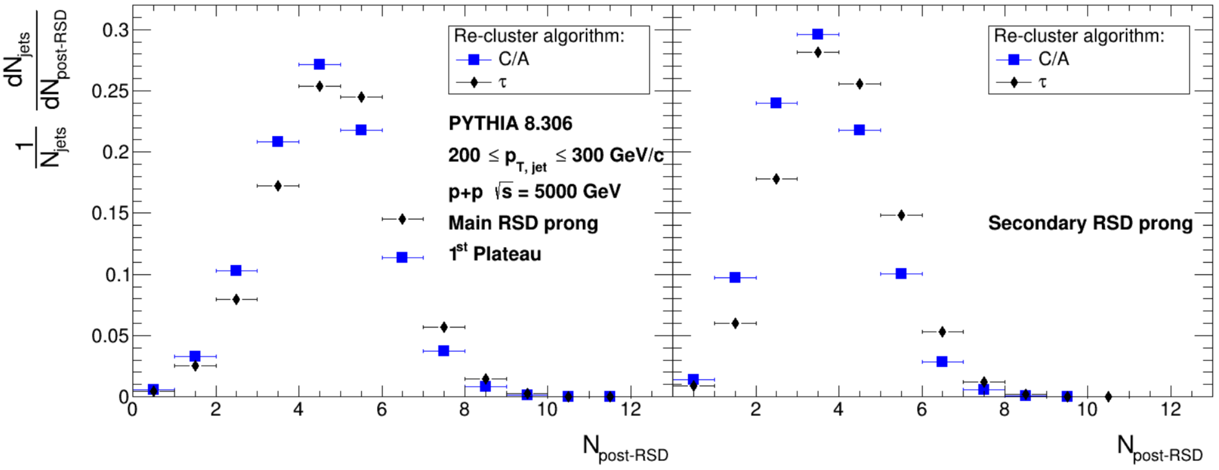


Figure 4.2.4: Jet distributions for the number of post-RSD splittings ($\mathrm{N_{post\text{-}RSD}}$) for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions and for whom the RSD splitting has formation time in the 1^{st} plateau range of the RSD/LCP ratio; they are represented by red square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show them for SD-splitting counting along the main and secondary RSD prongs, respectively.

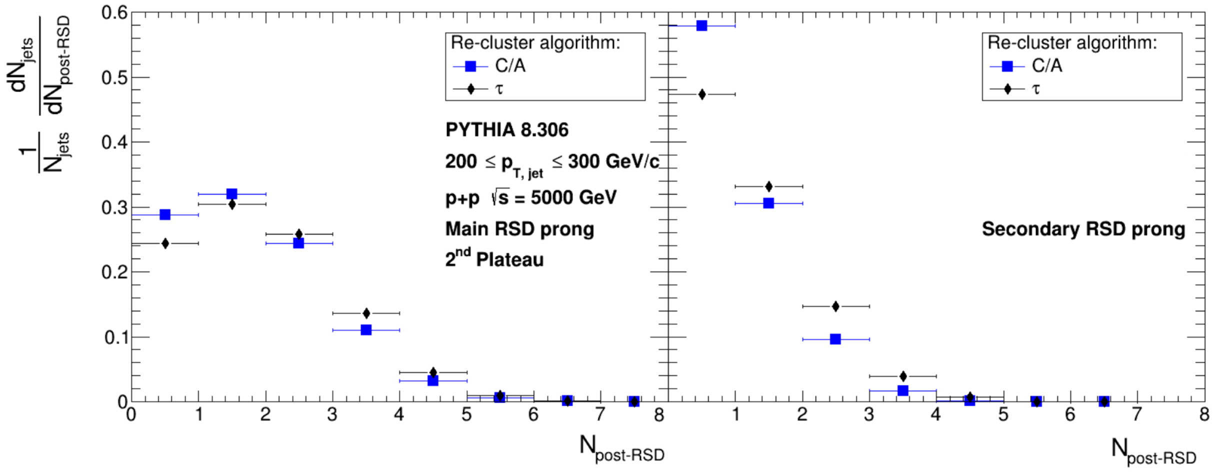


Figure 4.2.5: Jet distributions for the number of post-RSD splittings ($\mathrm{N_{post\text{-}RSD}}$) for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions and for whom the RSD splitting has formation time in the 2^{nd} plateau range of the RSD/LCP ratio; they are represented by red square and black diamond markers for C/A and τ re-clustering, respectively; left and right panels show them for SD-splitting counting along the main and secondary RSD prongs, respectively.

Ranges	Average $\mathrm{N_{post\text{-}RSD}}$			
	C/A re-cluster		τ re-cluster	
	Main	Secondary	Main	Secondary
Drop	2.740 ± 0.003	1.644 ± 0.002	2.475 ± 0.002	1.971 ± 0.002
Plateau	1.202 ± 0.002	0.502 ± 0.001	1.257 ± 0.002	0.653 ± 0.002

Table 4.1: Average number of soft-drop splittings after the resolved splitting takes place, both along the main and the secondary RSD prongs and both for C/A and τ re-clustered $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions, for the drop and plateau RSD formation time regions.

Ranges	Average $N_{\text{post-RSD}}$			
	C/A re-cluster		τ re-cluster	
	Main	Secondary	Main	Secondary
Drop	4.654 ± 0.006	3.934 ± 0.006	4.919 ± 0.006	4.167 ± 0.005
1^{st} Plateau	4.060 ± 0.005	3.054 ± 0.004	4.351 ± 0.005	3.446 ± 0.005
2^{nd} Plateau	1.302 ± 0.002	0.557 ± 0.001	1.470 ± 0.002	0.776 ± 0.002

Table 4.2: Average number of soft-drop splittings after the resolved splitting takes place, both along the main and the secondary RSD prongs and both for C/A and τ re-clustered $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions, for the drop, 1^{st} plateau and 2^{nd} plateau RSD formation time regions.

(smaller $\tau_{\text{form, RSD}}$) have RSD ≃ 1SD.

However, further investigation into jet substructure for the different $\tau_{\text{form, RSD}}$ cuts is needed in order to cement these conclusions.

4.2.2 Groomed Momentum Fraction and Jet Mass

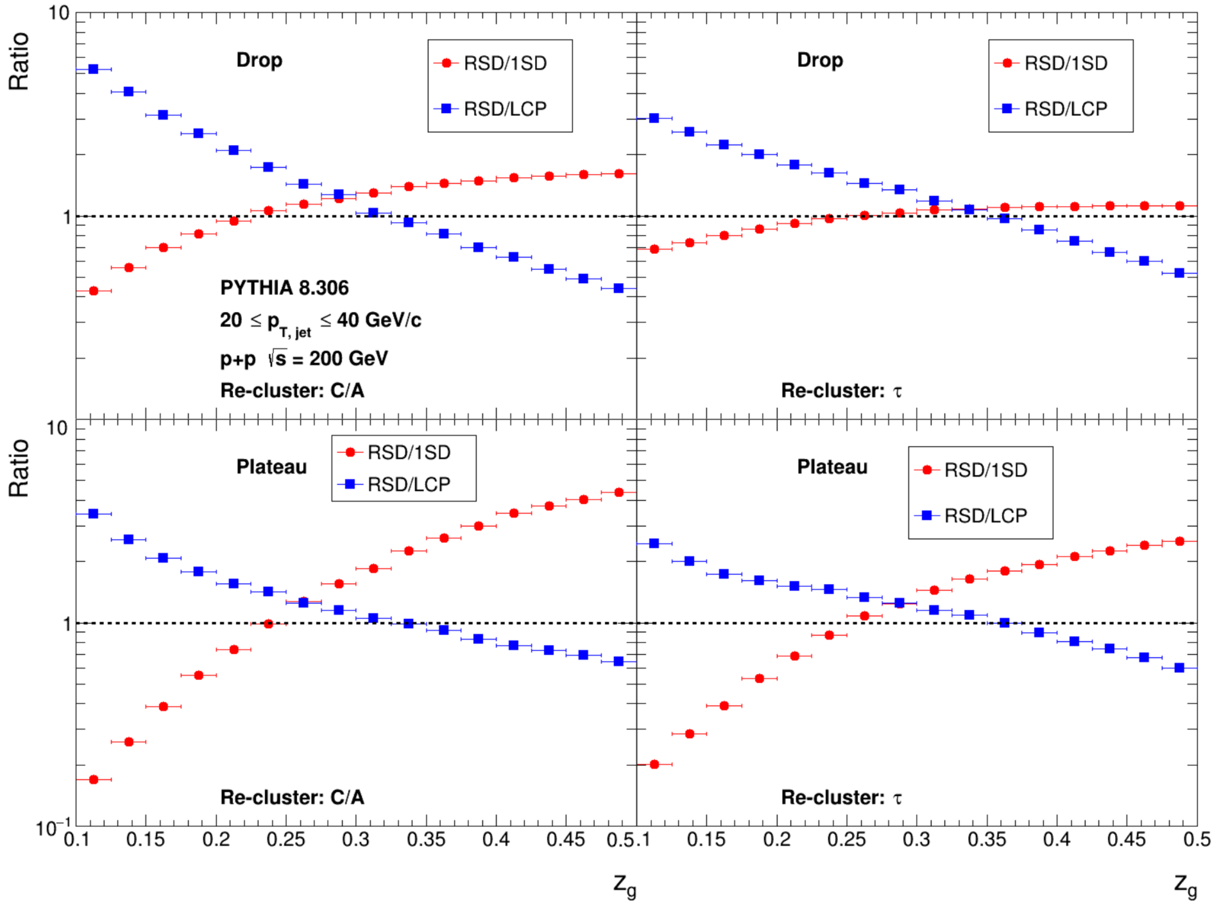


Figure 4.2.6: Ratios of the RSD momentum fraction distributions (z_g) with respect to the 1SD (red circle markers) and LCP (blue square markers) z_g distributions for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions; top and bottom panels sample jets with $\tau_{\text{form, RSD}}$ on the drop and plateau ranges of the RSD/LCP time ratios; left panels use C/A re-clustering and right one use τ.

The z_g distributions at the 1SD, RSD and LCP splittings for the previous $\tau_{\text{form, RSD}}$ selections are shown in Appendix D for RHIC and LHC kinematic settings and using both C/A and τ clustering trees. However, this Section focuses on the RSD/1SD and RSD/LCP ratios between those z_g distributions in order to study how RSD compares with both the 1SD and the LCP for each $\tau_{\text{form, RSD}}$ range of interest. These ratios can be observed in Figures 4.2.6 and 4.2.7 for RHIC and LHC kinematic settings, respectively. RSD/1SD ratios are represented by red circle markers and RSD/LCP ones by blue square markers.

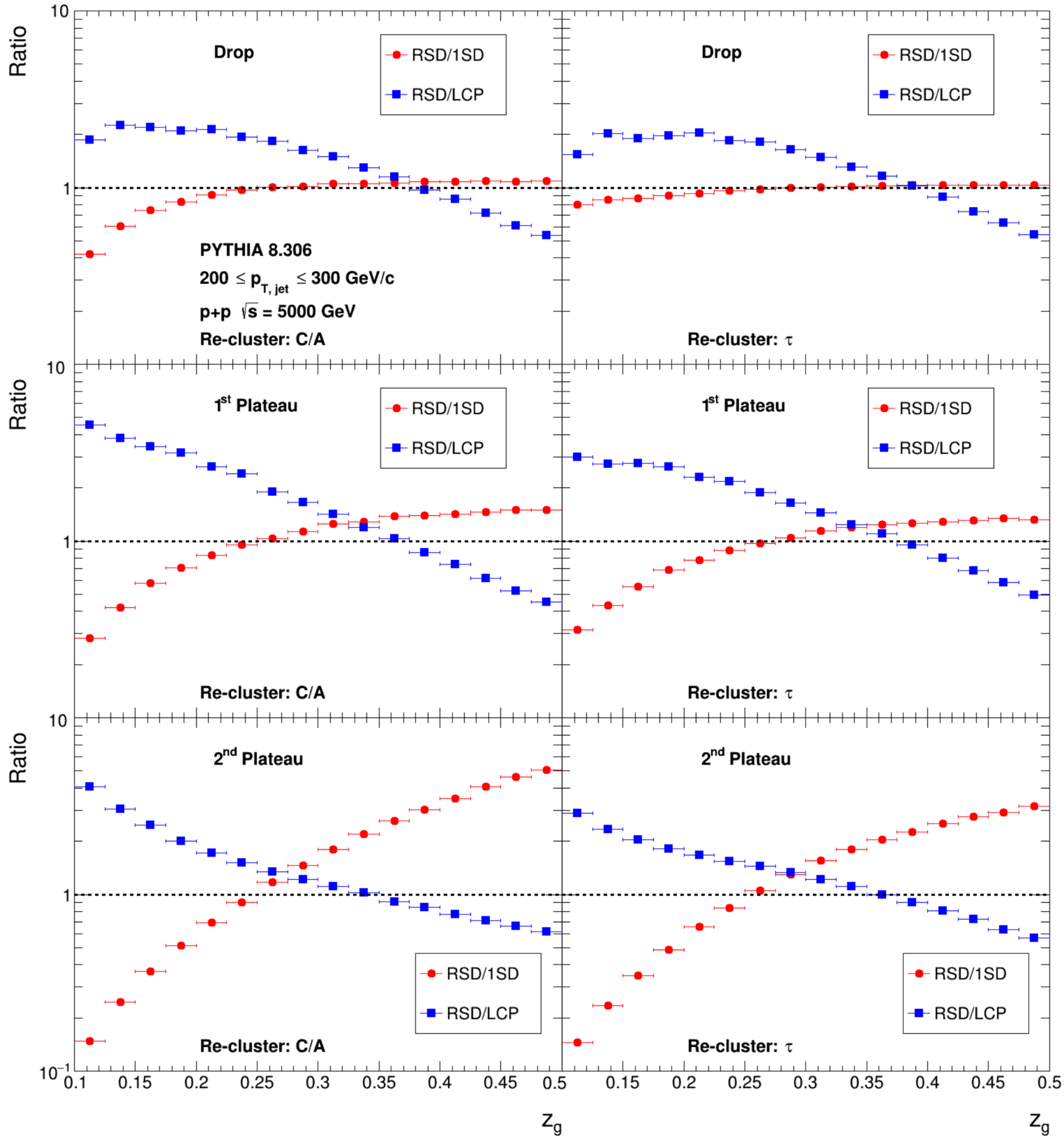


Figure 4.2.7: Ratios of the RSD momentum fraction distributions (z_g) with respect to the 1SD (red circle markers) and LCP (blue square markers) z_g distributions for $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions; top, middle and bottom panels sample jets with $\tau_{\text{form, RSD}}$ on the drop, 1st plateau and 2nd plateau ranges of the RSD/LCP time ratios; left panels use C/A re-clustering and right one use τ.

