Space-Time Structure of QCD jets

João Miguel Martins da Silva

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisors: Prof. Dr. José Guilherme Teixeira de Almeida Milhano
Prof. Dr. Liliana Marisa Cunha Apolinário

Examination Committee

Chairperson: Prof. Dr. Mário João Martins Pimenta
Supervisor: Prof. Dr. José Guilherme Teixeira de Almeida Milhano
Member of the Committee: Prof. Dr. João Nuno Ramalho Gonçalves Pires

December 2020

Para a minha mãe, Graça.

Acknowledgments

I first want to express my gratitude to my supervisors, professors Guilherme Milhano and Liliana Apolinário, for accepting my request of working with them and for making this journey full of fruitful conversations, sometimes including much needed moments of humour. Their patience and thoughtful comments were indispensable. I cannot wait to keep working with you both in the future.

To the people that helped me not taking Physics so seriously and, at the same time, maintaining the same level of curiosity about it throughout these last years - thank you Fernando Lima, João Rato and João Taquelim. To my long standing and beloved friends - Adriana, Ana, Inêses, Joana, Madalena and Pinto - you do you, that's all I need. Odeceixe awaits us.

To my girlfriend, Carolina Azeitona, I cannot express how thankful I am. Her infinite loving support is what kept me going and her patience in hearing me talk about things she's not that interested in - like jet physics - is unmatched.

Finally, I'm grateful for the warmth my family provided me, allowing for my feet to stay on the ground, each in its own way. I thank my father, José, for its inquisitive way of talking and curious questions about my work. I also thank my sister, Rita, not for being a pain in the but - which she is sometimes -, but for always being there to ask me how things are going. Last, but definitely not least, thank you, *mãe*. I (obviously) wouldn't be here without you, but I also wouldn't be as persistent and as hard-working if I had not learned that from you.

I acknowledge the financial support by Fundação para a Ciência e a Tecnologia (Portugal) under project CERN/FIS- PAR/0024/2019.

Resumo

Jatos hadrónicos representam uma oportunidade para explorar a física da interação forte a escalas variadas. A sua evolução desde um quark ou um gluão (colorido) altamente energético até a um conjunto de hadrões (não coloridos) é usualmente formulada no espaço dos momentos, sem fazer referência à imagem espacio-temporal que dela emerge. No entanto, em colisões de iões pesados, um referencial de espaço-tempo é introduzido pela interação de jatos com o *Plasma de Quarks e Gluões*, o meio denso e quente existente nos primeiros instantes do nosso Universo. Motivado por isto, este trabalho é uma investigação da estrutura espacio-temporal de jatos de QCD no vácuo.

Aproximando o desenvolvimento de um jato por uma história de clustering, as escalas de tempo (*tempos de formação*) envolvidas são calculadas e as emissões são ordenadas em tempo absoluto. É demonstrado que os tempos de formação cobrem múltiplas ordens de magnitude. É verificado não só que estes estão exatamente ordenados ao longo dos ramos de uma história atribuída pelo algoritmo de *reclustering* τ, mas também que estão logaritmicamente separados. É explicitamente demonstrado que algoritmos de *reclustering* diferentes resultam em estruturas de espaço-tempo diferentes para o mesmo jato. A escala de tempo da primeiro emissão, contudo, revela-se como independente desta escolha de algoritmo. Adicionalmente, é apresentada uma estratégia simples para estimar este tempo de formação através de quantidades mensuráveis experimentalmente.

Palavras-chave: Cromodinâmica Quântica; Jato; História de Fragmentação; Tempo de formação; Estrutura espacio-temporal.

Abstract

Hadronic jets provide us with a rich framework for the exploration of the physics of strong interactions at multiple scales. Their evolution from an highly energetic (colored) quark or gluon to an ensemble of (color-neutral) hadrons is usually formulated in momentum space, without making reference to the space-time picture that emerges from it. However, in heavy-ion collisions, a space-time reference frame is introduced by the interaction of jets with the *Quark-Gluon Plasma*, the hot and dense medium existing in the first instants of our Universe's lifetime. Motivated by that, this work is an investigation of the space-time structure of QCD jets that propagate in vacuum.

By approximating a jet's development by a clustering history, the timescales (*formation times*) involved are calculated and an absolute time order of the splittings is presented. Formation times are shown to span multiple orders of magnitude. It is verified not only that they are exactly ordered along the branches of a history assigned by the τ reclustering algorithm, but also that they are logarithmically separated. Furthermore, one makes explicit that different reclustering algorithms result in different space-time structures for the same jet. The first splitting's timescale, however, is shown to be independent of the choice of algorithm. Additionally, one presents a simple strategy to estimate this particular formation time through experimentally measurable quantities.

Keywords: Quantum Chromodynamics; Jet; Branching history; Formation time; Space-time structure.

Contents

Acknowledgments v

Resumo vii

Abstract ix

List of Figures xiii

Nomenclature xix

1 Introduction **1**

1.1 Motivation and Overview 1

1.2 Objectives 4

1.3 Thesis Outline 4

2 Theoretical Background **5**

2.1 QCD Basics 5

2.2 Jets and Logarithmic Divergences 9

2.3 Parton Showers and Jet Reconstruction 13

2.4 Hadron Collisions and Collider Variables 18

2.5 Jet Fragmentation Histories From Clustering Algorithms 20

2.6 From Parton Formation Time to Jet Space-Time Structure 23

3 Simulation and Reconstruction Setup **27**

4 Results Analysis **31**

4.1 Reconstruction efficiency and ISR contamination 31

4.2 First splitting's formation time 33

4.2.1 Sensitivity to ISR and hadronization 34

4.2.2 Correlation with Jet Mass 36

4.3 Reclustering algorithm comparison 38

4.4 Formation Time Estimates and Kinematical Dependences 41

4.4.1 Local Approach 41

4.4.2 Global Approach 46

4.5 Space-Time Structure Analysis 51

4.5.1 Formation Time Orderings 52

    4.5.2 Formation Time Distribution 56

5 Conclusions **63**

  5.1 Main Achievements and Future Work 63

Bibliography **67**

A Formation time calculation for a single emission **71**

  A.1 Formation time calculation for a single emission 71

List of Figures

1.1 Pictoric overview of the components of an heavy-ion collision. The soft part contributes to the creation of QGP between the two receding discs and the hard part in the form of back-to-back jets. [4] 2

