Jet Substructure Tools to Identify Hadronization Timescales

Nuno Olavo Gonçalves Mendes Madureira
nuno.olavo@tecnico.uliboa.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2022

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 - hadrons - lie within the npQCD regime. In order to investigate the transition from partons to hadrons - hadronization -, this work 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. This document studies jet substructure via the groomed momentum fraction, measuring an emission's symmetry in transverse momentum space, and 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

1. Introduction

The Standard Model (SM) of particle physics is a theory that describes the strong and the electroweak forces. Quantum Chromodynamics (QCD) is the quantum field theory (QFT) of the strong interaction, describing the color sector of the SM [1]. 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 [2].

Hadron and heavy-ion collisions are used to gain insight at the partonic 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) [3]. The concept of "hardness" is used as a qualitative scale based on the particle energy. Such experiments are performed at the Relativistic Heavy-Ion Collider (RHIC) from the Brookhaven National Laboratory (BNL) and at the Large Hadron Collider (LHC) from the European Organization for Nuclear Research (CERN). The typical collision energies at RHIC are $\sqrt{s} = 0.2$ TeV, at which both proton-proton (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.

However, partonic degrees of freedom cannot be measured directly in experiments. The process by which independent partonic degrees of freedom get coupled, resulting in confinement and hadron production, is what is designated as hadronization [4], marking the transition of QCD from its perturbative regime (pQCD) to its non-perturbative one (npQCD). Hadrons are the particles that eventually have their energies measured by the detectors.

Hadronization is the main focus of this work. Since it takes place at low momentum scales, the physics of hadronization cannot be derived from first principles. Therefore, this work attempts to gain insight into the mechanics of hadronization by grouping the particle aftermath of pp collisions into groups called jets. Jets are structures containing information about a collection of final-state particles that allows the recreation of its evolutionary history by recovering the steps originating each of the intermediate multi-particle states.

2. Background

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 coupling constant of the QFT (a measure of the strength of the interaction at a given scale

Q) is small. However, QCD's coupling constant $\alpha_{QCD}(Q)$ is prohibitively large for small Q scales for perturbative calculations to be performed with any degree of success. These scales include the hadronic ones, which leaves the problem of hadronization out of the reach of perturbation theory. The perturbative running coupling of QCD, calculated at one-loop, is given by

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

where $\beta_0 = 11 - \frac{2}{3}N_f$, with N_f the number of quark flavours active at scale Q^2 and Λ_{QCD} the Landau pole of QCD, at which the coupling constant diverges [5]. 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 running coupling of QCD in Equation (1) shows that partons are asymptotically free particles at large momenta (small distances) - pQCD regime -, while small momenta (large distances) makes partons become tightly bound in highly non-perturbative processes - npQCD regime.

Luckily, these ultra-relativistic collisions allow the factorization between their high-momenta regime, where particle-evolution is modeled by the so-called parton shower, and their low-momenta regime, encoding all the complexity of hadronic physics. The parton shower is an attempt to codify the higher than leading-order (LO) physics of the immediate aftermath of a hard process by creating a partonic-emission path across the full range of the pQCD regime. It links a hard scattered parton to a multi-parton final-state by a succession of $1 \to 2$ splittings, called parton branchings, obeying the QCD Feynman rules.

The branching probabilities are governed by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting kernels [7], $P_{a\to bc}$. The possible branchings in a parton shower are: quark into quark and gluon ($q \to q + g$), gluon into gluon pair ($g \to g + g$) and gluon into quark-antiquark pair ($g \to q + \bar{q}$). Therefore, the possible DGLAP (un-regularized) splitting functions (calculated to LO) are given by

$$\begin{aligned} P_{q\to qg} &= \frac{4}{3}\frac{1+z^2}{1-z} \\ P_{g\to q\bar{q}} &= \frac{1}{2}\left(z^2 + (1-z)^2\right) \\ P_{g\to gg} &= 3\frac{(1-z(1-z))^2}{z(1-z)} \end{aligned} \qquad (2)$$

where z is the fraction of the 4-momentum of the mother-parton carried by its softest daughter.

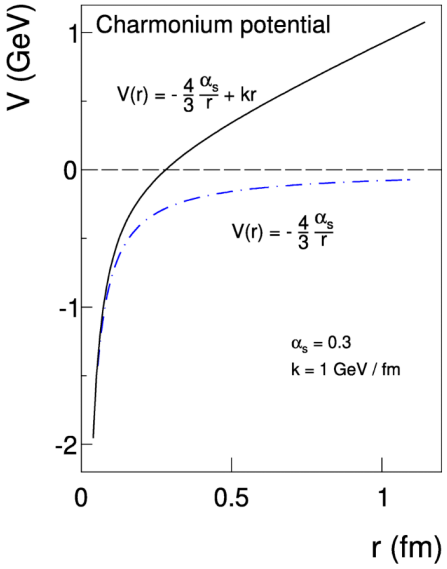

Figure 1: Typical QCD potential calculated at LO for a charmonium system (bound state of charm, c, and anti-charm, $\bar{c}$, quarks via strong interaction) taken from [8].

When it comes to hadronization, its treatment in this work follows the Lund String model. It is based on the assumption that the confining force field between a colour and an anticolour charge, such as the one in a color-singlet $q\bar{q}$, can be viewed as a flux tube with potential energy increasing linearly with the inter-charge distance, much like a string (shown in Figure 1 when neglecting the Coulomb term).