The bottom panels from Figure 4.2.6 clearly show that low $\tau_{\text{form, RSD}}$ means RSD splittings have z_g more similar to the 1SD emission. This observation was to be expected from the analysis in Appendix A.2, that shows the RSD takes place at the 1SD emission when $\tau_{\text{form, RSD}}$ is small. However, for higher $\tau_{\text{form, RSD}}$, shown on the bottom panels, the RSD momentum fraction clearly diverges from the 1SD. It is also slightly more similar to LCP z_g distributions, showing that it acquires npQCD behaviour for higher time cuts (Appendix D shows RSD becoming more symmetric for large $\tau_{\text{form, RSD}}$).

Figure 4.2.7 verifies for LHC the same observations made before for RHIC. The top panels, using jets with low $\tau_{\text{form, RSD}}$ from the LHC drop range, show almost identical z_g distributions for RSD and 1SD, with the RSD/1SD ratio being very close to unity. However, the higher $\tau_{\text{form, RSD}}$ from LHC's 2nd plateau range (bottom panels) shows RSD splittings diverging more from the 1SD. As for the intermediate 1st plateau range of the LHC time ratio, RSD/1SD z_g ratios are still relatively close to unity, similarly to what happens for the drop range. This hints, once again, that the 1st plateau of the LHC time ratio is not connected to anything in particular.

Furthermore, one also looks at the (un-groomed) jet mass M_{jet} distributions for jets whose RSD splitting falls on the relevant $\tau_{\text{form, RSD}}$ ranges of the RSD/LCP time ratios.

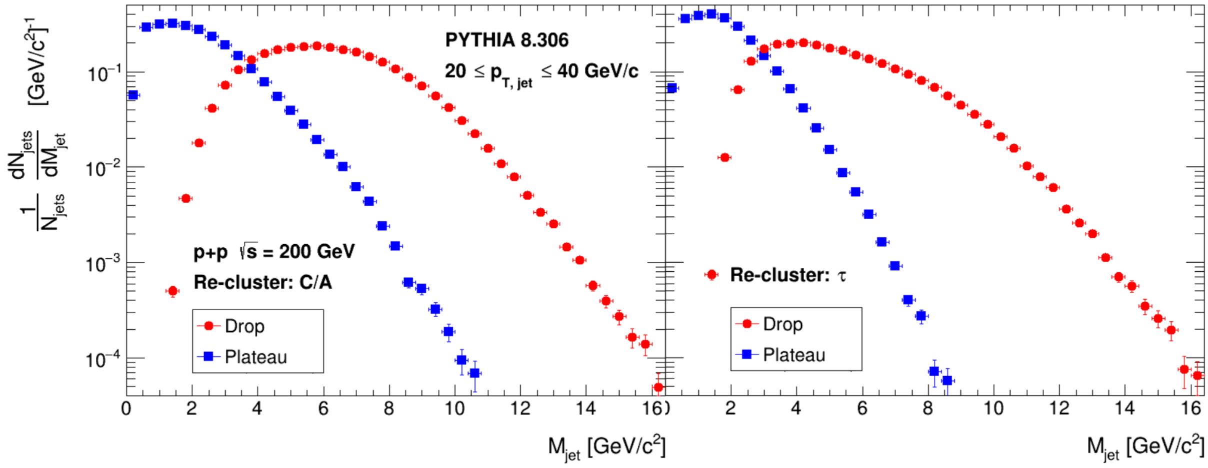


Figure 4.2.8: Distributions on the jet mass for $20 < p_{\text{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions and for whom the resolved soft-drop splitting has formation time in the drop and plateau regions from the RSD/LCP time ratio, represented by red circle and blue square markers, respectively; the left and right panels show results for C/A and τ re-clustering, respectively.

Figure 4.2.8 shows for RHIC jets a clear distinction in M_{jet} between the drop and "plateau" ranges. Jets with the higher $\tau_{\text{form, RSD}}$ of the "plateau" region have significantly smaller mass than the lower $\tau_{\text{form, RSD}}$ jets from the drop region. A clear separation in the M_{jet} distributions can also be observed in Figure 4.2.9 for LHC jets, with the jets from the 2nd "plateau" having much smaller mass than the jets from the 1st "plateau". Also, both of these sub-populations of jets have significantly smaller mass than the ones from the drop region. Since there is little overlap between M_{jet} distributions for each of the $\tau_{\text{form, RSD}}$ selections, the jet mass appears to be a good proxy for jet selection via formation time. Through a selection on higher jet masses, clustering trees will have less SD emissions and the RSD splitting will typically occur near LCP.

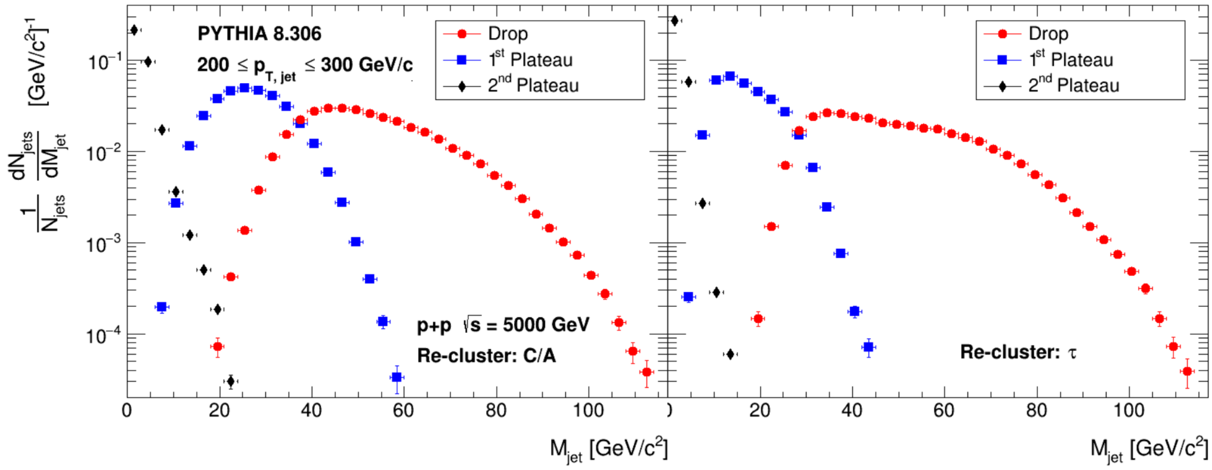


Figure 4.2.9: Distributions on the jet mass for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions and for whom the resolved soft-drop splitting has formation time in the drop, 1st plateau and 2nd plateau regions from the RSD/LCP time ratio, represented by red circle, blue square and black diamond markers, respectively; the left and right panels show results for C/A and τ re-clustering, respectively.

4.3 RSD Formation Time along the Clustering Tree

One could gain some insight into the "plateaus" of the RHIC and LHC RSD/LCP time ratios by computing separately the contributions to the RSD formation time distribution according to its location in the clustering tree.

From the study performed in Appendix A.1, it is clear that the vast majority of jets from both RHIC and LHC kinematic settings show a preference for early RSD splittings. More than 90% of RHIC and 80% of LHC jets have their respective RSD splitting happening either at $\mathrm{N_{RSD}} = 1$, $\mathrm{N_{RSD}} = 2$ or $\mathrm{N_{RSD}} = 3$, for both re-clustering algorithms, with later RSD splittings becoming progressively rarer. The percentages of RHIC and LHC clustering trees, found both with the C/A and τ algorithms, with their RSD splitting placed in the 1SD, 2SD, 3SD and later than 3SD emissions are displayed in Table 4.3. Therefore, studies regarding time-dependency on RSD placement can be restricted to the RSD splittings occurring in the first 3 soft-drop emissions along the main branch.

RSD Placement	RHIC Settings		LHC Settings	
	C/A trees	τ trees	C/A trees	τ trees
1SD	43%	53%	34%	38%
2SD	31%	27%	27%	27%
3SD	17%	14%	20%	18%
>3SD	9%	6%	19%	17%

Table 4.3: Percentage of jets with $15 < p_{\mathrm{T,jet}} < 20$ GeV/c produced in pp collisions of $\sqrt{s} = 200$ TeV (RHIC kinematic settings) and of jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c from $\sqrt{s} = 5$ TeV collisions (LHC kinematic settings) whose RSD splitting is located at the 1SD, 2SD, 3SD and later than 3SD emissions along the main branch of C/A and τ clustering trees.

Figure 4.3.1 shows, for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions, the

$\tau_{\text{form, RSD}}$ distribution for the RSD splitting in black diamond markers (displayed already in Figure 4.1.4) alongside the contributions from RSD at the n^{th} soft-drop emission along the main branch, with $n = 1, 2, 3$, represented in red circle, blue square and green cross markers, respectively. The generic RSD distribution is self-normalized, but the other 3 are normalized with respect to the first one. The left panel samples C/A clustering trees, while the right one samples τ trees. The same distributions are shown in Figure 4.3.2 for $200 < p_{\text{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions.

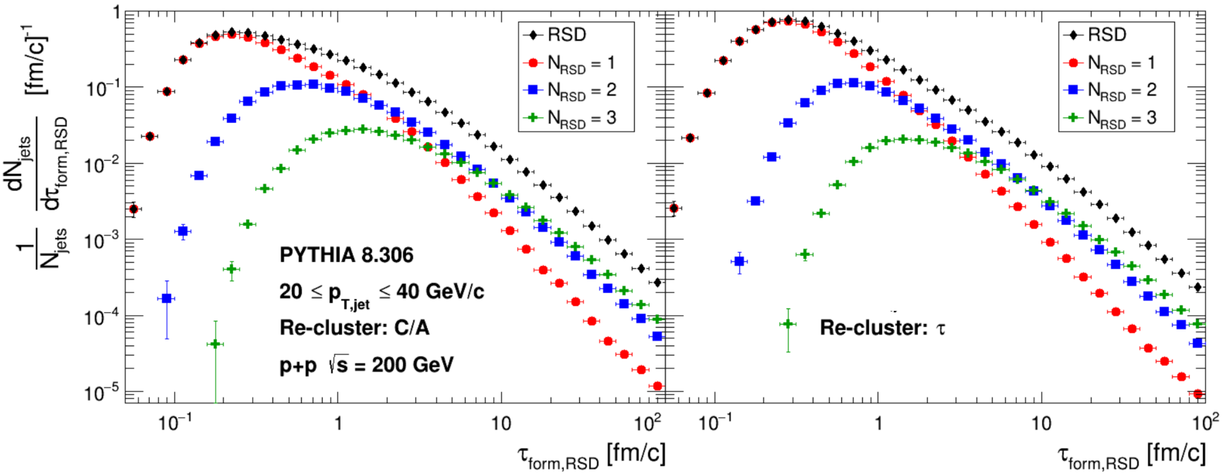


Figure 4.3.1: Formation time distributions for the generic resolved soft-drop splitting (RSD), in black diamond markers, and for the resolved splitting at $N_{\text{RSD}} = 1, 2, 3$, in red circle, blue square and green cross markers, respectively, for $20 < p_{\text{T,jet}} < 40$ GeV/c jets produced in pp collisions with energies of 200 GeV; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

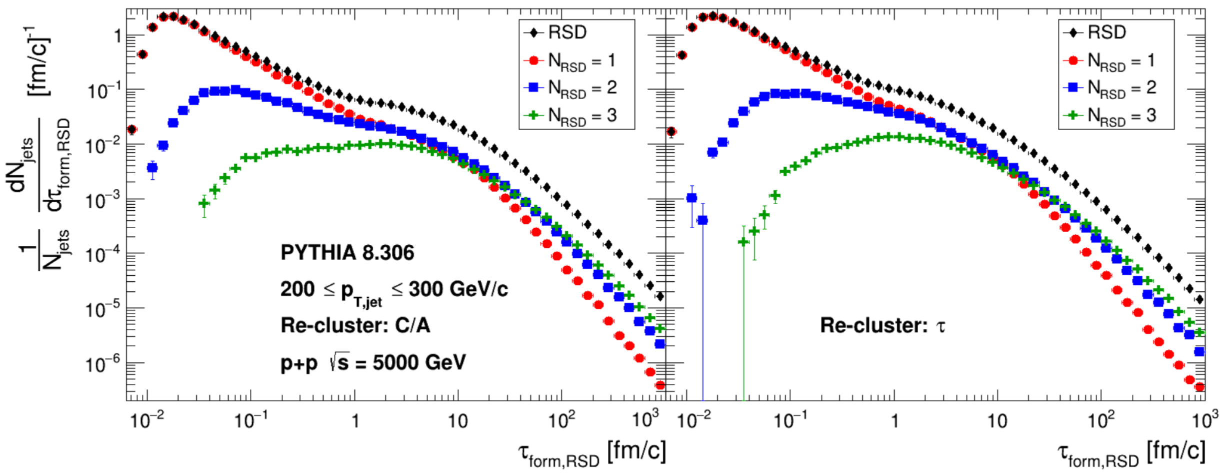


Figure 4.3.2: Formation time distributions for the generic resolved soft-drop splitting (RSD), in black diamond markers, and for the resolved splitting at $N_{\text{RSD}} = 1, 2, 3$, in red circle, blue square and green cross markers, respectively, for $200 < p_{\text{T,jet}} < 300$ GeV/c jets produced in pp collisions with energies of 5 TeV; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

The results presented in Figures 4.3.1 and 4.3.2 show, naturally, that the later the RSD splitting takes place in the clustering tree, the higher is the formation time of the RSD splitting. However, it is now relevant to check how the different RSD placements contribute to the RSD/LCP ratios presented in Figures 4.1.5 and 4.1.7, namely in the “plateau” regions. Therefore, the RSD/LCP time ratios are now calculated for RSD splittings taking place at the 1SD, 2SD and 3SD emissions, using the self-normalized $\tau_{\text{form, RSD}}$ distributions. These ratios are shown in Figures 4.3.3 and 4.3.4.