1.2 Snapshots of a central 2.76 TeV PbPb collision at different times with hadrons (blue and grey spheres) as well as QGP (red). The red lines indicate the approximate longitudinal location of particles with rapidity $y = 0$, $y = 1$, and $y = 6$. [3] 2

2.1 QCD branch of the Standard Model of Elementary Particles: 3 generations of fermions (quarks), represented by each column, and the boson (gluon). Taken from [16]. 5

2.2 On the left, taken from [19], one has the ratio of cross sections $R = \frac{\sigma(e^- e^+ \to \text{hadrons})}{\sigma(e^- e^+ \to \mu^- \mu^+)} = N_c \sum_q e_q^2$, where e_q is the electrical charge of quark with flavour q. On the right, taken from [24], a clear illustration of why the introduction of color was needed, aswell as the impact of considering different number of quarks flavours. Note that the top line, the result for 6 quark flavours, would only be accurate for center of mass energies high enough for the production of the heaviest quark (the top quark, see Fig. 2.1). 7

2.3 Running of the QCD coupling for $262 < Q < 1675$ GeV, taken from [25]. 8

2.4 Feynamn diagrams for $e^+e^- \to q\bar{q}$ at order $\mathcal{O}(\alpha_S)$. a) Full tree-level diagram. b), c) Real emission diagrams. d) Virtual 1-loop diagram. Adapted from [27]. 10

2.5 The $y - \phi$ plane for jets obtained with the C/A (left) and anti-k_t (right) algorithms, for jet radius $R = 1$. Taken from [27], originally from [40]. 17

2.6 Sketch of a typical hadronic collision. The focus is the red radiation - the final state shower - coming off the partons produced in the hard scattering (red blob). Taken from [42]. . . . 18

2.7 Example of a parton branching diagram $q \to qg$. We're set to calculate the formation time of the quark with momentum p^μ, which is assumed to come from some previous interaction, i.e, it is a virtual quark. 23

3.1 Examples of hard interaction diagrams for Z+jet, $qg \to Zq$ on the left, and Dijet, $gg \to gg$ on the right. For Z+jet there is a single hard scattered parton (the final quark) and for Dijet there are two hard scattered partons (the two final quarks). 28

4.1 Hard interaction diagrams for Z+jet event type, along with kinematics of the hard scattered particles. Because we assume the initial partonic collision carries no transverse momentum, we necessarily have $p_t = p_t^Z$. Hence, in this work, an idealized jet reconstruction would have $p_t^{jet}/p_t^Z = 1$. 32

4.2 Ratios p_t^{jet}/p_t^Z for various reconstruction radii, with and without Initial State Radiation. All distribution are normalized to unit integrals. 33

4.3 First splitting's formation time distributions for the 3 setups mentioned in the text and with reconstruction radii $R = 1$ and $R = 1.75$. As for most distributions encountered throughout the thesis and which cover more than 1 order of magnitude, the bins of each histogram are logarithmic, i.e., equally spaced in logarithmic space. All plots are normalized to unit integrals. 35

4.4 a) Jet mass m^{jet} spectrum. b) Jet transverse momentum p_t^{jet}spectrum. Both results are parton-level and are normalized to unit integrals. 36

4.5 a) Scattering plot of first splitting formation time τ_0^{seed} and jet mass m^{jet}. b) Average first splitting formation time for each jet mass in the range 1 GeV $< m^{\text{jet}} < 200$ GeV, along with the standard deviation calculated for each mass bin. Both results are parton-level. 37

4.6 Scattering plot of first splitting formation time τ_0^{seed} and jet mass m^{jet} for phase space cut a) 100 GeV $< p_t \cosh y < 200$ b) 200 GeV $< p_t \cosh y < 300$ GeV. Both results are parton-level 37

4.7 Average first splitting formation time for each jet mass (roughy the center of each marker) in the range 1 GeV $< m^{\text{jet}} < 100$ GeV, along with the standard deviation calculated for each mass bin. Results are presented for various $p_t \cosh y$ bins and at parton-level. 38

4.8 Space time structure for a branching history calculated with a) τ algorithm b) C/A algorithm. Red nodes/lines correspond to final partons, while the blue nodes (with corresponding black line) correspond to internal partons. Each formation time assigned to each black line is in fm/c and the 4-momenta of partons are not shown for readability's sake. The results are parton-level FSR. The figures were generated through a Python package for network algorithms [56], with information extracted from a single generated event. 39

4.9 Mean formation time for each splitting according to their order in absolute time, along with the 5% and 95% quantiles which are represented by the the lower and upper solid lines, respectively. The results are shown for the 4 different algorithms, for the first 5 splittings and for events of a) FSR and b) FSR+ISR+hadronization. Note that each point is plotted in the center of each bin, e.g., the first splitting corresponds to 0.5. 40

4.10 Ratios of the mean formation times of each splitting, having the τ algorithm as a reference. Results are presented for the remaining algorithms, for the first 10 splittings and for events of a) FSR and b) FSR+ISR+hadronization. 40

4.11 Branching history tree with 3 final partons. Each parton's 4-momentum is represented by p_i. The numbering follows no specific order. 41

4.12 a) Factor controlling the first order corrections of the square-root term in Eq. (4.5), plotted for each final parton. b) Factor controlling the smallness of the m_i^2 terms in Eq. (4.5), plotted for each parton of each children pair. Both plots are normalized to unit integrals. . 42