3. Jets

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 subsequently produced is what is called a jet [4].

Multiple algorithms were developed to group collections of final-state particles into jets. A jet clustering algorithm is a set of rules that dictate the order by which one clusters the particles, pairing them up according to their kinematic information. An important algorithm parameter is the jet radius R, which defines the maximum *angular* reach of the jet algorithm. 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 first-order 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 their substructure, where one searches for

the effects of hadronization.

The most commonly used jet algorithms (and the ones employed in this work) are the sequential recombination algorithms. They go beyond just finding jets and implicitly assign a clustering sequence to them [4], mimicking the heuristic parton branching pictures given by the DGLAP functions in Equations (2). They treat jet constituents at the pair-level by quantifying the probability of two particles having the same origin using a certain distance measure d_{ij} given by

$$d_{ij} \overset{\text{def}}{=} \min\left(p_{\text{T},i}^{2p}, p_{\text{T},j}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2} \quad , \tag{3}$$

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

$$\Delta R_{ij} \overset{\text{def}}{=} \sqrt{\left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2} \quad , \tag{4}$$

with $y \overset{\text{def}}{=} \frac{1}{2} \ln\left(\frac{E+p_L}{E-p_L}\right)$ the rapidity of a particle with energy E and longitudinal momentum p_L (with respect to the beam direction) and ϕ its azimuthal angle. In the high-energy limit approximation, for which quarks are assumed massless, rapidity becomes a new variable called the pseudorapidity, defined as $\eta \overset{\text{def}}{=} -\ln\left(tan\left(\frac{\theta}{2}\right)\right)$, where θ is the angle of the particle's 3-momentum with respect to the beamline.

The p is an algorithm-defining parameter that selects the Cambridge/Aachen (C/A) algorithm for $p = 0$, which clusters particles independently of particle momenta and therefore produces purely geometrical orderings, and the anti-k_t algorithm for $p = -1$, which favours clustering with at least one hard particle. Furthermore, setting $p = 0.5$ selects the τ algorithm [9], which will be discussed shortly.

One of the most recent techniques applied to jet analysis is the jet grooming. Jet grooming usually results in the removal of soft wide-angle 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 (DGLAP splitting functions), since they clean the jets from soft contamination (mainly npQCD effects). The most commonly used grooming algorithm and the one incorporated in this work is the soft-drop (SD) algorithm [10].

The SD algorithm works with jets identified using anti-k_t and re-clustered using the C/A algorithm. Then, following the main branch (branch of highest transverse momentum), the soft-drop criterion, expressed in Equation (5), is recursively applied to check, at each de-clustering step, if soft emissions occur, flagging them to be dropped if they do [10]. The SD criterion is expressed by

$$\frac{\min\left(p_{\text{T}1}, p_{\text{T}2}\right)}{p_{\text{T}1} + p_{\text{T}2}} > z_{cut} \left(\frac{\Delta R_{12}}{R}\right)^{\beta} \quad , \tag{5}$$

with ΔR_{12} the *angular* distance defined in Equation (4).

The fixed parameters z_{cut} (soft threshold) and β (angular exponent) regulate the severeness of the SD condition and, therefore, the effectiveness of the jet grooming. For this work, one sets $\beta = 0$, 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. One also sets $z_{cut} = 0.1$, 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 groomed jet can be characterized by the groomed jet momentum fraction z_g, given by

$$z_g \overset{\text{def}}{=} \frac{\min\left(p_{\text{T}1}, p_{\text{T}2}\right)}{p_{\text{T}1} + p_{\text{T}2}} \quad . \tag{6}$$

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 a SD emission.

SD grooming has been shown very recently in [9] to allow for the association of a timescale to the jet splittings that build up the fragmentation pattern. The concept of formation time was firstly developed as a jet analysis tool in the paper [9] and was recently measured at RHIC by the STAR collaboration [11]. The formation time τ_{form} is the time that a quantum state, such as a parton, takes to behave as two independent sources of additional radiation [12], like the 2 daugther-partons from a splitting. Applying the high-energy limit approximation, splitting formation time can be estimated as

$$\tau_{\text{form}} \sim \frac{1}{2\ E\ z\ (1-z)\ \ (1-\cos\theta_{12})} \quad , \tag{7}$$

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 2 daughters and $z = \frac{\min(E_1, E_2)}{E_1 + E_2}$ is the energy fraction.

From the contribution of the total energy to the denominator in Equation (7), it comes that the more energetic a parton is, the earlier it splits into 2 other partons. Then, since parton showers are p_{T}-ordered, splitting formation time has great potential to serve as a shower-ordering variable. This is, in fact, what happens for the τ algorithm. In the high-energy, soft ($z << 1$) and collinear ($\theta_{12} << 1$) limits, the main distance measure $d_{12}^{p=0.5}$ of the τ algorithm becomes a proxy for τ_{form}^{-1}. Clusterings that maximize the formation time are favoured over others, which will be very useful in the search for the hadronization timescales. This work extends SD-grooming to jets re-clustered with the τ algorithm.