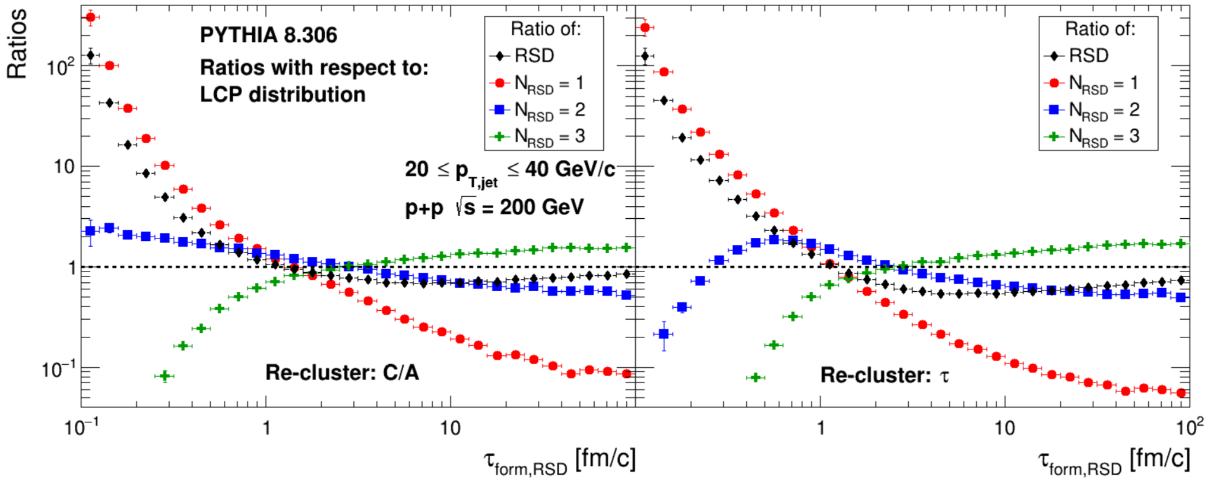


Figure 4.3.3: Ratios of the formation time distributions for the generic resolved splitting (RSD) and for the resolved splitting at $N_{\text{RSD}} = 1, 2, 3$, with respect to the formation time distribution for the leading charged particles splitting (LCP), evaluated for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from in $\sqrt{s} = 200$ GeV pp collisions, represented in black diamond, red circle, blue square and green cross markers, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

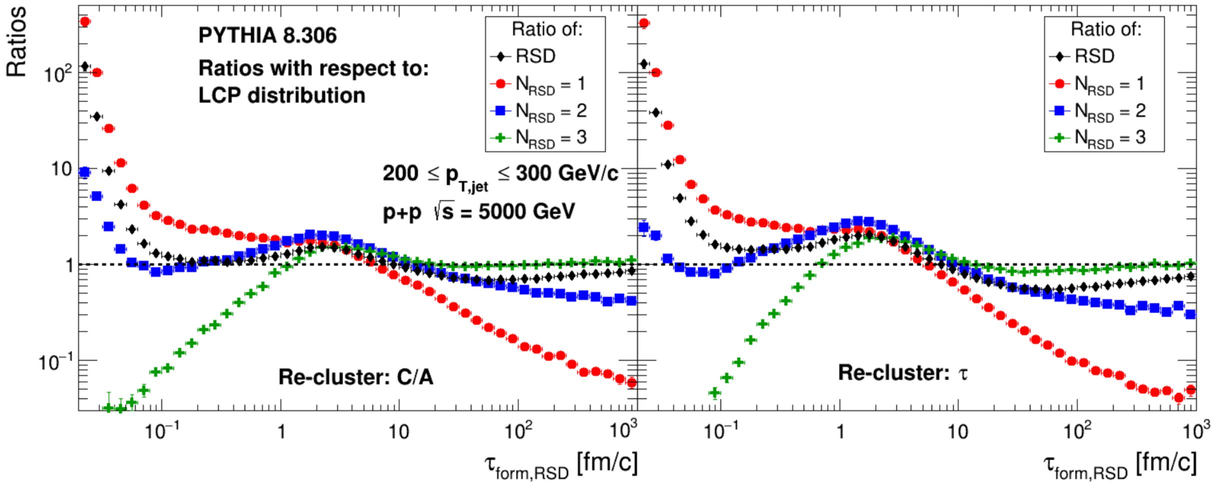


Figure 4.3.4: Ratios of the formation time distributions for the generic resolved splitting (RSD) and for the resolved splitting at $N_{\text{RSD}} = 1, 2, 3$, with respect to the formation time distribution for the leading charged particles splitting (LCP), evaluated for $200 < p_{\text{T,jet}} < 300$ GeV/c jets from in $\sqrt{s} = 5$ TeV pp collisions, represented in black diamond, red circle, blue square and green cross markers, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

Figure 4.3.3 (RHIC) clearly shows that the “plateau” region of the RSD/LCP time ratio gets its shape from the contributions of later RSD splittings, namely for $N_{\text{RSD}} = 2$ and $N_{\text{RSD}} = 3$. The red circle distribution, showing the RSD/LCP ratio for $N_{\text{RSD}} = 1$, is almost identical to the 1SD/LCP ratio presented in Figure 4.1.5, with very high ratios around small τ_{form} and small ratios for large τ_{form}. However, for $N_{\text{RSD}} = 2$ and $N_{\text{RSD}} = 3$, the ratios are shown to hit a “plateau” around the same time where the generic RSD/LCP “plateau” is located at.

As is evident from Figure 3.2.2, jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c from 200 GeV pp collisions have, on average, 3 to 4 soft-drop emissions in total along the main jet branch, meaning the ratio distribution for $\mathrm{N_{RSD}} = 3$ represented in green cross markers in Figure 4.3.3 shows properties of the ending limits of the parton shower.

Figure 4.3.4 shows the composition of the double-“plateau” structure of the RSD/LCP ratios for LHC jets. It reveals that the 1st “plateau” from Figure 4.1.7 comes from the balance of the contributions from early RSD splittings, namely for $\mathrm{N_{RSD}} = 1$ and $\mathrm{N_{RSD}} = 2$, as is the case for the drop region. For the 1st “plateau”, these observations confirm that the RSD similarity to the LCP in this time-range has no physical interpretation in terms of tagging the transition from parton to hadron level, since its built mainly out of early (more pQCD-like) SD emissions.

However, the 2nd “plateau” is clearly shown to be dominated by later RSD splittings: $\mathrm{N_{RSD}} \geq 3$. Furthermore, the contribution makeup of the 2nd “plateau” is very similar to the one observed in Figure 4.3.3 for the single-“plateau” of the RHIC RSD/LCP ratio. This is evidence that the RHIC “plateau” and the LHC 2nd “plateau” both probe similar phenomena of the jet history, taking place at large values of formation time and, therefore, largely non-pQCD-like splittings.

All and all, one concludes that the late distribution of $\tau_{\mathrm{form,RSD}}$ is the same as the late distributions of $\tau_{\mathrm{form},LCP}$, marking a possible transition from the pQCD to the npQCD regimes.

4.4 Pre- and Post-RSD Groomed Momentum Fraction

Previous results suggest the RSD can mark the transition to non-pQCD splittings when selecting late $\tau_{\mathrm{form,RSD}}$. As such, this Section investigates how the z_g of the RSD splitting fairs in comparison with the z_g of the SD emission taking place immediately before it (RSD-1) and the SD emission taking place immediately after it (RSD+1) along the main jet branch. To assure RSD has indeed a previous and a posterior SD emission in the clustering tree, the study is restricted to jets whose RSD splitting is neither the first nor the last SD emission along the main branch. This is to see if there is a clear shift towards hadron level splitting functions after the RSD.

Figure 4.4.1 shows the z_g distributions of the SD emission previous to the RSD splitting (RSD-1), for the RSD splitting itself and for the SD emission posterior to the RSD splitting (RSD+1) along the main branch, represented in red circle, blue square and black diamond markers, respectively. The top panels show z_g distributions for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions and the bottom ones for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions. Left panels used C/A re-clustered jets and the right ones used τ re-clustering.

Figure 4.4.1 reveals the SD emission prior to the RSD splitting is overwhelmingly asymmetrical for both kinematic settings and for both re-clustering algorithms, with z_g distributions significantly shifted towards small z_g. This is consistent with the 1SD (that, by convention, takes place before the RSD in this jet selection). The RSD, by itself, is symmetrical, as seen previously in Figure 4.1.3. However, RSD+1 z_g distributions suggest the SD emission after RSD becomes more asymmetric than the RSD itself, as opposed to what one would naively expect if RSD would indeed mark the pQCD-npQCD transition.

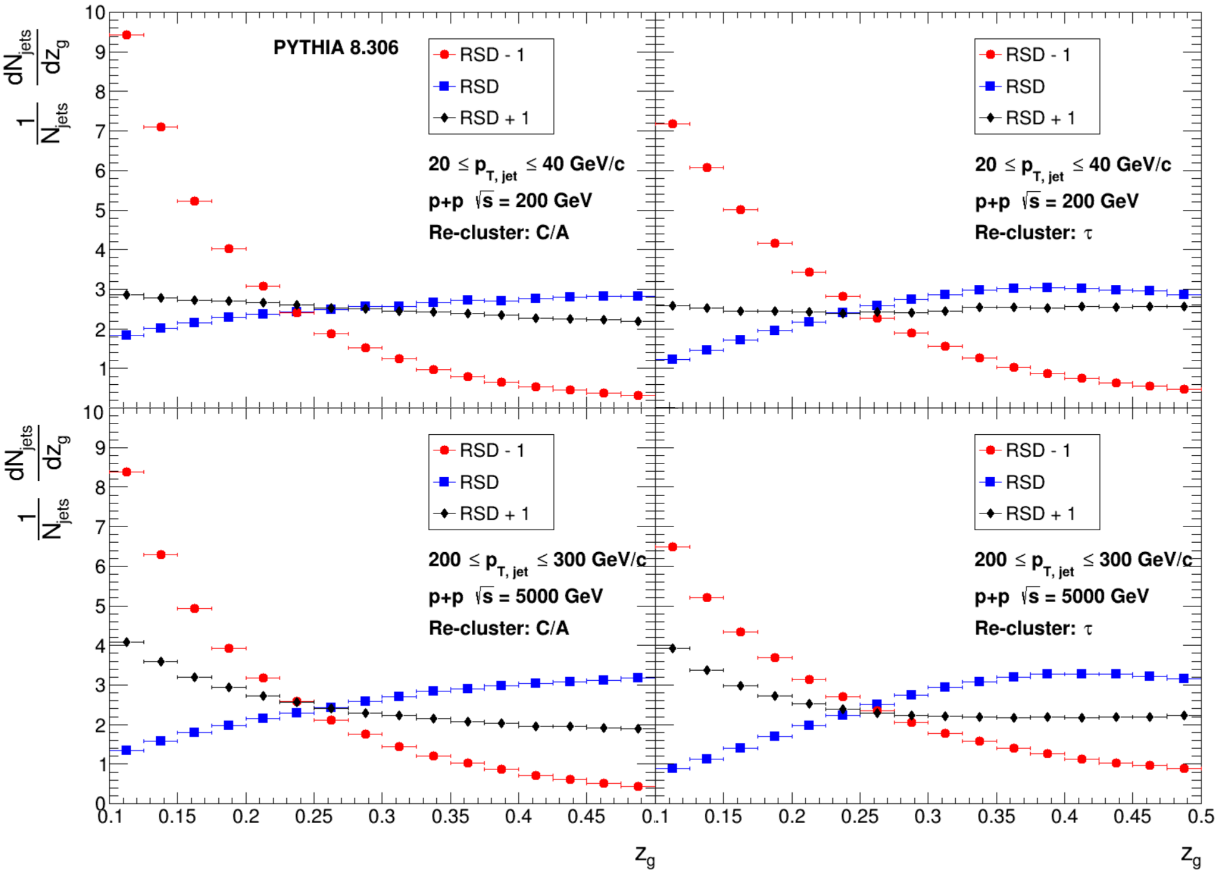


Figure 4.4.1: Groomed momentum fraction (z_g) distributions for the resolved soft-drop splitting (RSD), the previous SD emission (RSD-1) and the next SD emission (RSD+1) along the main jet branch, represented in blue square, red circle and black diamond markers, respectively, and evaluated for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from 200 GeV CM pp collisions, on the top panels, and for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from 5 TeV CM pp collisions, on the bottom panels; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

However, no $\tau_{\mathrm{form,RSD}}$ selection was performed on these jet populations. The selection that was indeed performed actually discards jets whose RSD sits at the end of the clustering tree, guaranteeing there is at least 1 soft-drop emission after the RSD along the main branch.