4.13 a) Angular distance ΔR^2 distribution for each pair of children partons. b) First order correction to the factor $\cosh y_i / \cosh y$ for each parton of each children pair. Both plots are normalized to unit integrals. 43

4.14 Softest transverse momentum fraction z distribution for each pair of children partons. Normalized to unit integral. 43

4.15 a) Factor controlling the major corrections to Eq. (4.9), which takes into account possible enhancements to the small $m_i^2/p_{t,i}^2$ term in the soft and collinear limit. The plot is normalized to unit integral. b) Zoomed distribution in the region $\frac{1}{(1-z_i)\Delta R^2}(m_i^2/p_{t,i}^2) > 1$. 44

4.16 a) Average formation time for the first 10 splittings, for the τ_i defined in Eqs. (4.5), (4.7) and (4.9). The approximations in these expressions were done sequentially. For instance, τ_4 in Eq (4.9) is the result of the low-virtuality, collinear and soft approximations. b) Ratio of the average formation times τ_2 and τ_4 with respect to the exact estimate τ_0. 46

4.17 a) Average formation time for the first 10 splittings, for the τ_0 and τ_1 defined in Eqs. (4.5) and (4.6). The remaining $\overline{\tau}_i$ are defined in Eq. (4.13) The remaining $\overline{\tau}_i$ are defined in Eq. (4.13) where the respective approximations are done individually. For instance, $\overline{\tau}_3$ is obtained by taking the collinear limit $\cosh(y_1 - y_2) \approx 1 + (y_1 - y_2)^2$, $\cos(\phi_1 - \phi_2) = 1 - (\phi_1 - \phi_2)^2$ and $\cosh y_i \approx \cosh y$ in Eq. (4.5). b) Ratio of the average formation times with respect to the exact estimate τ_0. 46

4.18 Branching history tree with 3 final partons. Each parton's 4-momentum is represented by p_i. The numbering follows no specific order. 47

4.19 a) Factor controlling the first order corrections of the square-root term in Eq. (4.16), plotted for each final parton. b) Factor controlling the smallness of the m_i^2 terms in Eq. (4.16). Plotted for each parton and for the minimum $p_j \cdot p_k$ possible from all possible jk pairs. Both plots are normalized to unit integrals. 48

4.20 a) Angular distance ΔR_{ij}^2 distribution for each possible pair of final partons. b) Factor $\cosh y_i / \cosh y^{jet}$ for each final parton. Both plots are normalized to unit integrals. 49

4.21 Ratio of the transverse momentum of the second hardest parton to that of the hardest parton in a jet. Normalized to unit integral. 49

4.22 a) Average formation time for the first 10 splittings, for the τ_i^g defined in Eqs. (4.5), (4.17) and (4.19). The approximations in these expressions were done sequentially. For instance, τ_4^g in Eq (4.9) is the result of the low-virtuality, collinear and soft approximations. b) Ratio of the average formation times τ_2^g and τ_4^g with respect to the exact estimate τ_0. . 51

4.23 a) Average formation time for the first 10 splittings. The τ_0 is defined in Eq. (4.5) and the remaining are defined in Eq. (4.20). The $\overline{\tau}_i^g$'s are the formation time estimates one obtains by doing the respective approximation directly in τ_0. For instance, $\overline{\tau}_3^g$ is obtained by taking the collinear limit $\cosh(y_i - y_j) \approx 1 + (y_i - y_j)^2$, $\cos(\phi_i - \phi_j) = 1 - (\phi_i - \phi_j)^2$ and $\cosh y_i \approx \cosh y^{jet}$ in Eq. (4.5). b) Ratio of the average formation times with respect to the exact estimate τ_0. 51

4.24 Branching history tree with 4 final partons and a single branch. Each internal parton (thicker lines) contributes with a formation time. The numbering follows no specific order. In this case, one would plot ratios τ^3/τ^2 and τ^2/τ^1. For trees with more than a single branch, this is done for each branch. 52

4.25 a) Ratio of the formation times, calculated with τ_0, of consecutive partons i and $i+1$ in each branch. b) Same plot with logarithmic y-axis with superimposed curves for τ_0, τ_1, τ_2 and τ_4 as given by Eqs. (4.5), (4.6), (4.7) and (4.9). All plots are normalized to unit integrals and resulted from data of all generated events. 53

4.26 Branching history tree with 4 final partons and two branches. Each internal parton (thicker lines) contributes with a formation time. The numbering follows the specific sequence of reclustering, i.e., τ_3 is the first reclustering step, τ_2 is the second and so on. 53

4.27 a) Ratio of the formation times, calculated with τ_0, of consecutive reclustering steps i and $i-1$. b) Same plot with logarithmic y-axis with superimposed curves for τ_0, τ_1, τ_2 and τ_4, as as given by Eqs. (4.5), (4.6), (4.7) and (4.9). All plots are normalized to unit integrals and resulted from data of all generated events. 54

4.28 a) Ratio of the absolute times, calculated with τ_0, of consecutive reclustering steps i and $i-1$. b) Ratio of the absolute times, calculated with $\tau_5 = 1/d_\tau$, of consecutive reclustering steps i and $i-1$. All plots are normalized to unit integrals and resulted from data of all generated events. 55

4.29 Branching history tree with 5 final partons and two branches. Each internal parton (thicker lines) has a corresponding inter-particle distance d_i^τ and contributes with a formation time $\tau_5^i = 1/d_i^\tau$. The numbering follows the specific sequence of reclustering, e.g., τ_5^4 is the first reclustering step's formation time. 55

4.30 a) Scattering plot for the formation time of each splitting according to their order in absolute time. b) Jet particle multiplicity spectrum. All results are parton-level and the multiplicity spectrum is normalized to unit integral. 57