4. Implementation

In particle physics, a hard scattering performed in a collider like RHIC or the LHC is called an event. In order to develop and guide new research in this field, many physicists are dedicated to develop numerical algorithms that apply theoretical principles and experiment-based optimizations to generate simulated events. Such a framework of algorithms is called an event generator. The event generator used in this work is PYTHIA 8.306 [13]. This program is a standard tool for the generation of high-energy collision events.

When simulating pp events with PYTHIA, the user has to set a number of minimal initialization parameters. In this work, PYTHIA is alternately set (in the CM frame) with $\sqrt{s} = 0.2$ TeV (RHIC collisions) or $\sqrt{s} = 5$ TeV (LHC collisions). Another input parameter is a minimum transverse momentum requirement $\hat{p_T}^{min}$ that the user imposes on the outgoing hard scattering partons to enhance sampling efficiency on the jet transverse momentum ($p_{\mathrm{T,jet}}$) ranges of interest. One sets $\hat{p_T}^{min} = 16$ GeV/c when sampling jets with $20 < p_{\mathrm{T,jet}} < 40$ GeV/c (from RHIC collisions) and $\hat{p_T}^{min} = 170$ GeV/c when sampling jets with $200 < p_{\mathrm{T,jet}} < 300$ GeV/c (from LHC collisions). Furthermore, $2 \cdot 10^6$ pp collisions are simulated for each setup of kinematic conditions and all QCD $2 \to 2$ quark/gluon scattering processes (excluding heavy-quarks starting from charm [13]) are enabled.

The 4-momentum and the electric charge of the particles produced in the simulated event is provided by PYTHIA. This particle information is then transferred to the jet finder package *FastJet* [14], which provides extensively tested implementations of the clustering algorithms described in Section 3. The initial jet-finding clustering is performed with the anti-k_t jet algorithm and requires only one parameter: the jet radius, set to $R = 0.6$ for all simulation settings. Anti-k_t is chosen due to its ability of 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 from outside the $-1 < \eta_{\mathrm{jet}} < 1$ interval.

However, 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, the jet particles are submitted to 2 distinct re-clustering processes, one using the C/A algorithm and the other using the τ algorithm, both with jet radius set to $R_{\mathrm{new}} = 1.0$ to guarantee the inclusion of particles in the jet borders. 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 z_g. However, the τ algorithm performs better at identifying correctly the formation time, since it is built around τ_{form} being (roughly) its ordering variable. The re-clustered jets are then groomed by the soft-drop algorithm.

All the graphical results are plotted using the ROOT [15] framework. Histograms presented in Section 5 are self-normalized and bin contents are divided by the width of the respective bins.

5. Results

To advance the study of jet substructure towards non-perturbative hadronization scales, one selects 3 distinct splittings from a jet: the 1^{st} soft-drop emission (1SD), the leading charged particles splitting (LCP) and the resolved soft-drop splitting (RSD). The 1SD emission corresponds to the first de-clustering step that satisfies the SD criterion. It physically translates to the first hard (generally partonic) pQCD-like emission within that jet (hardness scale set by $z_{\mathrm{cut}} = 0.1$). The LCP splitting assumes the 2 final-state charged particles with the highest p_{T} from a jet (leading and sub-leading charged particles) are produced in a hadronic splitting. It will serve as a baseline for effects taking place at hadron level (manifestly non-perturbative). The RSD splitting is the SD emission that resolves the 2 leading charged particles, corresponding to the de-clustering step where the sources of the leading and sub-leading charged particles get separated into 2 different prongs. These 3 splittings are schematically represented in Figure 2. The comparison between the 3 holds the potential of revealing how to use jet substructure observables to flag the transition from partonic to hadronic scales. Results presented in this Section are merely a selection of the ones produced for my thesis.

Starting with the groomed momentum fraction z_g introduced in Equation (6), Figure 3 shows the 1SD emission as being highly asymmetrical, as are the pQCD parton splitting functions in Equations (2). This translates into a significantly uneven momentum share between the main and secondary jet branches. However, Figure 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 other 2 splittings, 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 is shown to be tendentiously symmetrical, more to the likes of the LCP splitting, diverging