As such, it is important to finally look to the z_g for the different selections RSD=1SD, RSD=2SD and RSD=3SD, which are performed in the next Section.

4.5 Groomed Momentum Fraction with RSD Selections

To gain more insight into the results presented in Section 4.4, this Section studies how the z_g distributions of the first 3 soft-drop emission along the main branch change between the jet sub-populations for whom the RSD splitting is situated in the 1SD emission (RSD=1SD), in the 2SD emission (RSD=2SD) and the 3SD emission (RSD=3SD). Figure 4.5.1 shows the z_g distributions for the 1SD, 2SD and 3SD emissions along the main branch in red circle, blue square and black diamond markers, respectively, for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions (RHIC). The top panels show these distributions for jets whose RSD splitting coincides with the 1SD emission, the middle panels for jets whose RSD splitting coincides with the 2SD emission and the bottom ones for jets whose RSD splitting coincides with the 3SD emission. Left panels use C/A re-clustered jets and the right ones use τ re-clustering. The same z_g distributions are shown in Figure 4.5.2 for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions (LHC).

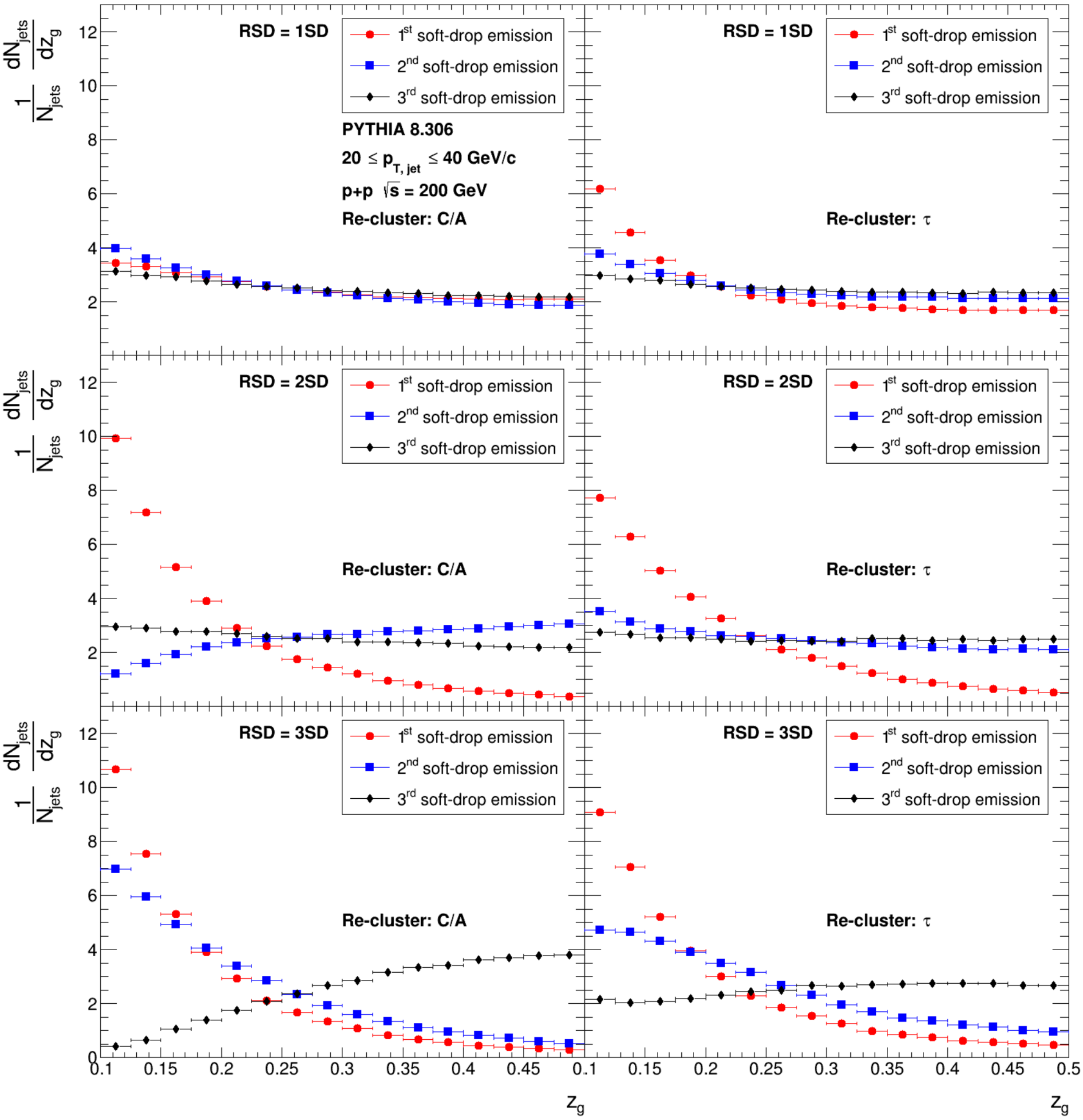


Figure 4.5.1: Groomed momentum fraction (z_g) distributions for the 1SD, 2SD and 3SD emissions, represented by red circle, blue square and black diamond markers, respectively, for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions; distributions for RSD=1SD,2SD,3SD are displayed in the top, middle and bottom panels, respectively; left panels sample C/A clustering trees and the ones on the right sample τ ones.

The top panels of Figure 4.5.1 reveal that when RSD=1SD the z_g distributions of the 1SD, 2SD and 3SD emissions are all notoriously flat. Since RSD=1SD, the 1SD z_g distributions were themselves expected to be flat. However, the 2SD and 3SD z_g distributions are significantly flatter than the generic 2SD and 3SD z_g distributions shown in Figure 3.4.2. As for the middle panels, where RSD=2SD, they show the 1SD is significantly asymmetrical, the 2SD emission, being the resolved one, has a flat z_g distribution, and the posterior 3SD has also a flat z_g distribution. Finally, the bottom panels, having RSD=3SD, show highly asymmetrical 1SD and 2SD emissions, even more asymmetrical than the generic 1SD and 2SD emissions shown in Figure 3.4.2. These observations hold independently of the re-clustering algorithm and are very similar to the ones drawn for Figure 4.5.2, with LHC kinematic settings. RSD separates the previous and more asymmetrical SD emissions from the subsequent and more symmetrical ones.

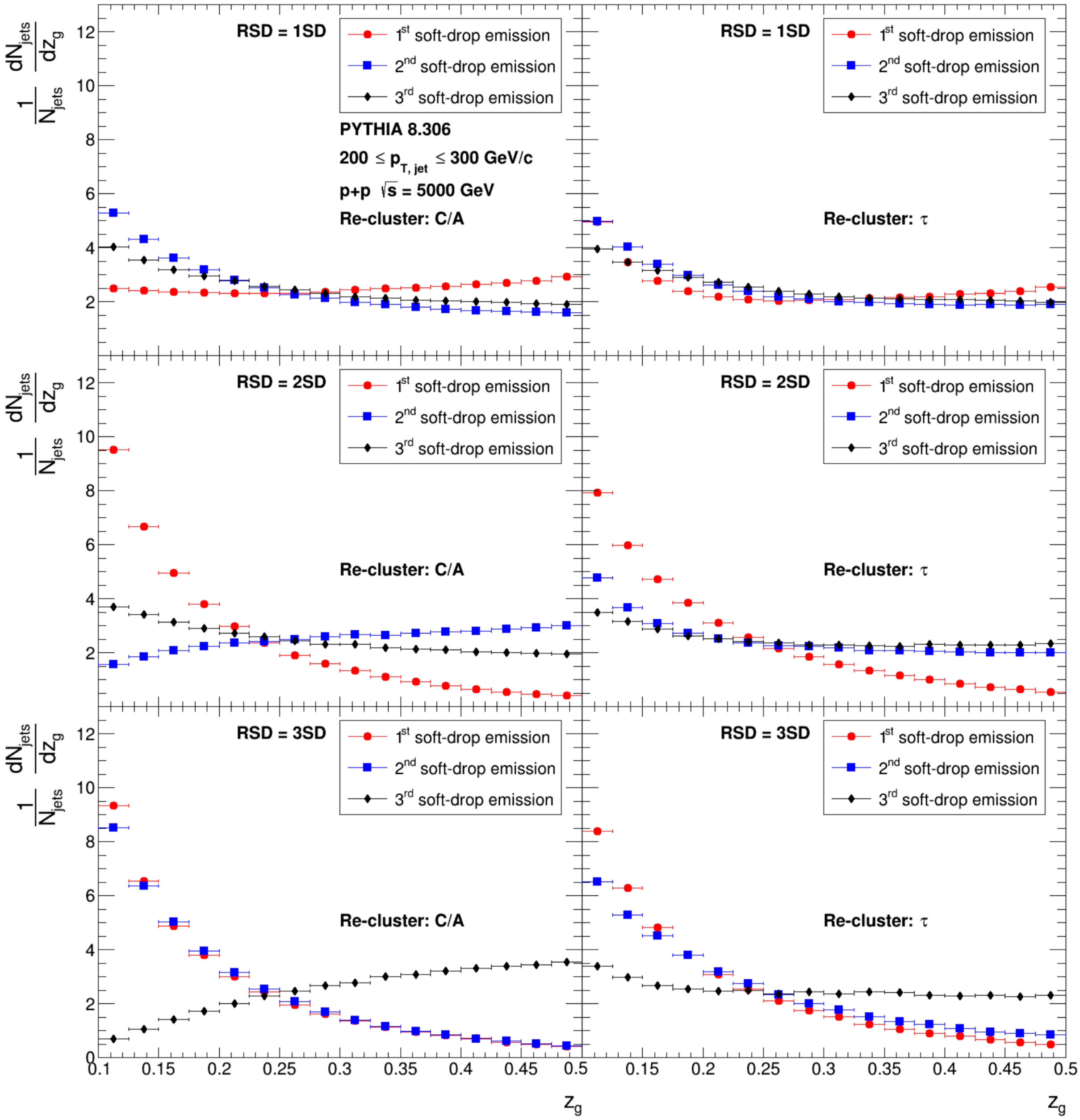


Figure 4.5.2: Groomed momentum fraction (z_g) distributions for the 1SD, 2SD and 3SD emissions, represented by red circle, blue square and black diamond markers, respectively, for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions; distributions for RSD=1SD,2SD,3SD are displayed in the top, middle and bottom panels, respectively; left panels sample C/A clustering trees and the ones on the right sample τ ones.

These observations suggest the RSD splitting indeed marks the transition between the asymmetric (pQCD-like) splitting functions and the symmetric (non-pQCD-like) SD emissions. Furthermore, for jets with later $\tau_{\mathrm{form,RSD}}$, the RSD coincides with the moment the sub-leading charged particle ends its partonic-like cascade.

Chapter 5

Conclusions

The purpose of this thesis was set out to be the identification of hadronization timescales in vaccum proton-proton collisions through the means of jet substructure observables sensitive to the transition between perturbative and non-perturbative QCD-regimes. The observables were selected according to their potential in the search for the space-time substructure of jets - the opening angles, the groomed momentum fraction and the formation time -, namely for 3 particle-production processes of interest - the 1^{st} soft-drop emission (1SD), the resolved soft-drop splitting (RSD) and the leading charged particles splitting (LCP). These correspond to the 1^{st} pQCD-like emission in the clustering tree, to the splitting where the 2 leading charged particles are resolved and to the hadron-production process of these 2 leading charges, respectively. Furthermore, substructure was studied with differently-clustered trees throughout the entire thesis, using both the Cambridge/Aachen and the τ jet algorithms. Following Sections summarize the most important results and conclusions achieved in this thesis and also paths forward for future research.

5.1 Achievements

Firstly, it became clear that selecting jets based on the collision centre-of-mass energy and jet transverse momentum strongly determines the substructure features of those jets. These cuts are a means of selecting the resolution scale at which one is probing the proton. Higher center-of-mass energies and softer jets generally select outgoing gluons, while the opposite selects more quarks. Although not providing a full picture of the phenomena at play, it was possible to show that hard scattered gluons tend to initiate wider, more symmetric and earlier jets than quarks.

Afterwards, a study of how substructure evolves along the early stages of the clustering tree was performed for the 1SD, 2SD and 3SD emissions along the main jet branch. It started by showing a desirable angular ordering (prescribed by pQCD) for consecutive emissions and for both algorithms selected for the re-clustering procedure: Cambridge/Aachen and τ. It also showed an increase in the transverse momentum sharing between the two prongs along the clustering tree, with momentum fraction distributions getting significantly flatter for consecutive SD emissions. This behaviour is also consistent with ending the parton shower phase and with approaching hadronization scales. An increase in SD formation time was also revealed, with larger timescales being associated to later SD emissions along the tree. Using C/A re-clustering, the emission formation times are significantly overestimated with respect to τ re-clustering but both of them seem to translate into time-ordered clustering trees.

Then, by counting how many SD emissions are left along the main branch after the RSD splitting, it was found that the majority of RHIC and LHC jets have their RSD up to 2 soft-drop emissions before the end of the clustering tree. However, since these trees usually have a total of 3 to 5 soft-drop emissions, the vast majority of jets also have their RSD splitting close to the beginning of the clustering tree. This leads the RSD to inherit both pQCD-like behaviour, dominant at the 1SD, and non-pQCD-like behaviour, dominant at the LCP. This is shown in the study of the groomed momentum fraction, where the tendentiously flat RSD distributions differ from the manifestly asymmetrical 1SD and from the symmetrical LCP, being nevertheless more similar to the latter.