4.31 a) Plot of the relevant quantiles (blue lines) and the mean values (black markers) for each each bin. The inner box represents the 25%, 50% (the median formation time) and 75% quantiles. The lower and upper bounds represent the 5% and the 95% quantiles, respectively. This is plotted for the first 20 splittings. b) Scattering plot for the formation time of each splitting according to their order in absolute time. Only the portion of each bin's distribution between the corresponding 5% and 95% quantiles is ploted. All results are parton-level. 57

4.32 a) Scattering plot for the formation time of each splitting according to the reclustering sequence b) Scattering plot of the formation time of each splitting and corresponding ΔR of the subsequent partons. All results are parton-level. 58

4.33 a) Scattering plot for the formation time of each splitting according to their absolute time and only for splittings which result in partons separated by, at least, $\Delta R = 10^{-2}$. b) Average of formation times for each splitting, with errors given by the standard error of the mean (SEM), for the first 5 splittings of the history. The SEMs are almost unperceptible, only being relevant for the last two splittings on the plot. All results are parton-level. 59

4.34 Left column: Scattering plot for the formation time of each splitting according to their absolute time. Right column: Average of formation time for each splitting, with errors given by the standard error of the mean (SEM). The jet multiplicity is fixed in each plot at 9, 10 and 11 splittings. Results are shown for the first 12 splittings and are all parton-level. . . . 61

A.1 Generic Feynman diagram for a single parton emission off a virtual parton. 71

Nomenclature

pp	Proton-Proton
DGLAP	Dokshitzer–Gribov–Lipatov–Altarelli–Parisi
FSR	Final-State Radiation
ISR	Initial-State Radiation
LHC	Large Hadron Collider
LL	Leading Logarithmic
LO	Leading Order
QCD	Quantum Chromodynamics
QGP	Quark-Gluon Plasma
RHIC	Relativistic Heavy-Ion Collider

Superscripts

$\dagger$	Transpose conjugate.

Chapter 1

Introduction

1.1 Motivation and Overview

Over the past few decades, we have witnessed numerous ultra-relativistic hadronic collision experiments with increasing energy, whether it be between proton beams [1] (pp) or heavy nuclei beams [2], usually Lead (Pb) or Gold (Au). This experimental endeavour, carried out most recently by the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory and the Large Hadron Collider (LHC) at CERN, allowed us to study, among a wide range of aspects of particle physics, the theory of the strong interaction, namely Quantum Chromodynamics (QCD).

The analytical equations governing the evolution of QCD systems are formulated in momentum space. Likewise, the usual parton shower Monte Carlo implementations (approximate pictures of this evolution) follow calculations in momentum space. They both respect a given ordering variable which relates to the QCD logarithmic divergences and which allows for soft/hard physics factorization. This ordering variable serves as the quantity through which one can order the evolution of a dynamic system that goes from one physical scale to another. An example of such evolution is that of hard scattered *partons* (quarks or gluons) in hadronic collisions, which fragment sucessively producing a multi-particle final state. This is introduced in a more detailed way in Chapter 2.

All of this is usually studied without ever detailing what the actual space-time evolution is. Where and when does the first particle emission happen after a quark or a gluon has been produced in the hard scattering of a collision event? Is this question relevant for a QCD jet propagating freely in vacuum, like in typical pp collisions? No. Is it relevant for a QCD jet produced in heavy-ion collisions? Absolutely. In such events, a background medium is produced - the *Quark-Gluon Plasma* (QGP) - which introduces a space-time reference frame in which jets propagate. Let us introduce this complex QCD state of matter for the purpose of motivating the study of the space-time structure of QCD jets.

In a typical heavy-ion collision, two nuclei collide at ultra-relativistic energies and are Lorentz-contracted into disks which form very complex systems of a multitude of partons. As these discs overlap they will interact via strong force between their constituents. Aside from the hard perturbative scatterings which give rise to hadronic jets, the majority of these interactions will involve little transverse momentum transfer and will be carried out by color fields and color charge exchanges. These are called ”soft” processes and they are the main reason why both the energy and particle densities rise abruptly shortly after (~ 1 fm/c) the collision [3] - the QGP is created (Fig. 1.1).

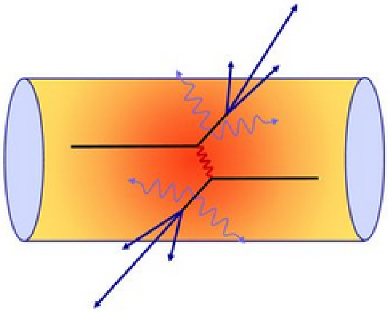


Figure 1.1: Pictoric overview of the components of an heavy-ion collision. The soft part contributes to the creation of QGP between the two receding discs and the hard part in the form of back-to-back jets. [4]

This deconfined, hot and dense state of matter, with energy densities of about 20 times those inside typical hadrons [3], characterizes the aftermath of a heavy-ion collision. It has been thoroughly investigated in distinct experiments and it is supported by unequivocal evidence (see, for example, measurements at RHIC [5] or at LHC [6]). One of the reasons why this medium is such an interesting study subject is because it is understood that this was the state of matter that existed up to about 1 μs after the beginning of our Universe, with temperatures above $\Lambda_{QCD} \sim 200$ MeV [7]. Although one refers to this medium as being in a deconfined QCD state, its constituent partons are not, in any way, independent from each other. In fact, they strongly interact and form a collective medium which resembles a liquid with very small viscosity to entropy ratio [8]. This collective fluid, driven by internal pressure forces, expands and cools down after its creation. It holds its distinct properties until the local energy density falls below that of a typical hadron, at which points it turns into a collection of hadrons. The typical lifetime of the QGP is ~ 10 fm/c [9] (Fig. 1.2). For an exposition of the QGP's interesting physical properties, the reader should refer to [3] and references therein.