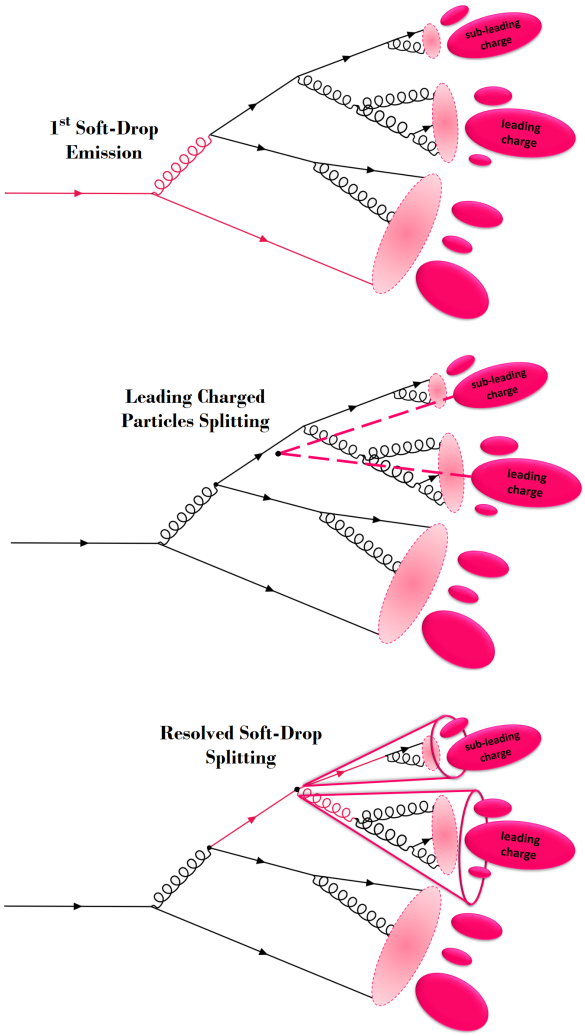

Figure 2: First soft-drop emission (top), leading charged particles splitting (middle) and resolved soft-drop splitting (bottom), highlighted in magenta for the hypothetical branching scheme of a jet produced in a hard scattering event; arrowed lines represent quarks, spiral lines gluons, translucent magenta ellipses the hadronization process and opaque magenta ellipses the outgoing hadrons.

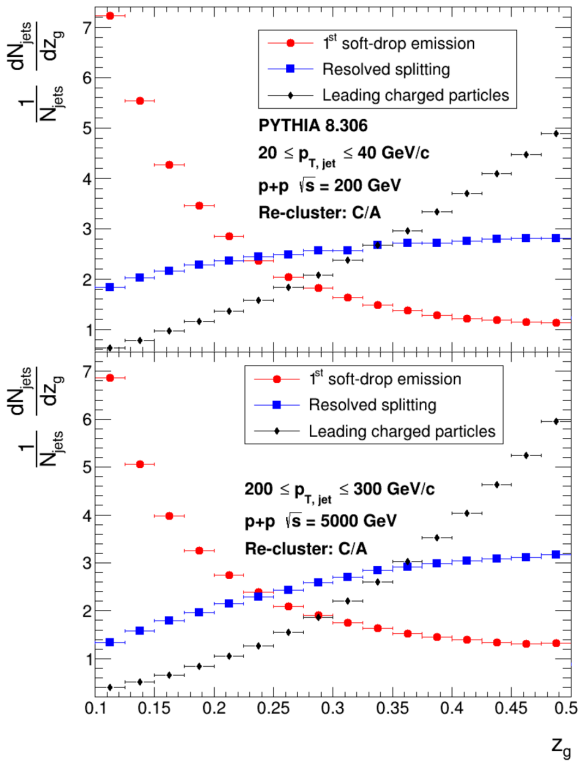

Figure 3: z_g distributions for the splittings of $20 < p_{\mathrm{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions (top panel) and of $200 < p_{\mathrm{T,jet}} < 300$ GeV/c jets produced $\sqrt{s} = 5$ TeV pp collisions (bottom panel); distributions are represented in red circle, blue square and black diamond markers for the 1SD, RSD and LCP splittings, respectively; re-clustering is performed with the C/A algorithm.

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 both RHIC and LHC jets (and also for both C/A and τ re-clustering).

When it comes to the formation time at each of the 3 splittings of interest, τ_{form} is calculated using Equation (7). The results in Figure 4 show a shift to higher formation time values caused by the 1SD $\rightarrow$ RSD $\rightarrow$ LCP transition for RHIC kinematic settings. The RSD distribution follows the 1SD one more closely for small τ_{form} and then the LCP one for large τ_{form}. In fact, the $\tau_{\mathrm{form,\ RSD}}$ distribution is significantly flatter than the $\tau_{\mathrm{form,\ 1SD}}$ one. This 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 $\tau_{\mathrm{form,\ RSD}}$ distribution.

To make a more rigorous analysis of the relative behaviour of the τ_{form} distributions, it is interesting to look at the ratio between them. The ratios are 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 5 shows that the shape of the 1SD and LCP τ_{form} distributions are never the same. Small τ_{form} is dominated by 1SD, while large τ_{form} is dominated, naturally, by the LCP distribution, leading to a sharp decrease of the 1SD/LCP ratio.

However, when it comes to the RSD/LCP ratio, Figure 5 reveals that RSD dominates at smaller τ_{form} but it quickly decreases and stagnates at ratios below but very close to unity. For that time range, RSD and LCP seem to very closely match when

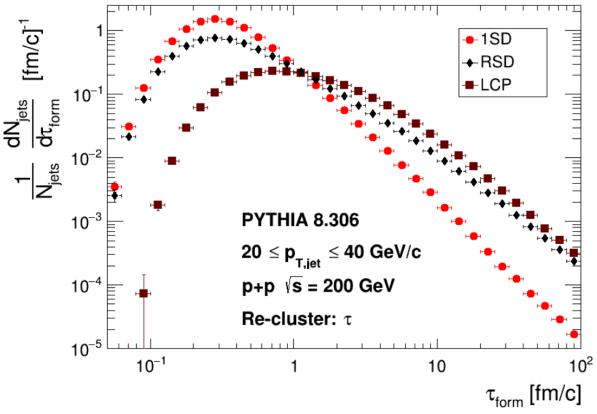

Figure 4: $\tau_{\rm form}$ distributions for $20 < p_{\rm T,jet} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions, corresponding to the 1SD, RSD and LCP splittings in red circle, black diamond and brown square markers, respectively; re-clustering is performed with the τ algorithm.