The following study of the formation times of these 3 splittings helps to understand further the connection between them. The RSD and LCP formation time (τ_{form}) distributions are even shown to practically match in ranges of large τ_{form}, with the ratio between them stabilizing close to unity. To realise what this meant, jet substructure was subsequently analysed by applying cuts in time ranges where the RSD/LCP ratios hit a plateau close to unity for both RHIC and LHC. With this method, it was possible to identify one plateau for RHIC ($\tau_{\text{form}} > 1.3$ fm/c, approximately), while LHC displayed the presence of a double plateau structure ($0.1 < \tau_{\text{form}} < 0.8$ fm/c for the 1^{st} plateau and $\tau_{\text{form}} > 7$ fm/c for the 2^{nd} one, approximately).

Jets selected with large RSD formation time were firstly found to have RSD splittings very close to the end of the clustering tree, while smaller timescales select jets with still a few SD emissions along both prongs. It was verified that higher $\tau_{\text{form, RSD}}$ selects jets with RSD momentum fraction significantly different to 1SD momentum fraction and closer to the LCP one. Finally, it was shown that time selections strongly translate into a sharp separation in jet mass, with higher $\tau_{\text{form, RSD}}$ jets having significantly smaller jet masses. Overall, these substructure studies revealed the 1^{st} LHC plateau to have very similar features with respect to the smallest time cuts for both RHIC and LHC jets, hinting that this plateau is not in fact marking the shift of the RSD splitting towards non-pQCD-like regimes.

To further investigate the nature of the plateaus from the RSD/LCP time-ratios, these ratios were computed separately for jets with RSD splittings coinciding with the 1^{st}, 2^{nd} and 3^{rd} soft-drop emissions along the main tree branch. These showed that the main contributions to the RHIC plateau and the 2^{nd} LHC plateau come from RSD splittings placed higher in the clustering tree, namely after the 2SD emission.

Finally, by comparing how the groomed momentum fraction varies in the vicinity of the RSD splitting, the clearest evidence was found that RSD is a robust flagger of the transition from parton to hadron levels. It firstly showed the SD emission immediately prior to the RSD one to be manifestly asymmetrical, while the RSD itself and the one immediately after it have notably flat momentum fraction distributions, suggesting the RSD is marking the end of tendentiously pQCD behaviour in the clustering tree. This observation is further corroborated afterwards, where jets with RSD=1SD have flat momentum fraction distributions for all 1SD, 2SD and 3SD emissions, while jets with RSD=3SD have asymmetrical 1SD and

2SD emissions.

5.2 Future Work

This thesis is the first to propose the use of jet substructure to tag the transition from partons to final-state particles. The results suggest that the proposed method (selecting late RSD jets) is robust against clustering algorithm choices or parton shower settings. While this method marks the transition, the absolute time at which this transition occurs is still inconclusive and will be left for future studies. In addition, it will be interesting to see how this can be extrapolated to the case of a quark-gluon plasma, where a fast-expanding medium is developing along the shower.

Bibliography

[1] D. Castelvecchi and E. Gibney. Cern makes bold push to build €21-billion supercollider. *Nature*, ISSN 1476-4687, June 2020. doi: https://doi.org/10.1038/d41586-020-01866-9.

[2] T. Muta. *Foundations of Quantum Chromodynamics: An Introduction to Perturbative Methods in Gauge Theories*. World Scientific Lecture Notes in Physics: Volume 5, 1987. doi: https://doi.org/10.1142/0022.

[3] R. Ellis and W. Stirling. *QCD and Collider Physics*. Cambridge University Press, 1996. doi: https://doi.org/10.1017/CBO9780511628788.

[4] K. Johnson. The m.i.t. bag model. *Acta Phys. Polon. B*, 1975. doi: https://doi.org/10.1038/d41586-020-01866-9.

[5] G. Salam. Towards jetography. Eur.Phys.J.C67:637-686, 2010. doi: https://doi.org/10.48550/arXiv.0906.1833.

[6] P. D. Group. *Review of Particle Physics*. Chin.Phys.C 40, 2016. doi: https://doi.org/10.1088/1674-1137/40/10/100001.

[7] C. Bierlich and et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. SciPost, mar 2022. doi: https://doi.org/10.48550/arXiv.2203.11601.

[8] P. Ramond. *Group Theory: A Physicist's Survey*. Cambridge University Press, 2010. doi: https://doi.org/10.1017/CBO9780511781865.

[9] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, 1995. doi: https://doi.org/10.1201/9780429503559.

[10] P. A. Zyla and et al. (Particle Data Group). *Review of Particle Physics*. Progress of Theoretical and Experimental Physics, 2020. doi: https://doi.org/10.1093/ptep/ptaa104.

[11] G. jiong Ni, G. hong Yang, and R. tang Fu. Running coupling constants of fermions with masses in quantum electro dynamics and quantum chromo dynamics. *International Journal of Modern Physics A*, 1999. doi: https://doi.org/10.48550/arXiv.hep-ph/9906364.

[12] A. Deur, S. J. Brodsky, and G. F. de Teramond. *The QCD Running Coupling*. Prog. Part. Nuc. Phys. 90 1, 2016. doi: https://doi.org/10.1016/j.ppnp.2016.04.003.

[13] J. J. Ethier and E. R. Nocera. *Parton Distributions in Nucleons and Nuclei.* Annual Review of Nuclear and Particle Science, Vol.70, 2020. doi: https://doi.org/10.1146/annurev-nucl-011720-042725.

[14] I. Abt. The proton as seen by the hera collider. *Annual Review of Nuclear and Particle Science Vol. 66:377-399*, 2016. doi: https://doi.org/10.1146/annurev-nucl-102115-044544.

[15] J. W. Harris. Introduction to hard scattering processes and recent results from hard probes at rhic and lhc. *Phys.: Conf. Ser. 630 012052*, 2015.

[16] V. Gribov and L. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys. 15*, 1972. doi: https://doi.org/10.1016/0370-2693(71)90576-4.

[17] B. Andersson, G. Gustafson, and B. Söderberg. A general model for jet fragmentation. *Z. Phys. C 20, 317*, 1983. doi: https://doi.org/10.1007/BF01407824.

[18] T. Sjöstrand. Jet fragmentation of nearby partons. *Nucl. Phys. B 248, 469*, 1984. doi: https://doi.org/10.1016/0550-3213(84)90607-2.

[19] M. Bahr and et al. *Herwig++ Physics and Manual.* Eur.Phys.J.C58:639-707, 2008. doi: https://doi.org/10.48550/arXiv.0803.0883.

[20] Z. Liu. Four-quark matter—a new era of spectroscopy. *AAPPS Bulletin volume 31, 8*, 2021. doi: https://doi.org/10.1007/s43673-021-00007-2.

[21] B. Andersson and et al. Is there screwiness at the end of the qcd cascades? *Journal of High Energy Physics*, 1998. doi: https://doi.org/10.1088/1126-6708/1998/09/014.

[22] T. Sjostrand and et al. Pythia 6.4 physics and manual. JHEP 0605:026, 2006. doi: https://doi.org/10.1088/1126-6708/2006/05/026.

[23] A. Larkoski and et al. Soft drop. JHEP 1405 146, 2014. doi: https://doi.org/10.48550/arXiv.1402.2657.

[24] L. Apolinário, A. Cordeiro, and K. Zapp. Time reclustering for jet quenching studies. *The European Physical Journal C volume 81, Article number: 561*, 2021. doi: https://doi.org/10.1140/epjc/s10052-021-09346-8.

[25] M. Cacciari, G. P. Salam, and G. Soyez. Fastjet user manual. Eur. Phys. J. C 72, 2012. doi: https://doi.org/10.1140/epjc/s10052-012-1896-2.

[26] G. Sterman and S. Weinberg. Jets from quantum chromodynamics. *Phys. Rev. Lett. 39, 1436*, 1977. doi: https://doi.org/10.1103/PhysRevLett.39.1436.

[27] S. Catani, Y. Dokshitzer, and B. Webber. The $k_\perp$ clustering algorithm for jets in deep inelastic scattering and hadron collisions. *Physics Letters B, Volume 285, Issue 3*, ISSN 0370-2693, 1992. doi: https://doi.org/10.1016/0370-2693(92)91467-N.

[28] M. Wobisch and T. Wengler. Hadronization corrections to jet cross sections in deep-inelastic scattering. *High Energy Physics - Phenomenology*, 1998. arXiv:hep-ph/9907280.

[29] H. Collaboration, C. Adloff, and et al. Measurement and qcd analysis of jet cross sections in deep-inelastic positron proton collisions at $\sqrt{s} = 300$ gev. *The European Physical Journal C - Particles and Fields volume 19, pages 289–311*, 2001. doi: https://doi.org/10.1007/s100520100621.

[30] M. Cacciari, G. Salam, and G. Soyez. The anti-k$_t$ jet clustering algorithm. *J. High Energy Phys. 0804, 063*, LPTHE-07-03, 2008. doi: https://doi.org/10.48550/arXiv.0802.1189.

[31] A. Collaboration, M. Aaboud, and et al. A measurement of the soft-drop jet mass in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector. *Phys. Rev. Lett. 121, 092001*, CERN-EP-2017-231, 2018. doi: https://doi.org/10.48550/arXiv.1711.08341.

[32] C. Collaboration, A. M. Sirunyan, and et al. Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ tev. *JHEP 11 113*, CMS-SMP-16-010, CERN-EP-2018-180, 2018. doi: https://doi.org/10.48550/arXiv.1807.05974.

[33] A. Collaboration and T. A. collaboration. Measurement of the lund jet plane using charged particles with the atlas detector from 13 tev proton-proton collisions. *Phys. Rev. Lett. 124, 222002*, CERN-EP-2020-030, 2020. doi: https://doi.org/10.48550/arXiv.2004.03540.

[34] C. Frye and et al. Factorization for groomed jet substructure beyond the next-to-leading logarithm. JHEP 07, 064, 2016. doi: https://doi.org/10.48550/arXiv.1603.09338.

[35] S. Marzani, L. Schunk, and G. Soyez. A study of jet mass distributions with grooming. JHEP 07, 132, 2017. doi: https://doi.org/10.48550/arXiv.1704.02210.

[36] Z.-B. Kang and et al. The groomed and ungroomed jet mass distribution for inclusive jet production at the lhc. JHEP 10, 137, 2018. doi: https://doi.org/10.48550/arXiv.1803.03645.

[37] Z.-B. Kang and et al. Soft drop groomed jet angularities at the lhc. *Phys. Lett. B793, 41-47*, 2019. doi: https://doi.org/10.48550/arXiv.1811.06983.

[38] R. Elayavalli. Jet substructure and hadronization with star. 2021.

[39] Y. Dokshitzer and et al. *Basics of Perturbative QCD*. Fong and Sons Printers Pte. Ltd, 1991.

[40] N. Cabibbo and G. Parisi. Exponential hadronic spectrum and quark liberation. *Physics Letters B, vol. 59, no. 1, pp. 67-69*, 1975. doi: https://doi.org/10.1016/0370-2693(75)90158-6.

[41] R. Brun and F. Rademakers. Root: An object oriented data analysis framework. Nucl. Instrum. Meth. A 389, 81, 1997. doi: https://doi.org/10.1016/S0168-9002(97)00048-X.

[42] A. Accardi and et al. Electron-ion collider: The next qcd frontier. *The European Physical Journal A 52, 268*, 2016. doi: https://doi.org/10.1140/epja/i2016-16268-9.

Appendix A

RSD Placement

A.1 Overall Position of the RSD Splitting

Since the RSD splitting has significantly different substructure properties than the 1SD and even later SD emissions, it is important to determine the placement of the RSD splitting in the clustering tree. This is done in 2 ways: through the numbering of the SD emissions until the RSD splitting takes place, using the $\mathrm{N_{RSD}}$ variable, and through the relative placement of the RSD splitting with respect to the end of the clustering tree, using the $\mathrm{N_{post\text{-}RSD}}$ variable (introduced in Section 4.2.1), now only evaluated along the main RSD prong. Then, to extract $\mathrm{N_{RSD}}$, one counts the number of soft-drop emissions along the main jet branch until the resolved one is reached. So, for instance, if for a given jet it is found that $\mathrm{N_{RSD}} = 1$, the first soft-drop approved splitting in the clustering tree corresponds simultaneously to the 1SD emission as well as to the RSD splitting. For $\mathrm{N_{RSD}} = 2$, the RSD splitting coincides with the 2SD emission along the main jet branch and so on and so forth.

To get insight into the RSD placement in the clustering tree for given sets of kinematic conditions, Figures A.1.1 and A.1.2 show the 2-dimensional correlation histograms for the total number of jet SD splittings, $\mathrm{N_{splits}}$, found along the main branch of the clustering tree versus the number of SD emissions that come after the RSD splitting, $\mathrm{N_{post\text{-}RSD}}$, also along the main branch, until the clustering tree ends. Figure A.1.1 is based on $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV RHIC pp collisions, while Figure A.1.2 refers to $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV LHC pp collisions. The left panels used C/A re-clustering, while the right ones used τ re-clustering. The z-axis scale is given by the colour bar presented on the rightmost part of the Figures.

These histograms are very clearly limited from above by the $\left(\mathrm{N_{post\text{-}RSD}}\right) = \left(\mathrm{N_{splits}}\right) - 1$ line or, equivalently, $\mathrm{N_{RSD}} = 1$, since the RSD splitting cannot, by definition, take place before the 1SD emission. This condition represents the limit case where the RSD precisely coincides with the 1SD, with all the two-dimensional bins below it representing jets were the RSD splitting happens at later stages of the clustering tree.