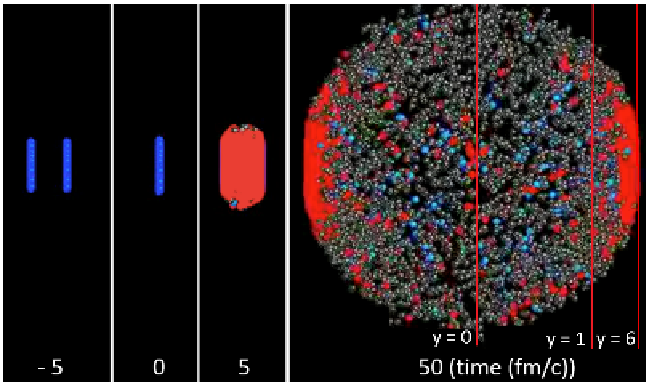


Figure 1.2: Snapshots of a central 2.76 TeV PbPb collision at different times with hadrons (blue and grey spheres) as well as QGP (red). The red lines indicate the approximate longitudinal location of particles with rapidity $y = 0$, $y = 1$, and $y = 6$. [3]

During the medium's lifetime, its fluid-like expansion occurs not only in the direction of the beam but also in the one transversal to it, with velocities reaching about half of the speed of light after 1 fm/c [3]. It is this characteristic of the QGP that allows for particle jets to be candidate probes of its inner structure and properties, by interacting with it while it evolves in time and space. In fact, jets suffer a wide range of modifications when they travel through the QGP by scattering off the constituents of the plasma - *jet quenching* [10]. The degree of quenching of the jet, its change in orientation, composition and substructure are then great footprints for what happened inside the "fireball" when the jet had to plough through it. By searching for jet observables which signal the presence of the medium, one can then infer information about its properties when drawing comparisons to a well established baseline, i.e. pp collisions, where there is no QGP and the jets properties are well understood. Detailed measurements of jet quenching, not just related to jet observables, are shown in, e.g., [11].

If, in particular, one wishes to access the spatio-temporal evolution of the QGP, then one needs to be sensitive to its characteristic timescales. Most probes seem to be sensitive only to the integrated time evolution of the medium. Jets, however, present themselves as probes at multiple different scales. In fact, modified in-medium jets signal that the evolving partonic wave-function and medium interacted somewhere in space-time during its short lifetime. Hence, jet-quenching may provide insights on the relevant space-time regions for jet-plasma interaction. To leverage this information and extract knowledge about the QGP's spatial and temporal profiles, it is imperative that one calculates a causality structure for QCD jets, i.e., a space and time meaning for each emission. The key for this calculation lies in the parton formation time, which is derived from the Heisenberg's uncertainty principle.

The probing of the temporal evolution of the QGP has already been approached in a number of works. For instance, motivated by experimental evidence of unchanged fragmentation patterns of in-medium jets, Casalderrey-Solana *et al.* [12] calculate the parton formation times inside a jet by using information directly from the Monte Carlo event (as opposed to only using experimentally accessible information, as is done in this work). By using a simple model for nuclear matter density, the authors concluded that roughly 70% of the relevant branching process will always occur outside the medium. In a more recent work, Apolinário *et al.* [13] worked on using the top quark as the QGP's time structure probe. The motivation for this novel approach lies on the fact that the interaction of jets originated from the decay products of the W-boson, which arises from the top quark's decay, with the QGP occurs at later times (i.e. not simultaneously with the collision). By relating this time delay with the reconstructed top quark's transverse momentum and using a simple energy loss model, the authors concluded they could extract meaningful insights about the QGP's time structure. Lastly, recent efforts were carried out in [14] and [15], where the authors explored the jet reclustering algorithms that could better reproduce the time structure of QCD jets in the presence of the medium. In particular, Apolinário *et al.* [15] introduce the τ algorithm and conclude, using the first emission's formation time as a metric, that it maximizes the correlation between the reclustered branching history and the corresponding Monte Carlo parton shower information.

1.2 Objectives

The main aim of this thesis is to explore the timescales involved in the transition from a single hard parton to a multi-parton ensemble. The framework for this study is the production of QCD jets - multi-particle final states which are footprints of this quantum evolution - in high-energy proton-proton (pp) collisions, i.e., *vacuum* jets. One formulates this evolution in a simplistic manner and solely by resorting to final state information, i.e., the jet's constituents. First, one assigns a *branching history* to each jet and then quantifies the relevant timescales through the calculation of *formation times* of intermediate virtual states. The result is a space and time meaning to the parton shower interpretation of multiple parton emissions. The space-time structure that is obtained from this process is often put into perspective by having the lifetime of the *Quark-Gluon Plasma* as a reference timescale.

1.3 Thesis Outline

In Chapter 2, one introduces the fundamental theoretical ideas supporting the conceptual part of this work. In particular, the concept of logarithmic divergences and jets, which serve as a basis for most of the reasoning in the thesis, are elaborated on. Afterwards, in Chapter 3, one details the setup of the generation of collision events and how each event's data is handled so as to produce the results shown in this work. These results are analysed in detail in Chapter 4. There, one starts off by analysing the compromise between jet reconstruction efficiency and ISR contamination in Section 4.1, followed by a study of the formation time of a jet's first splitting in Section 4.2. Next, one compares space-time structures for branching histories obtained with different algorithms in Section 4.3, followed by a study of the kinematic dependences of the parton formation time and ends with a comprehensive investigation of the behaviour of the formation time distributions at partonic level and with a branching history fixed by the τ algorithm. Lastly, in Chapter 5, one reflects on the main achievements of this work, while pointing out directions for further future studies which can build on it.