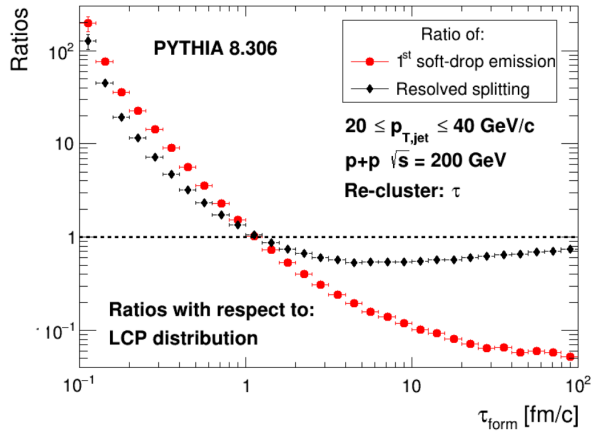

Figure 5: Ratios of the $\tau_{\rm form}$ distributions of the 1SD and RSD splittings with respect to the $\tau_{\rm form}$ distribution of the LCP, represented in red circle and black diamond markers, respectively, for $20 < p_{\rm T,jet} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions; re-clustering is performed with the τ algorithm.

it comes to the shape of their $\tau_{\rm form}$ distributions. Taking the $\tau_{\rm form}$ where the RHIC RSD/LCP ratio becomes smaller than unity as reference, the time range for which this ratio behaves as a "plateau" near unity is identified as (for both re-clustering algorithms)

$$\text{Plateau:} \quad \tau_{\rm form} > 1.3 \pm 0.3 \text{ fm/c} \quad . \tag{8}$$

This hints that cuts in formation time will select jets with differently-placed RSD splittings. If the cut is made on $\tau_{\rm form,RSD}$ for the range in (8), the RSD of the selected RHIC jets seem to have the hadron-like time-structure of the LCP splitting.

For LHC kinematic settings, Figure 6 reveals the same time-ordering for the different splittings, with the average $\tau_{\rm form}$ increasing significantly for 1SD $\rightarrow$ RSD $\rightarrow$ LCP. It also shows the shape of the RSD distribution to resemble the 1SD one for small $\tau_{\rm form}$ and then the LCP one for large $\tau_{\rm form}$.

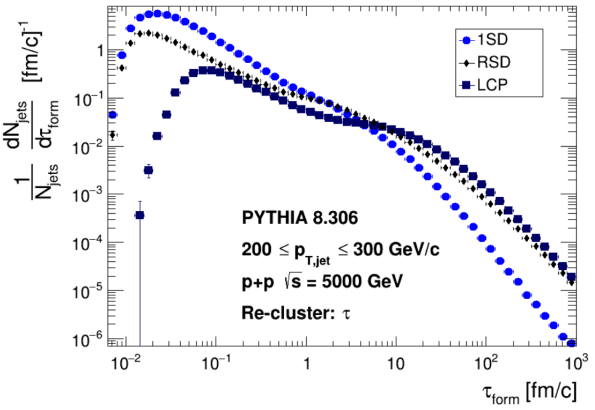

Figure 6: $\tau_{\rm form}$ distributions for $200 < p_{\rm T,jet} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions, corresponding to the 1SD, RSD and LCP splittings in red circle, black diamond and brown square markers, respectively; re-clustering is performed with the τ algorithm.

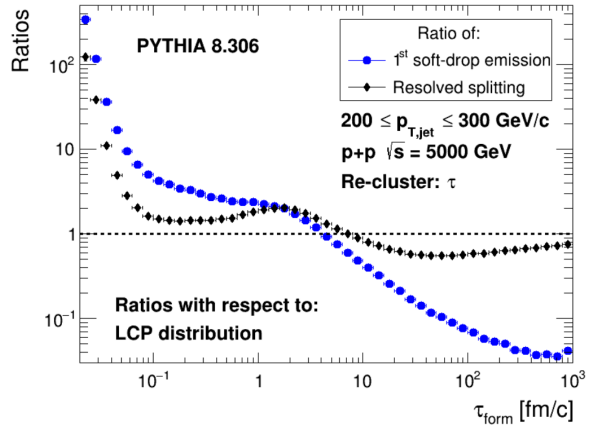

Figure 7: Ratios of the $\tau_{\rm form}$ distributions of the 1SD and RSD splittings with respect to the $\tau_{\rm form}$ distribution of the LCP, represented in red circle and black diamond markers, respectively, for $200 < p_{\rm T,jet} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions; re-clustering is performed with the τ algorithm.

In Figure 7, the 1SD/LCP ratio 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 larger $\tau_{\rm form}$. However, the RSD/LCP ra-

tio 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 the 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

$$\begin{aligned} &1^{\text{st}}\text{ Plateau:} \quad 0.10 \pm 0.03 < \tau_{\text{form}} < 1.0 \pm 0.3 \text{ fm/c} \\ &2^{\text{nd}}\text{ Plateau:} \quad \tau_{\text{form}} > 8 \pm 2 \text{ fm/c} \end{aligned} \tag{9}$$

for C/A re-clustering and

$$\begin{aligned} &1^{\text{st}}\text{ Plateau:} \quad 0.10 \pm 0.03 < \tau_{\text{form}} < 0.6 \pm 0.2 \text{ fm/c} \\ &2^{\text{nd}}\text{ Plateau:} \quad \tau_{\text{form}} > 6 \pm 2 \text{ fm/c} \end{aligned} \tag{10}$$

for τ re-clustering.