Figures A.1.1 and A.1.2 reveal that the preferred placement for the RSD splitting is precisely the 1SD emission, with the highest probability density values (warmest colours) being achieved for the upper

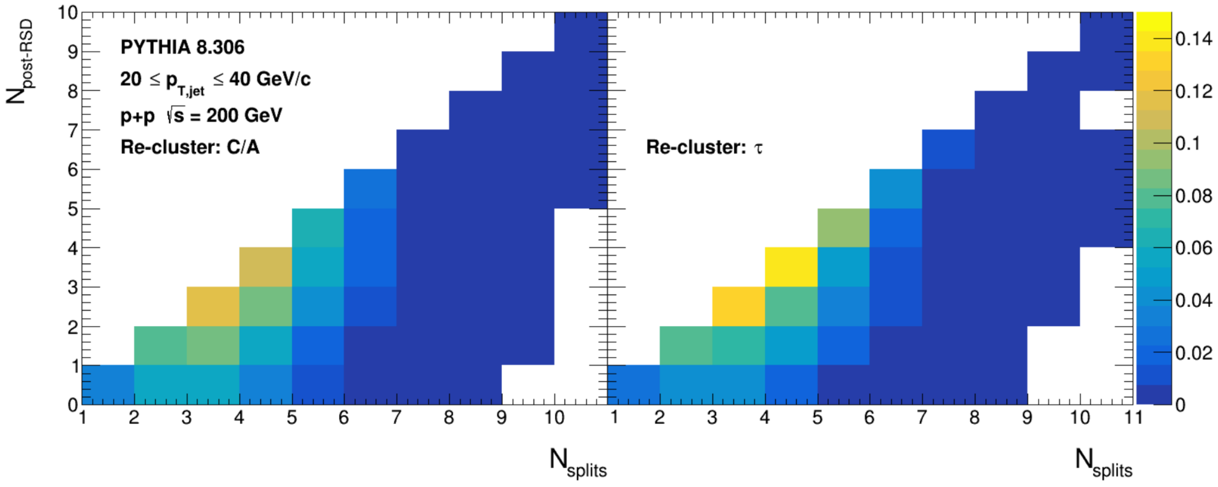


Figure A.1.1: Two-dimensional histograms for the $\mathrm{N}_{\text{splits}}$ versus $\mathrm{N}_{\text{post-RSD}}$ jet distributions, evaluated for jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c produced in pp collisions with energies of 200 GeV, represented by the color-scheme outlined in the color bar on the rightmost side of the Figure; left and right panels show these distributions found for C/A and τ clustering trees, respectively.

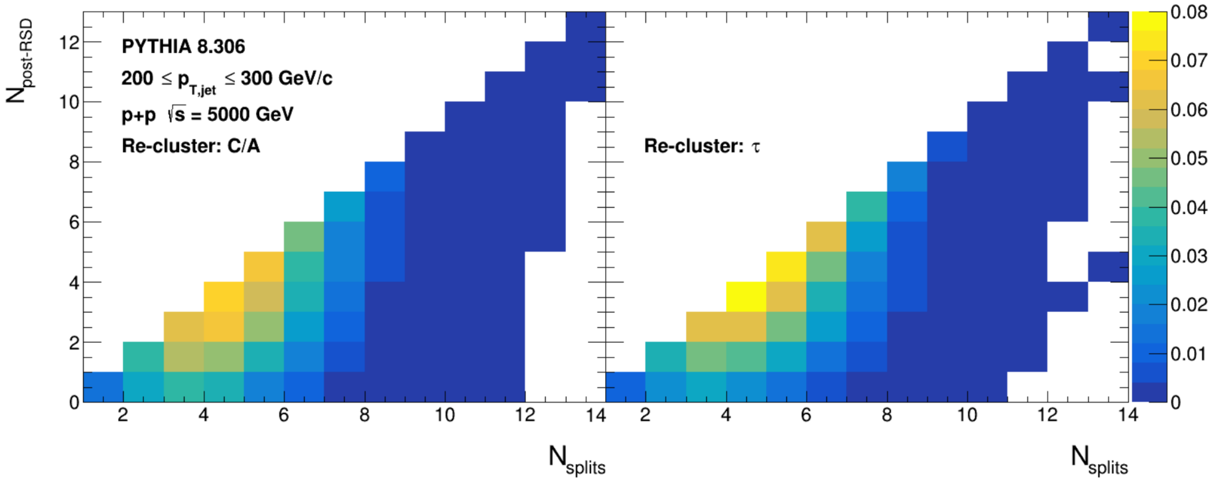


Figure A.1.2: Two-dimensional histograms for the $\mathrm{N}_{\text{splits}}$ versus $\mathrm{N}_{\text{post-RSD}}$ jet distributions, evaluated for jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c produced in pp collisions with energies of 5 TeV, represented by the color-scheme outlined in the color bar on the rightmost side of the Figure; left and right panels show these distributions found for C/A and τ clustering trees, respectively.

bound $\left(\mathrm{N}_{\text{post-RSD}}\right) = \left(\mathrm{N}_{\text{splits}}\right) - 1$. For instance, looking only at τ re-clustering, approximately 15% of all $20 < p_{\mathrm{T,jet}} < 40$ GeV/c RHIC jets and 8% of all $200 < p_{\mathrm{T,jet}} < 300$ GeV/c LHC jets have a total of 4 soft-drop splittings along the main branch and the RSD splitting coinciding with the 1SD emission (yellowest bin).

However, there are still significant portions of the jet population for which the RSD splitting sits at higher positions of the clustering tree. Figure A.1.1 shows that RHIC jets can reach a total of 10 soft-drop emissions along the main branch and have their RSD splitting placed up to the 9th soft-drop emission. As for LHC jets, Figure A.1.2 shows they can go up to a total of 13 soft-drop emissions along the main branch and have their RSD splittings taking place all the way up to the 11th SD emission, for C/A clustering trees, or the 10th SD emission, for τ clustering trees.

A.2 Cuts in Formation Time

It is also relevant to check how the RSD placement varies according to jet selections based on RSD formation time cuts (see (4.1), (4.2) and (4.3)). Therefore, Figures A.2.1 and A.2.2 show the 2-dimensional correlation histograms for N$_{\text{splits}}$ versus N$_{\text{post-RSD}}$, the first one for the RHIC drop and “plateau” ranges of the RSD/LCP ratio and the second for the LHC drop, 1st “plateau” and 2nd “plateau” ranges of the RSD/LCP ratio.

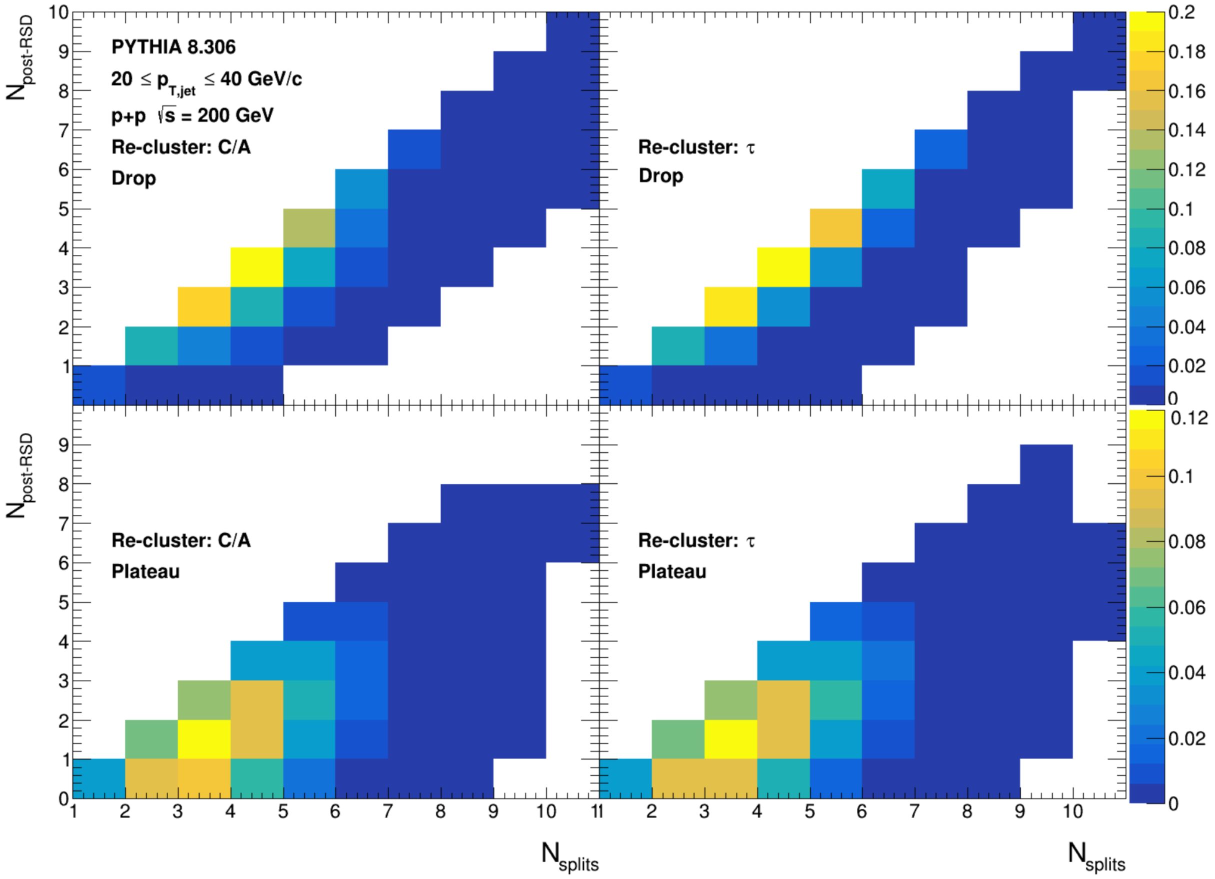


Figure A.2.1: Two-dimensional histograms for the N$_{\text{splits}}$ versus N$_{\text{post-RSD}}$ jet distributions, evaluated for jets with $20 < p_{\text{T,jet}} < 40$ GeV/c produced in pp collisions with energies of 200 GeV, represented by the color-scheme outlined in the color bar on the rightmost side of the Figure; left and right panels show these distributions found for C/A and τ clustering trees, respectively; top panels have a RSD formation time cut for the drop range of the RSD/LCP ratio and the bottom ones for the plateau range.

Figure A.2.1 shows that small $\tau_{\text{form, RSD}}$ jets have their RSD splitting very early on, namely on the 1SD emission for the vast majority of them. These jets still go up to 10 soft-drop emissions in total, but the RSD is limited to the first 5. However, it also shows that, overall, large $\tau_{\text{form, RSD}}$ jets have their RSD splitting later in the clustering tree, mostly located at the 2SD and 3SD emissions.

Figure A.2.2 reinforces that jets with small $\tau_{\text{form, RSD}}$ (the ones in the drop and 1st “plateau” ranges) have very early RSD splittings. In the drop range, trees can go up to 13 soft-drop emissions in total, but the RSD splitting only happens for the first 4 ones. Also notable is the similarity between the LHC plots at the 1st “plateau” and the RHIC plots at the drop range. This shows indeed that LHC’s 1st “plateau” is unrelated to the proximity of the RSD splitting to hadronization. For the 2nd LHC “plateau”, plots are also very similar to the ones for the RHIC "plateau", showing that most high $\tau_{\text{form, RSD}}$ jets have their RSD splitting at the end of the clustering tree or 1 soft-drop emission away from it.

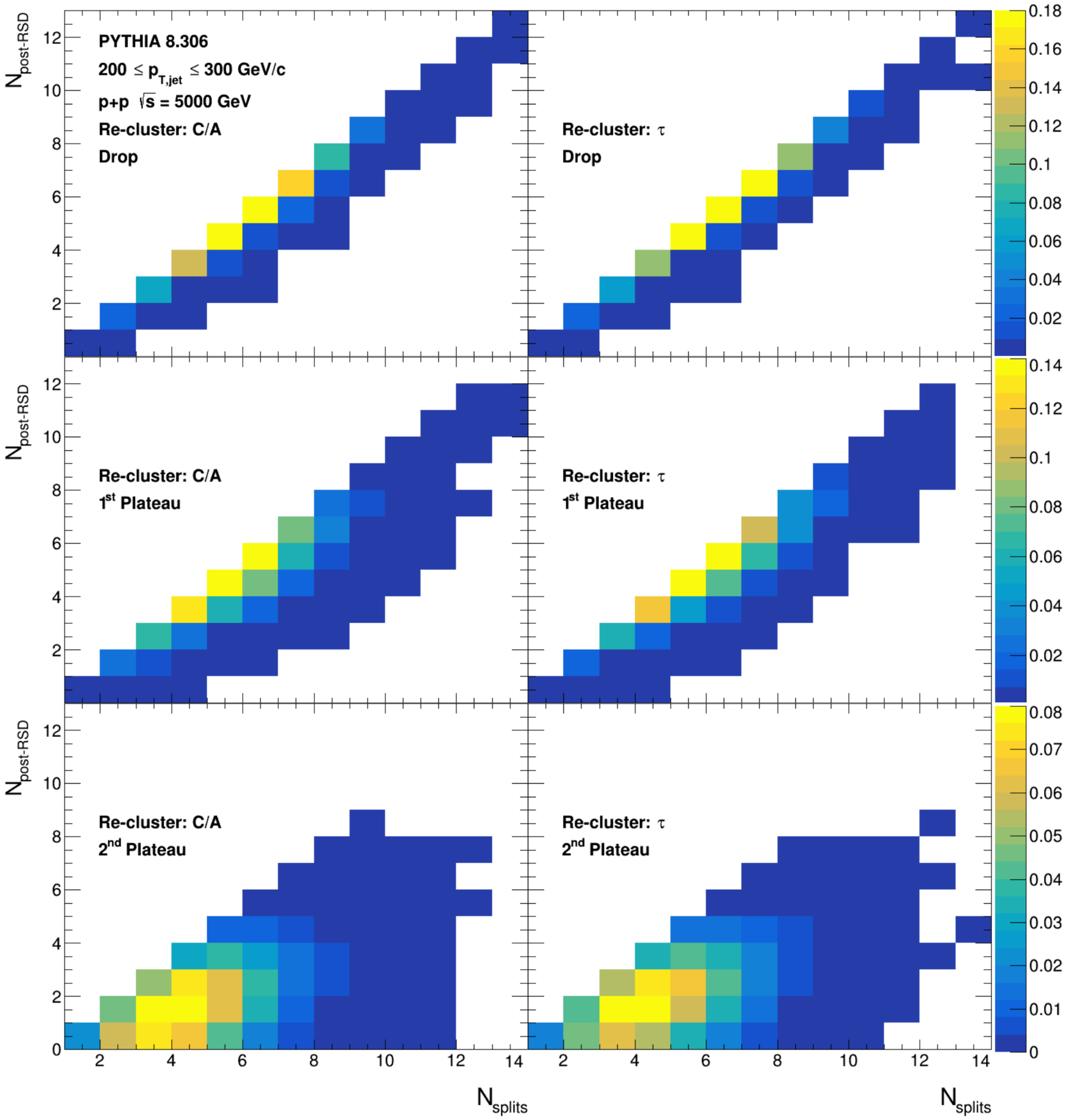


Figure A.2.2: Two-dimensional histograms for the N_{splits} versus $\text{N}_{\text{post-RSD}}$ jet distributions, evaluated for jets with $200 < p_{\text{T,jet}} < 300$ GeV/c produced in pp collisions with energies of 5 TeV, represented by the color-scheme outlined in the color bar on the rightmost side of the Figure; left and right panels show these distributions found for C/A and τ clustering trees, respectively; top panels have a RSD formation time cut for the drop range of the RSD/LCP ratio, middle one for the 1^{st} plateau range and the bottom ones for the 2^{nd} plateau range.