Chapter 2

Theoretical Background

In this chapter, one will introduce key concepts which form the basis for the remaining of the work. A brief overview of the theory of strong interactions is given, followed by an exposition about particle jets and the logarithmic divergences which are unavoidable in Quantum Chromodynamics (QCD) calculations. The parton shower interpretation of jet evolution is introduced with a detour through Sudakov form factor and its link to the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equations. Jet reconstruction and relevant collider variables are detailed and a broad picture of hadronic collisions is painted. One ends with the two centerpiece concepts - assignment of a *branching history* to a jet's constituents and the *parton formation time*.

2.1 QCD Basics

Quantum Chromodynamics (QCD) is the theory that describes the strong interaction between the constituents of the subatomic particles called hadrons, e.g., protons and neutrons, the components of atomic nuclei. These constituents are quarks, fermions with fractional electric charge, and gluons, the massless bosons mediating the strong force (Fig. 2.1).

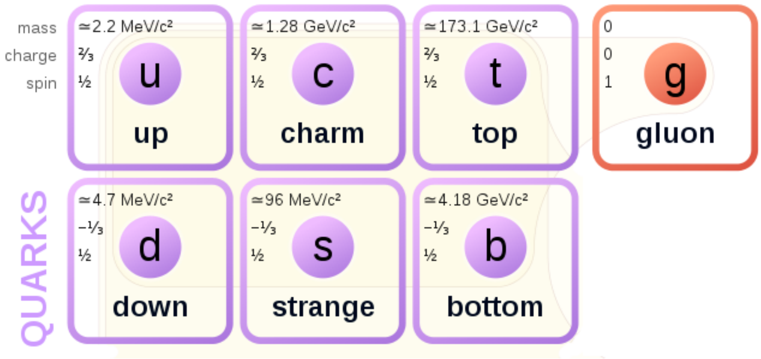


Figure 2.1: QCD branch of the Standard Model of Elementary Particles: 3 generations of fermions (quarks), represented by each column, and the boson (gluon). Taken from [16].

Such theory arises when one imposes invariance of the free fermion Lagrangian under the local gauge transformation group of SU(3), a symmetry strongly motivated both by theory ([17], [18]) and experiment [19] (Fig. 2.2) and which corresponds to an additional quantum number - color. This implies the existence of quark fields in 3 different colors and the necessity of the introduction of 8 gauge fields - the gluon fields - that have to be massless. These quantum fields are defined by their transformation properties under representations of SU(3). The fact that the symmetry group of the strong interaction is non-commutative demands that theory be non-Abelian and, as a consequence of that, the gluon field has self-interactions (three and four-gluon vertices).

The theory's Lagrangian density is [20]

$$\mathcal{L} = \sum_q \bar{\psi}_{q,a} \left(i\not{D} - m_q \right)_{ab} \psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{\mu\nu}_A \tag{2.1}$$

where repeated indices are summed over. The first term, with spinor indices omitted, contains both the kinetic and mass terms of the quark field and the quark-gluon interaction, while the last one is the kinetic term for the gluon field. The $\psi_{q,a}$ are the quark field spinors for a quark of flavour q (running over the n_f quark flavours, Fig. 2.1) and mass m_q, with the color index running from $a = 1$ to $N_c = 3$, the total number of colors. The covariant derivative, in the fundamental representation of SU(3), is given by $(D_\mu)_{ab} = \partial_\mu \delta_{ab} + i g_s (t_C \mathcal{A}^C_\mu)_{ab}$, where $t_C = \lambda_C/2$ and λ_C are the eight Gell-Mann matrices. The field tensor $F^A_{\mu\nu}$ is given by

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f^{ABC} \mathcal{A}^A_\mu \mathcal{A}^B_\nu \tag{2.2}$$

where $\mathcal{A}^A_\alpha$ is the gluon field, with the gluon color index running from $A = 1$ to $N_c^2 - 1 = 8$, g_s is the QCD coupling constant, determining the strength of the interaction, and f^{ABC} are the structure constants of the SU(3) group. The third term is where gluon self-interactions show up and the key point distinguishing QCD from Quantum Electrodynamics (QED), the theory describing light-matter interaction, and the main reason why the former is computationally much more complex than the latter.

The fact that QCD is a non-Abelian gauge theory has further implications on how the strong force behaves. As shown in [18], this type of theories has the property of *asymptotic freedom*. Put simply, such property dictates that the strength of the interaction decreases with increasing momentum and with corresponding decreasing distance, i.e, if we look closer and closer inside an hadron, its constituents will appear to be freely moving.

This assertion was in agreement with Deep Inelastic Scattering (DIS) experiments ([21], [22]) done in the 1960's. This type of experiment is an extension of the Rutherford scattering to higher energies using an electron to probe the internal structure of the proton. The aim was that photons with high virtualities $Q^2 = -q^2 > 0$ (q its four-momentum) could probe small regions of space, i.e., their wavelengths $\lambda_\gamma \sim 1/Q$ needed to be smaller than the typical proton charge radius (~ 1 fm) in order to interact with the hypothetical constituents. The product result of this short distance interaction was measured at large deviation angles of the scattered electron and at a rate much higher than was expected for a structureless proton. This, together with the discovery of *Bjorken scaling*, strongly supported the claim that hadrons were made up of point-like constituents - partons, a term introduced at the time of the parton model [23] but which is still used today as a collective name for quarks and gluons. Such a term will be widely used throughout this work. Later on, the scaling property was verified to be violated by slowly (logarithmically) varying QCD corrections coming from the running of the coupling, which one describes ahead.