Once again, these cuts in formation time should select jets with more LCP-like RSD splittings, indicating they might be taking place at hadronic scales.

Given the "plateau" structure of RSD/LCP ratios found at RHIC and LHC energies, one proceeds to separately sample jets according to the RSD formation time cuts presented in Equations (8), (9) and (10). 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 prior to the LHC 1^{st} "plateau" ($\tau_{\text{form,RSD}} < 0.10$ fm/c).

The number of SD emissions that occur after the RSD splitting ($\text{N}_{\text{post-RSD}}$) provides a measure of how far the RSD splitting is from the end of clustering tree. This number can be traced along both the main and the secondary RSD prongs. The average $\text{N}_{\text{post-RSD}}$ for RHIC and LHC kinematics are shown in Tables 1 and 2, respectively.

Ranges	Average $\text{N}_{\text{post-RSD}}$	
	Main	Secondary
Drop	2.740 ± 0.003	1.644 ± 0.002
Plateau	1.202 ± 0.002	0.502 ± 0.001

Table 1: Average $\text{N}_{\text{post-RSD}}$ both along the main and the secondary RSD prongs and for C/A re-clustered $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions, selected from the drop and plateau $\tau_{\text{form, RSD}}$ regions.

Table 1 shows 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 shown in (8) (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. This hints that large $\tau_{\text{form, RSD}}$ selects RHIC jets whose RSD splitting effectively marks the end of the clustering tree along the secondary RSD prong, meaning RSD is in the vicinity of the hadronization scales.

Ranges	Average $\text{N}_{\text{post-RSD}}$	
	Main	Secondary
Drop	4.654 ± 0.006	3.934 ± 0.006
1^{st} Plateau	4.060 ± 0.005	3.054 ± 0.004
2^{nd} Plateau	1.302 ± 0.002	0.557 ± 0.001

Table 2: Average $\text{N}_{\text{post-RSD}}$ both along the main and the secondary RSD prongs and for C/A re-clustered $200 < p_{\text{T,jet}} < 300$ GeV/c jets from $\sqrt{s} = 5$ TeV pp collisions, selected from the drop, 1^{st} plateau and 2^{nd} plateau $\tau_{\text{form, RSD}}$ regions.

The LHC averages in Table 2 also show that jets sampled from smaller RSD formation time cuts still have, on average, multiple SD emissions after the RSD along both RSD prongs. In fact, the $\text{N}_{\text{post-RSD}}$ averages found for LHC's 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. However, for the higher $\tau_{\text{form, RSD}}$ range of the 2^{nd} "plateau", most clustering trees keep evolving along the main RSD prong (although to a lesser extent than jets from the drop and 1^{st} "plateau" ranges), but they effectively stop along the secondary prong.

Furthermore, the similarity between the $\text{N}_{\text{post-RSD}}$ averages at the "plateau" of the RHIC ratio and at the 2^{nd} "plateau" of the LHC ratio is noteworthy (last row of Tables 1 and 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 (verified for both C/A and τ re-clustering).

The effect of different RSD formation time selections on the (un-groomed) jet mass M_{jet} is shown in Figures 5 and 7 for RHIC and LHC kinematic settings, respectively.

Figure 8 shows a clear distinction in M_{jet} between the drop and "plateau" ranges of the RHIC ratio. Jets with the higher $\tau_{\text{form, RSD}}$ of the "plateau" region have significantly smaller mass than the lower

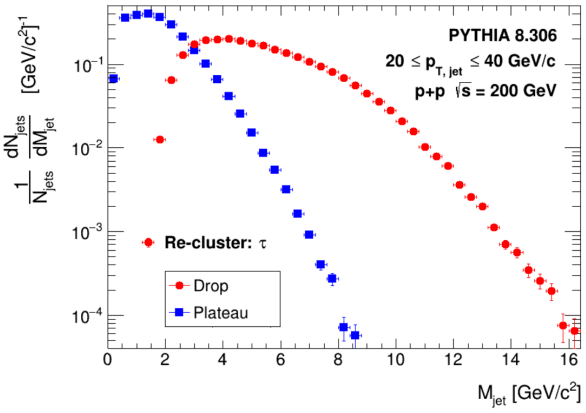

Figure 8: M_{jet} distributions for $20 < p_{\text{T,jet}} < 40$ GeV/c jets produced in $\sqrt{s} = 200$ GeV pp collisions and for whom $\tau_{\text{form, RSD}}$ falls in the drop and plateau regions from the RSD/LCP time-ratio, represented by red circle and blue square markers, respectively; re-clustering is performed with the τ algorithm.