Appendix B

Splitting Opening Angle

This Appendix focuses specifically on the opening angles of each of the 3 splittings studied in Chapter 4 - 1SD, RSD and LCP - and how they compare with each other. The opening angle is introduced in Equation 2.17 and is evaluated for each splitting using the *delta_R()* method in the *FastJet* framework.

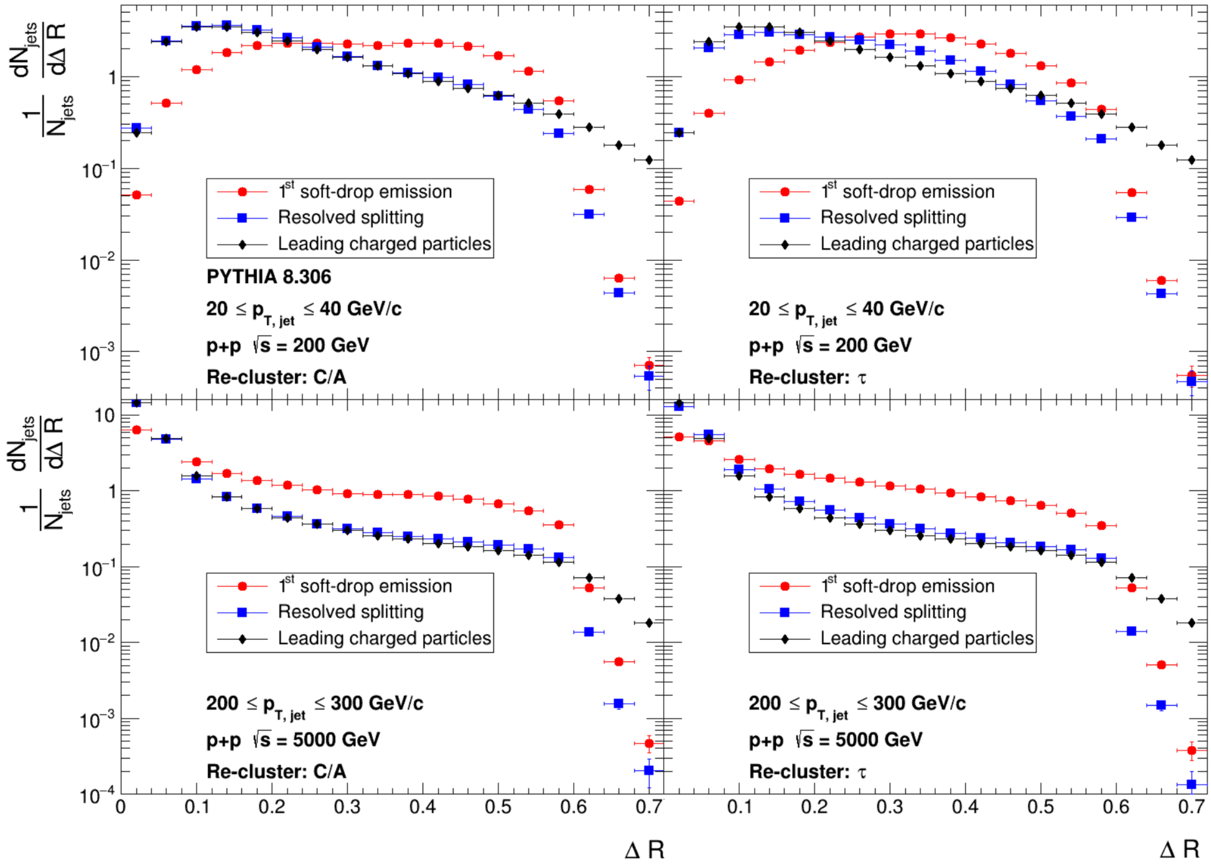


Figure B.0.1: Opening angle (ΔR) distributions for splittings from jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c produced in pp collisions with energies of 200 GeV, in the top panels, and from jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c produced in pp collisions with energies of 5 TeV, in the bottom ones; ΔR distributions are represented in red circle, blue square and black diamond markers for the 1st soft-drop emission, the resolved soft-drop splitting and the leading charged particles splitting, respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively.

Figure B.0.1 shows the ΔR distributions for the 1SD, RSD and LCP splittings, represented in red circle, blue square and black diamond markers, respectively. The top panels have these distributions for jets falling in the $20 < p_{\text{T,jet}} < 40$ GeV/c range from $\sqrt{s} = 200$ GeV pp collisions, while the bottom ones show them for $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions. Jets are re-clustered using the C/A and τ algorithms on the left and right panels, respectively.

The comparison between the 3 splittings shows, for all panels of Figure B.0.1, a narrowing of the splittings from the 1SD to the RSD and then to the LCP splittings. However, the RSD ΔR distributions are relatively similar in shape to the LCP ones, which does not happen with respect to to the 1SD. This indicates that, when it comes to angular openings, RSD appears to be more like the hadronic LCP splitting than it does with respect to the generally partonic 1SD emission, signalling the RSD splitting to be significantly less pQCD-like than the earliest soft-drop emission (and even than the next 2 soft-drop emissions shown in 3.4.1).

Appendix C

Charge Studies

The choice of the 2 leading charged particles (LCP) for the substructure studies performed in Chapter 4 is motivated by the straightforward detection of electrically charged particles in the calorimeters from RHIC and LHC colliders. The impact of this charge-requirement on the leading particles is studied in C.1 in the context of formation time (τ_{form}), comparing temporal features between the LCP and the (no charge-requirement) leading particles splitting (LP). The τ_{form}-dependency on the sign of the leading charges is shown afterwards in C.2.

C.1 Leading Particles Charge-Requirement

Figure C.1.1 shows, on the top panels, the τ_{form} distributions for the LCP (already shown in Chapter 4) and for the new LP splitting, represented by diamond and square markers, respectively. The bottom panels show the LP/LCP ratios The panel. Distributions on the left panels sample $20 < p_{\text{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions and the ones on the right sample $200 < p_{\text{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions.

From Figure C.1.1, one can observe for both kinematic settings that LP distributions dominate for very small and for large τ_{form}, while intermediate values are dominated by the LCP splitting. Most notably, these ratio plots show that LP distributions do not have the same shape as LCP distributions for the "plateau" range of the RHIC RSD/LCP ratio (in (4.1)) and for the 2nd "plateau" range of the LHC RSD/LCP ratio (in (4.2) and (4.3)). In fact, the LP/LCP ratios are not constant for these ranges of interest, meaning these "plateaus" could be significantly deformed if they were looked for in RSD/LP ratios. Nevertheless, the 1st "plateau" of the LHC RSD/LCP ratio (also in (4.2) and (4.3)) would continue to be a plateau for the RSD/LP ratio, although with ratios even more above unity. One would also, however, have to get the RSD distributions when RSD is considered to resolved the LP and not the LCP, which could compensate this differences.

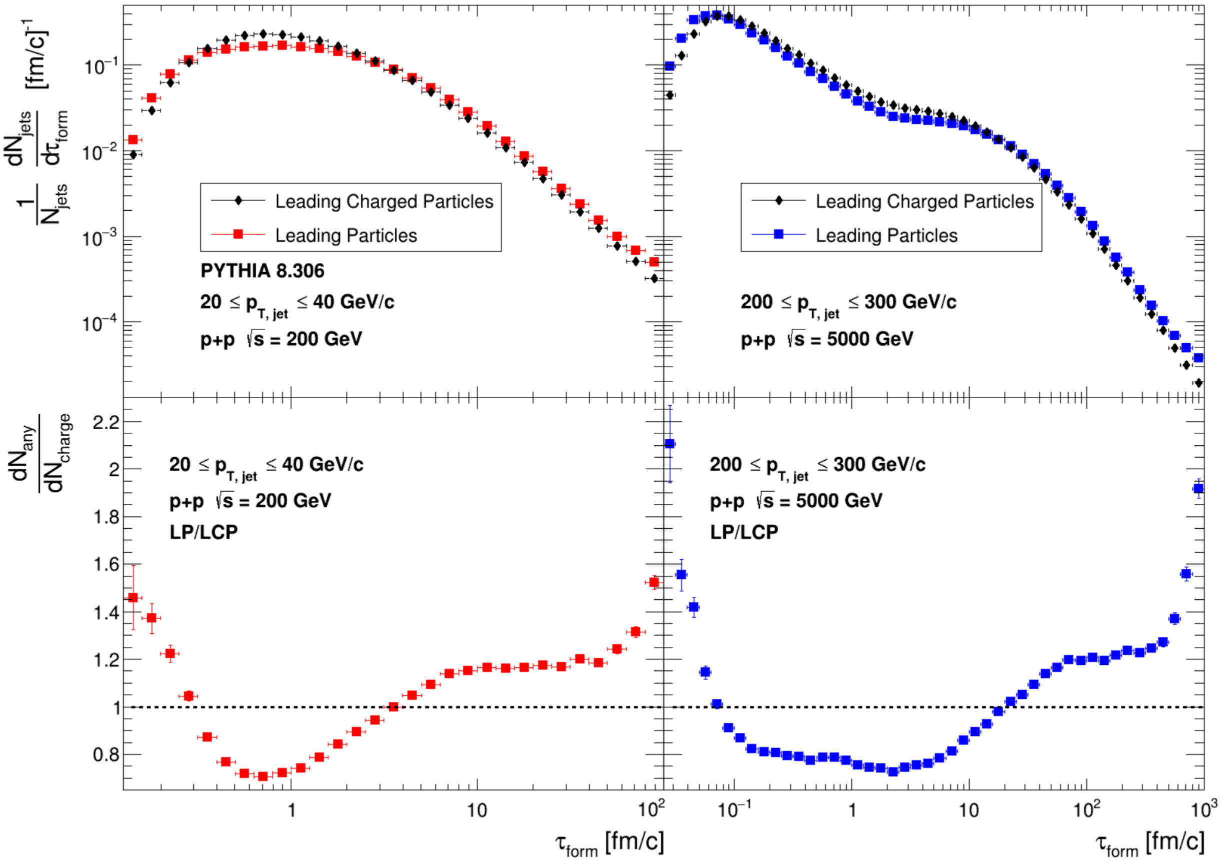


Figure C.1.1: Formation time distributions for the leading charged particles splitting (LCP) and for the leading particles splitting (LP), shown on the top panels separately through diamond and square markers and on the bottom panels via their ratio (LP/LCP); the left panels show them for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions (LCP in black and LP in red) and the right panels show them for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions (LCP in black and LP in blue).

C.2 Charges of the 2 Leading Charged Particles

Looking now at the charge signs of the leading charged particles, Figure C.2.1 shows the $\tau_{\text{form, LCP}}$ distributions for jets with generic LCP, with equally-charged LCP and with oppositely-charged LCP (normalized with respect to the first one), represented by black diamond, red circle and blue square markers, respectively. Distributions on the left panel sample $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions and the one on the right $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions. The blue and red distributions naturally add up to the black one.

The $\tau_{\text{form, LCP}}$ distributions from Figure C.2.1 show that, overall, the two leading charged particles from a jet are much more likely to have the opposite charge signs than to have the same one. Furthermore, jets with oppositely-charged LCP have on average larger $\tau_{\text{form, LCP}}$ than jets with equally-charged LCP. To investigate the $\tau_{\text{form, LCP}}$-dependency of the fraction of jets with equally-charged and oppositely-charged LCP, Figure C.2.2 shows the ratios between these distributions with respect to the generic $\tau_{\text{form, LCP}}$ distribution.