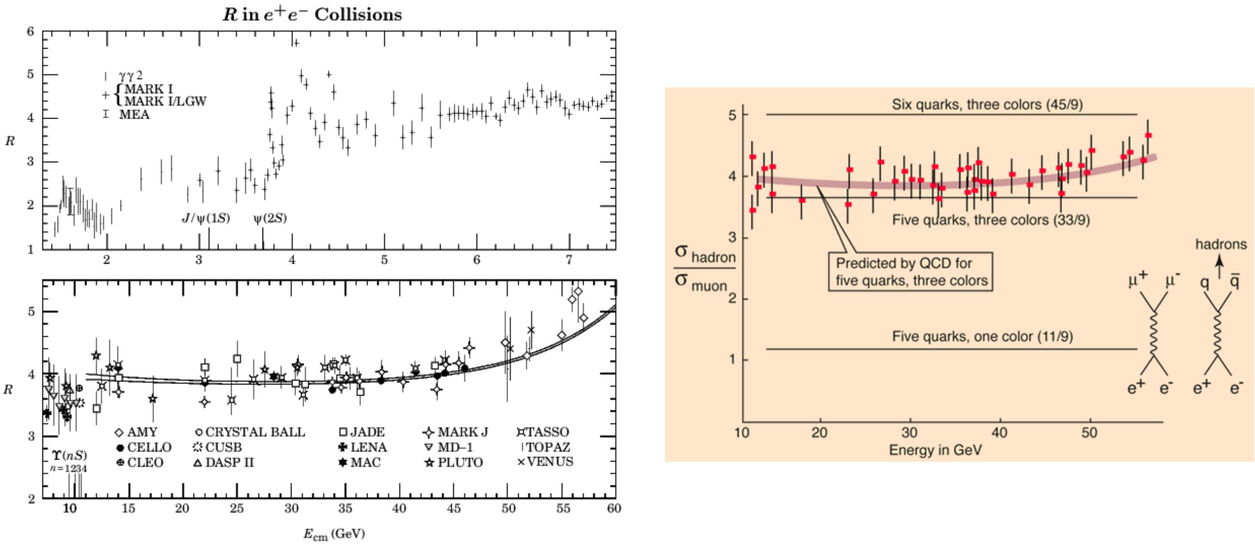


Figure 2.2: On the left, taken from [19], one has the ratio of cross sections $R = \frac{\sigma(e^- e^+ \to \text{hadrons})}{\sigma(e^- e^+ \to \mu^- \mu^+)} = N_c \sum_q e_q^2$, where e_q is the electrical charge of quark with flavour q. On the right, taken from [24], a clear illustration of why the introduction of color was needed, aswell as the impact of considering different number of quarks flavours. Note that the top line, the result for 6 quark flavours, would only be accurate for center of mass energies high enough for the production of the heaviest quark (the top quark, see Fig. 2.1).

Formally, since QCD is a renormalizable theory (i.e. its infinities are mathematically tractable), there exists an energy scale (μ, an arbitrary valued renormalization scale) dependence attached to all theory parameters. For a given observable quantity F, the total dependence on μ must vanish, since the parameter is unphysical, and one demands that $\mu \frac{dF}{d\mu} = 0$ be satisfied. This materializes on a coupling constant which evolves with the energy scale - the running coupling -, giving rise to an equation of the type

$$\mu^2 \frac{\partial \alpha_s}{\partial \mu^2} = \beta(\alpha_s) = -(b_0 \alpha_s^2 + b_1 \alpha_s^3 + ...) \tag{2.3}$$

where $\alpha_s = \frac{g_s^2}{4\pi}$ and $\beta(\alpha_s)$ is a function to be determined. The expansion on the right-hand side comes from assuming the validity of a perturbative calculation, i.e., assuming small α_s and finite b_i coefficients.

If one just takes the first term of the perturbative expansion in Eq. (2.3), we get the one-loop coupling constant of QCD

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + b_0 \alpha_s(\mu^2) \ln(Q^2/\mu^2)} \tag{2.4}$$

with $b_0 = \frac{11N_c - 2n_f}{12\pi}$, Q the energy scale of the process, usually the momentum transfer involved in it, and μ the renormalization scale. It is precisely the first factor in b_0 (i.e. the existence of the color quantum number) that makes it > 0, therefore producing a decrease of the coupling constant with the scale Q^2

(Fig. 2.3). Often, one writes Eq. (2.4) in terms of a mass scale which is generated as a consequence of renormalization, a process called dimensional transmutation. That is the QCD scale $\Lambda_{QCD} \approx 200$ MeV, which allows (2.4) to be written as $\alpha_s(Q^2) = \frac{1}{b_0 \ln(Q^2/\Lambda^2)}$ and sufficiently above which, i.e., in high energy processes, one can use perturbative QCD (pQCD). Below and this scale, non-perturbative dynamics dominates.

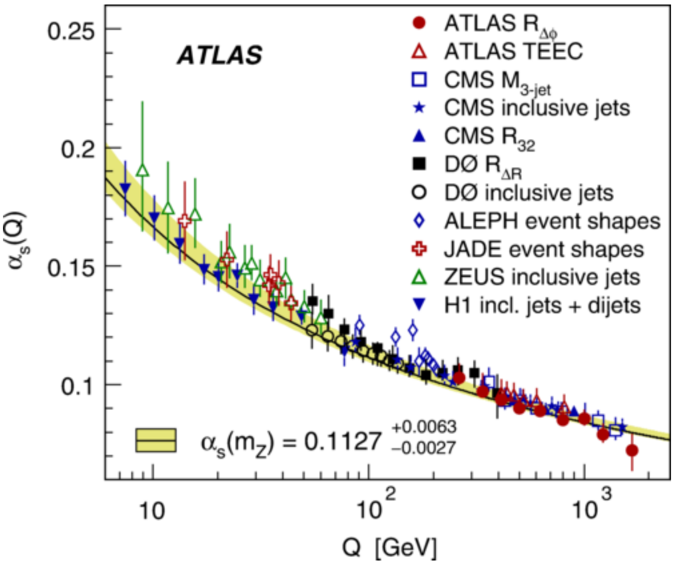


Figure 2.3: Running of the QCD coupling for $262 < Q < 1675$ GeV, taken from [25].