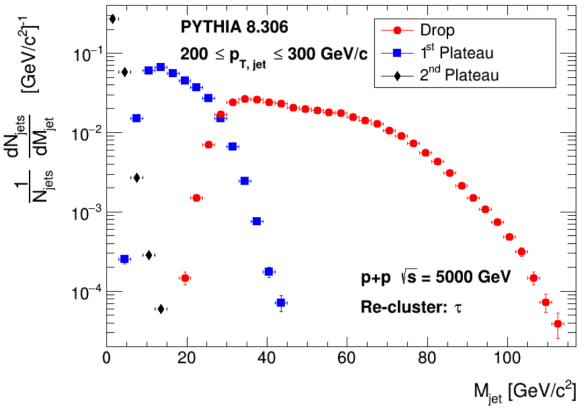

Figure 9: M_{jet} distributions for $200 < p_{\text{T,jet}} < 300$ GeV/c jets produced in $\sqrt{s} = 5$ TeV pp collisions and for whom $\tau_{\text{form, RSD}}$ falls in the drop, 1^{st} plateau and 2^{nd} plateau regions from the RSD/LCP time-ratio, represented by red circle, blue square and black diamond markers, respectively; re-clustering is performed with the τ algorithm.

$\tau_{\text{form, RSD}}$ jets from the drop region. A clear separation in the M_{jet} distributions can also be observed in Figure 9, with the jets from the 2^{nd} "plateau" of the RSD/LCP ratio from the LHC settings having much smaller mass than the jets from the 1^{st} "plateau" and drop regions. Since there is little overlap between distributions for each of the $\tau_{\text{form, RSD}}$ ranges of interest, the jet mass appears to be a good proxy for jet selection via formation time (verified for both C/A and τ re-clustering). Through a selection on higher jet masses, clustering trees will have less SD emissions and the RSD splitting will typically occur near LCP.

It is relevant to check how the different RSD placements contribute to the RSD/LCP ratios presented in Figures 5 and 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 ($N_{\text{RSD}} = 1, 2, 3$), separately. These ratios are shown in Figures 10 and 11.

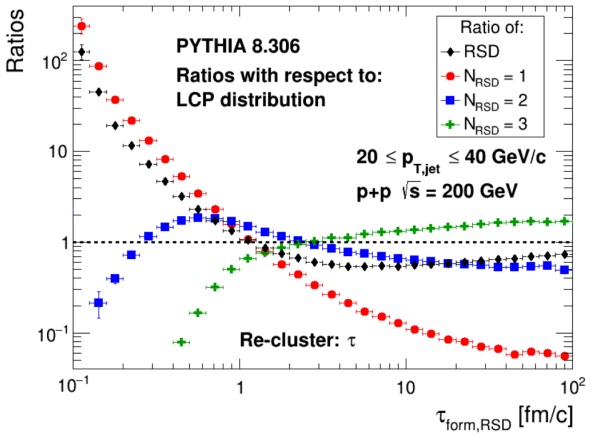

Figure 10: RSD/LCP time-ratios for the generic RSD and for RSD located at $N_{\text{RSD}} = 1, 2, 3$ for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from in $\sqrt{s} = 200$ GeV pp collisions, represented in black diamond, red circle, blue square and green cross markers, respectively; re-clustering is performed with the τ algorithm.

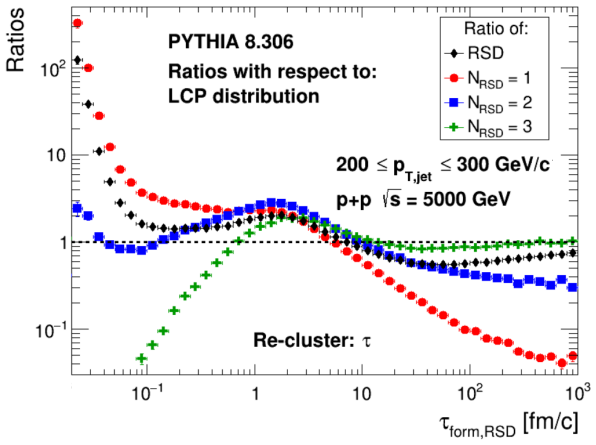

Figure 11: RSD/LCP time-ratios for the generic RSD and for RSD located at $N_{\text{RSD}} = 1, 2, 3$ for $200 < p_{\text{T,jet}} < 300$ GeV/c jets from in $\sqrt{s} = 5$ TeV pp collisions, represented in black diamond, red circle, blue square and green cross markers, respectively; re-clustering is performed with the τ algorithm.

Figure 10 clearly shows that the "plateau" region of the RHIC RSD/LCP time-ratio gets its shape from the contributions of later RSD splittings, namely for $N_{RSD} = 2$ and $N_{RSD} = 3$. However, for $N_{RSD} = 2$ and $N_{RSD} = 3$, the ratios are shown to hit a "plateau" around the same time where the generic RSD/LCP time-ratio is located at. Since RHIC jets have an average of 3 to 4 soft-drop emissions in total along the main jet branch, the ratio distribution for $N_{RSD} = 3$ shows properties of the ending limits of the parton shower.

Figure 11 shows the composition of the double-"plateau" structure of the RSD/LCP ratios for the LHC setup. It reveals that the 1st "plateau" from Figure 7 comes from the balance of contributions from early RSD splittings, namely $N_{RSD} = 1$ and $N_{RSD} = 2$, as is the case for the drop region. These observations show that the RSD similarity to the LCP in the 1st "plateau" 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: $N_{RSD} \geq 3$. Furthermore, the contribution makeup of the 2nd "plateau" is very similar to the one observed in Figure 10 for the single-"plateau" of the RHIC 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 $\tau_{\text{form, RSD}}$. All and all, one concludes that the late distribution of $\tau_{\text{form,RSD}}$ is the same as the late distribution of $\tau_{\text{form},LCP}$, marking a possible transition from the pQCD to the npQCD regimes (verified for both C/A and τ re-clustering).