The ratios in Figure C.2.2 show that small $\tau_{\text{form, LCP}}$ sees an equitable fraction of jets with equally- and oppositely-charged LCP, but higher $\tau_{\text{form, LCP}}$ quickly translates into a high predominance of jets with

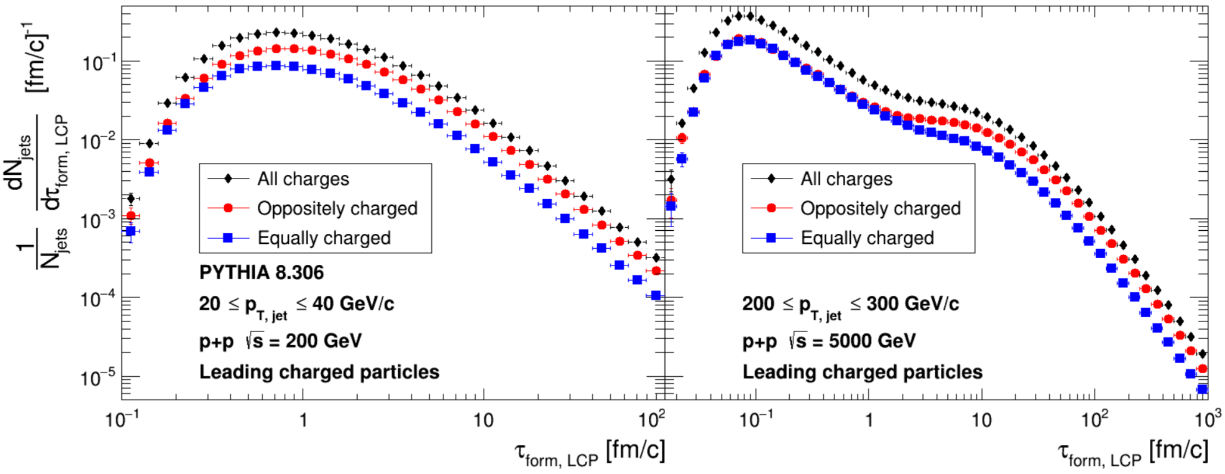


Figure C.2.1: Formation time distributions for the generic leading charged particles splitting (LCP), for the equally-charged leading charged particles splitting and for the oppositely-charged leading charged particles splitting, represented by black diamond, red circle and blue square markers, respectively; the left panel shows them for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions and the right panel shows them for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions; the generic LCP distribution is self-normalized, while the others are normalized with respect to it.

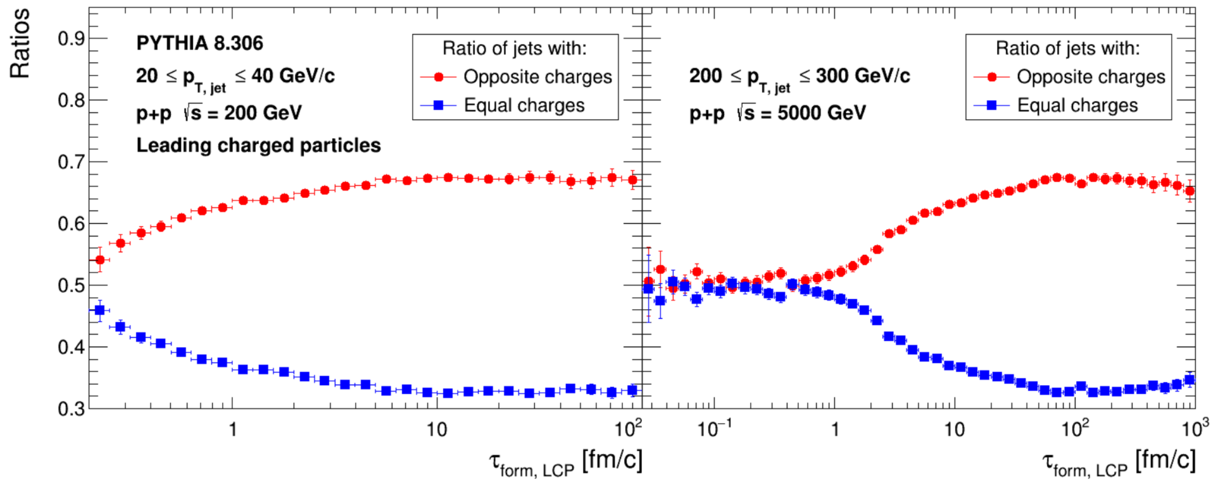


Figure C.2.2: Ratio of the LCP formation time distributions for jets with equally-charged leading charged particles and oppositely-charged leading charged particles with respect to the generic LCP distribution, represented by red circle and blue square markers, respectively; the left panel shows them for $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions and the right panel shows them for $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions.

oppositely-charged LCP. The fraction of jets with equally- and oppositely-charged LCP seems to stabilize for high $\tau_{\text{form, LCP}}$ around $1/3$ and $2/3$, respectively. This is generally the case for the “plateau” range of the RHIC RSD/LCP ratio and for the 2nd “plateau” range of the LHC RSD/LCP ratio.

Appendix D

Groomed Momentum Fraction for Formation Time Cuts

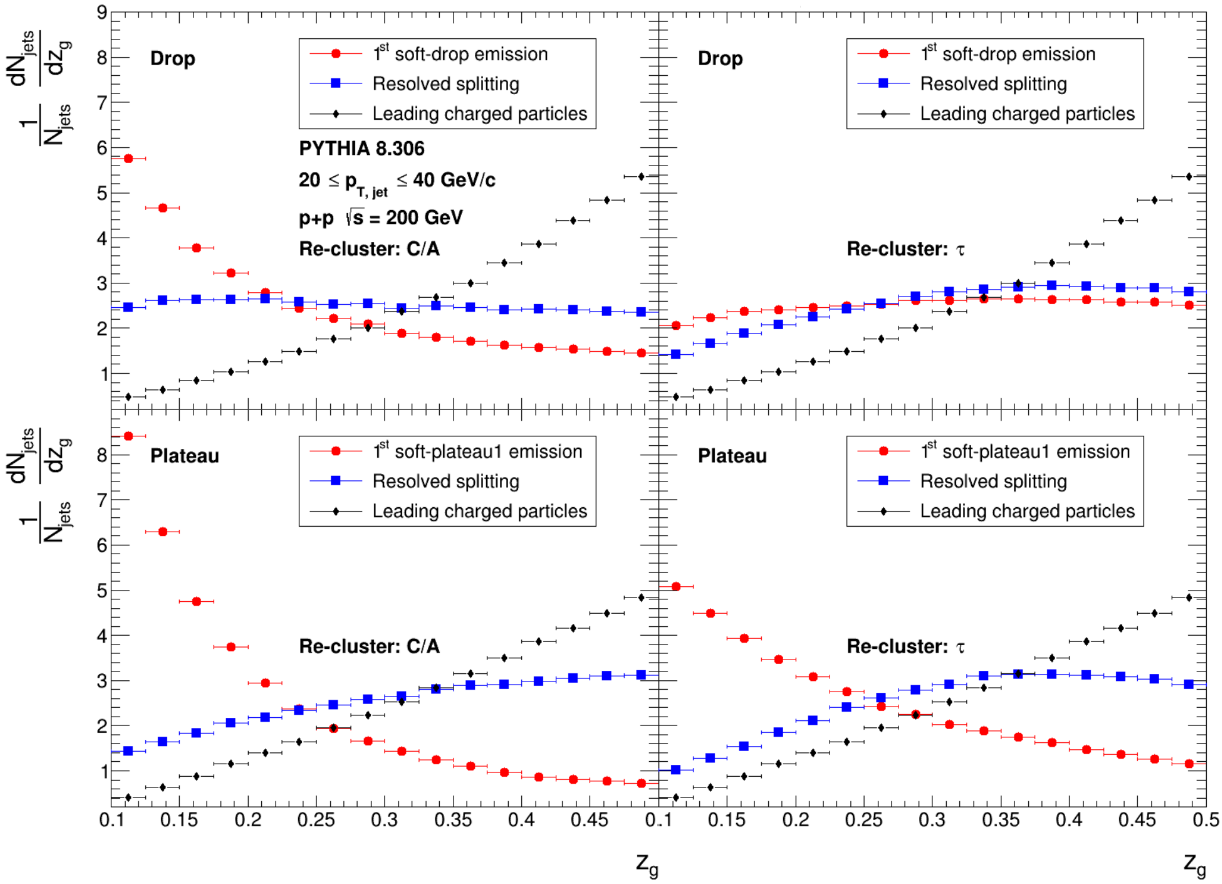


Figure D.0.1: Groomed momentum fraction distributions (z_g) for splittings from jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c produced in pp collisions with energies of 200 GeV; z_g distributions are represented in red circle, blue square and black diamond markers for the 1st soft-drop emission (1SD), the resolved soft-drop splitting (RSD) and the leading charged particles splitting (LCP), respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively; top and bottom panels sample jets with $\tau_{\text{form, RSD}}$ on the drop and plateau ranges of the RSD/LCP time-ratios, respectively.

Figures D.0.1 and D.0.2 show the groomed momentum fraction distributions evaluated at the 1SD, RSD and LCP splittings, represented by red circle, blue square and black diamond markers respectively, sampling jets from different RSD formation time ranges. Figure D.0.1 shows, on the top panels, the z_g distributions for RHIC jets with $\tau_{\text{form, RSD}}$ on the drop range of the RSD/LCP time-ratio, while the bottom panels show them for jets with $\tau_{\text{form, RSD}}$ on the "plateau" range. On the other hand, Figure D.0.1 shows, on the top, middle and bottom panels, the z_g distributions for LHC jets with $\tau_{\text{form, RSD}}$ on the drop, 1^{st} plateau and 2^{nd} plateau ranges of the RSD/LCP time-ratio, respectively.

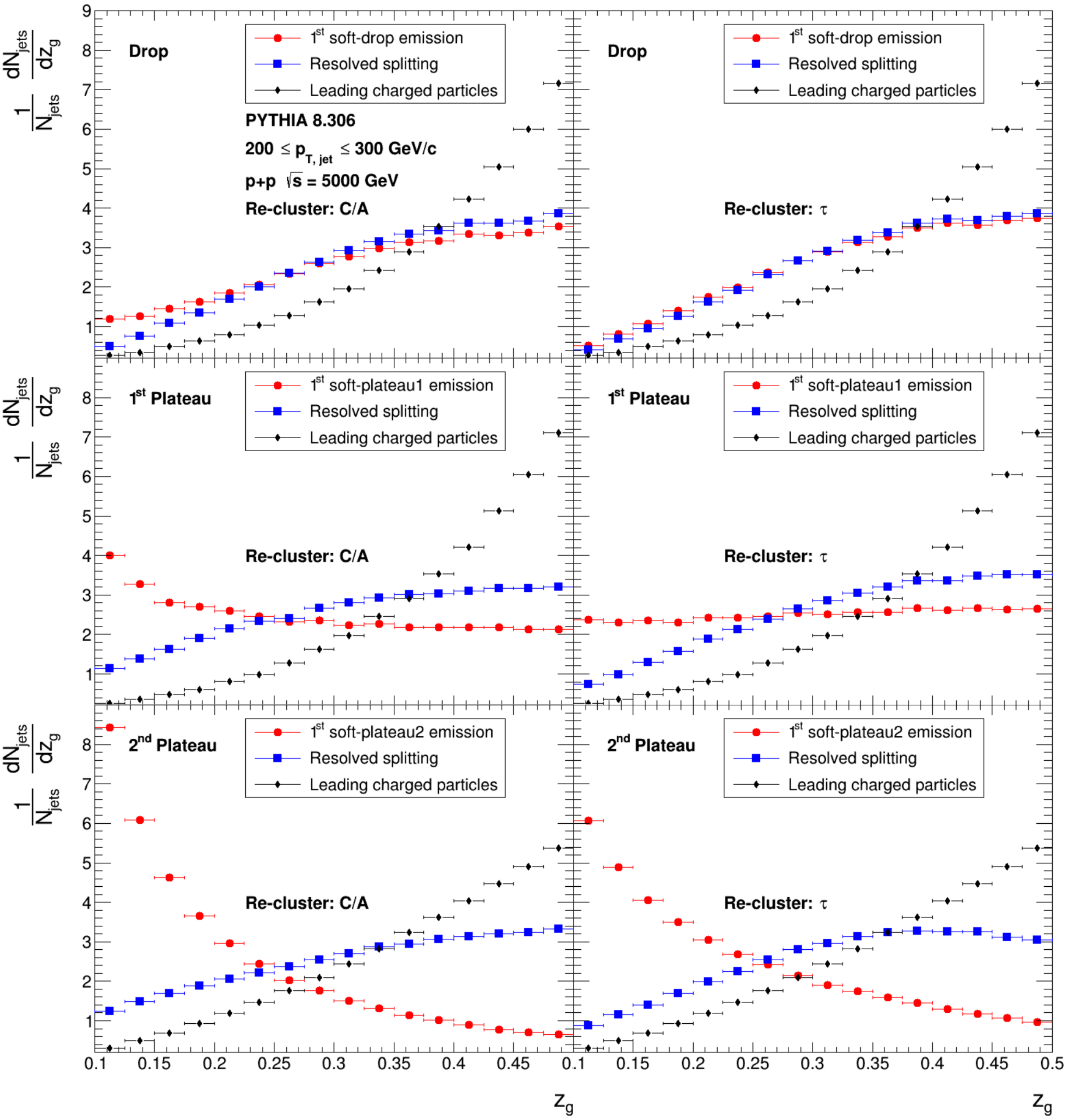


Figure D.0.2: Groomed momentum fraction distributions (z_g) for splittings from jets with $200 < p_{\text{T,jet}} < 300$ GeV/c produced in pp collisions with energies of 5 TeV; z_g distributions are represented in red circle, blue square and black diamond markers for the 1st soft-drop emission (1SD), the resolved soft-drop splitting (RSD) and the leading charged particles splitting (LCP), respectively; left and right panels show these distributions found for C/A and τ re-clustered jets, respectively; top, middle and bottom panels sample jets with $\tau_{\text{form, RSD}}$ on the drop, 1^{st} plateau and 2^{nd} plateau ranges of the RSD/LCP time-ratios, respectively.