To further investigate substructure at the RSD splitting, Figure 12 shows the RHIC z_g distributions for the 1SD, 2SD and 3SD emissions along the main branch for jets selected with RSD situated in the 1SD emission (RSD=1SD), in the 2SD emission (RSD=2SD) and in the 3SD emission (RSD=3SD).

The top panel from Figure 12 reveals 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 distribution was itself expected to be flat (similarly to what is observed in Figure 3), but the following 2SD and 3SD emissions are shown to have a similar behaviour. As for the middle panel, where RSD=2SD, the 1SD emission is shown to be significantly asymmetrical, the 2SD emission, being the resolved one, to have a flat z_g distribution and the posterior 3SD emission to also have a flat z_g distribution. Finally, the bottom panel, having RSD=3SD, shows highly asymmetrical 1SD and 2SD emissions. These observations hold independently of the re-clustering algorithm and are very similar to the ones observed for

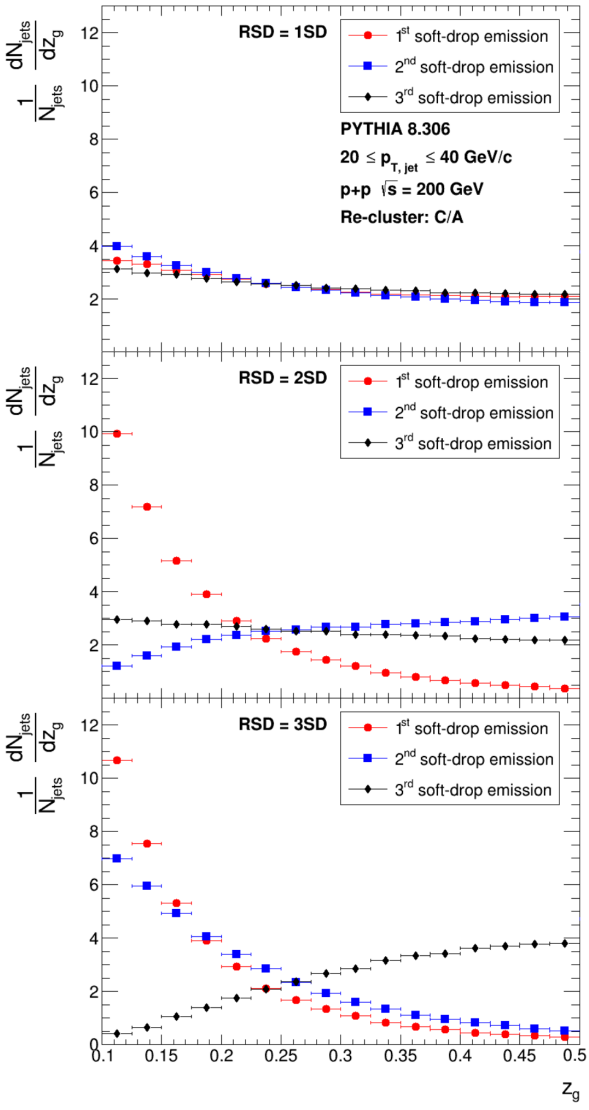

Figure 12: z_g distributions for the 1SD, 2SD and 3SD emissions, represented by red circle, blue square and black diamond markers, respectively, for $20 < p_{\text{T,jet}} < 40$ GeV/c jets from $\sqrt{s} = 200$ GeV pp collisions; distributions for jets with RSD=1SD, 2SD and 3SD are displayed in the top, middle and bottom panels, respectively; re-clustering is performed with the C/A algorithm.

LHC kinematic settings. Therefore, from the RSD splitting onward, the z_g distributions flatten for the subsequent SD emissions along the main branch, while the previous SD emissions conserve the highly asymmetric typical shape.

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

particle ends its partonic-like cascade.

6. Conclusions

The RSD splitting was found 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 the 1SD, RSD and LCP helps to understand further the connection between them. The RSD and LCP time distributions are shown to practically match in ranges of large τ_{form}, with the ratio between them stabilizing close to unity. It is possible to identify one plateau for RHIC settings ($\tau_{\text{form}} > 1.3$ 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 $\tau_{\text{form, RSD}}$ are 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 is also 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. On the other hand, the main contributions to the RHIC plateau and the 2^{nd} LHC plateau come from jets with RSD placed higher in the clustering tree, namely after the 2SD emission.

Finally, the RSD splitting is shown to flag the transition from parton-like to hadron-like splittings.

Acknowledgements

I would like to thank Dr. Liliana Apolinário and Dr. Raghav Kunnawalkam Elayavalli for the indispensable guidance and support provided in the making of this project.

References

[1] 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.

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

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

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

[5] 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.

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

[7] V.N. Gribov and L.N. 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.

[8] Zhiqing 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.

[9] 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.

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

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

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

[13] 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.

[14] 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.

[15] 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.