As a consequence of the running coupling, the theory exhibits two interesting phenomena. At large distances, i.e., small energy scales, QCD is strongly interacting, with quarks and gluons bounded within hadrons, leading to *color confinement*. At small distances, i.e., high energy scales, it is weakly interacting, leading to *asymptotically free* partons. Qualitatively, this is caused by the *anti-screening* effect, whereby a colored charge has its field enhanced at large distances. The reason for this is that, although there's a cloud of virtual quark-antiquark pairs ($q\overline{q}$) canceling the color charge, this is countered by the force mediators - the gluons - which also carry color charge and whose net effect, due to the self-interactions, is to enhance the color field. In terms of Feynman diagrams, the vacuum polarization has more contributions than the one coming from the quark loop. Note that this field enhancing is in opposition with the well known QED coupling constant behaviour - from a large charge in a given region, one gets a small observable influence far away. The reason for such contrast is that the boson of the theory - the photon - carries no charge, hence there is nothing countering the effect of electric charge cancelling caused by fermion-antifermion pairs (mostly electron-positron pairs) - the *screening* effect.

An interesting repercussion of *confinement* is how quarks and gluons can never be detected as free particles, for they are bounded within hadrons. Therefore, only these color-singlet (color-neutral) combinations are asymptotic states of the theory and can be detected.

2.2 Jets and Logarithmic Divergences

In particle collisions, whenever there's the possibility of producing partons with high enough energy and virtuality, one detects relatively isolated bunches of energy in the form of collimated sprays of hadrons. We call these collections of particles *QCD jets* and they can be both detected and calculated in, e.g., e^-e^+ , e^-p (DIS experiments), pp (our main framework in this thesis) and heavy-ion collisions (e.g. Au+Au or Pb+Pb).

Jets can be thought of as footprints of the theory's colored degrees of freedom produced in rare short-distance interactions, carrying information about physics at multiple scales. In pp and heavy-ion collisions, these processes are $2 \to 2$ partonic hard interactions, e.g., $gg \to q\overline{q}$ or $q\overline{q} \to gZ$. Such scatterings are characterized by the exchange of a particle with virtuality high enough to probe, directly or indirectly, the hadronic structure. This amounts to $Q \gg \Lambda_{QCD}$ - the pQCD regime. The reason why, in the detectors, we get hadrons instead of the assymptotically free, high energy partons, has already been mentioned in Section 2.1 - *confinement*.

The process whereby one goes from a single parton being produced off the hard collision at large angles to a multitude of hadrons being detected as a jet, can be phrased as a conjugation of higher order corrections to the hard process - the emission of an arbitrary number of partons [26] - and of an *hadronization* phase. In going from one stage to another, one evolves the system's wavefunction between two physical scales: a hard scale, calculable through pQCD at a given precision level, and a soft, inherently non-perturbative, scale. The perturbative phase, which can be *factorized* (separated) from both the hard process and the hadronization, is characterized by the sucessive fragmentation of a primary parton into a multi-parton final state all the way down to the hadronic mass scales $Q \approx 1$ GeV. Classicaly, such a phenomenon is a consequence of the radiation off an acelerated (color) charge. The fragmentation is usually modeled by a *parton shower*, an approximate representation of a jet's partonic content and evolution, which contains the dominant QCD contributions at all orders. Before we elaborate more on this, let's first introduce some relevant, theoretical concepts.

Consider a $q\overline{q}$ pair emerging from a hard process. An elementary process often used as example is e^-e^+ annihilation into a highly virtual photon. The tree-level matrix element (Fig. 2.4) a)) is easily calculated,

$$\mathcal{M}_{q\overline{q}} = A^{\mu}\overline{u}_a(k_1) i e_q \gamma_\mu \delta_{ab} v_b(k_2) \tag{2.5}$$

where $\overline{u}_a(k_1)$ and $v_b(k_2)$ are the massless spinors for the outgoing quark with color index a and anti-quark with color index b, respectively, and γ_μ are the 4×4 Dirac matrices. The quark's electric charge is e_q. The A^μ represents the remaining part of the matrix element which, in this example, is the photon propagator and the electron-photon vertex. We won't need to write this out explicitly in the context of our calculation.

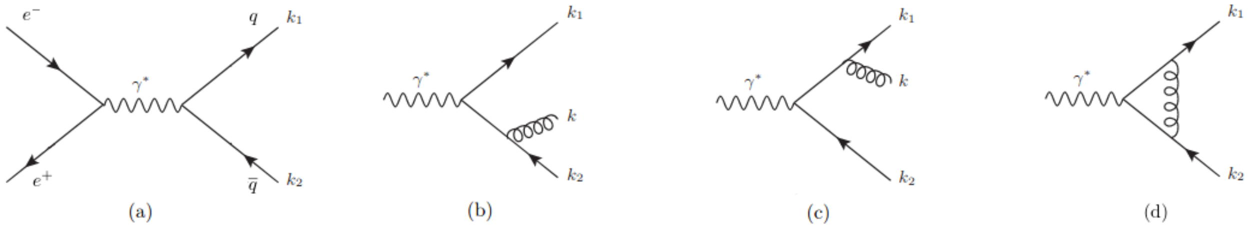


Figure 2.4: Feynamn diagrams for $e^+e^- \to q\bar{q}$ at order $\mathcal{O}(\alpha_S)$. a) Full tree-level diagram. b), c) Real emission diagrams. d) Virtual 1-loop diagram. Adapted from [27].

Now consider the emission of a gluon off either the quark or the anti-quark leg (Fig. 2.4 b), c)). By squaring the sum of these two matrix element contributions and then averaging over color indices and polarization, one gets [28]

$$|\mathcal{M}_{q\bar{q}g}|^2 \approx |\mathcal{M}_{q\bar{q}}|^2 \left(C_F g_s^2 \frac{2k_1 \cdot k_2}{(k_1 \cdot k)(k_2 \cdot k)} \right) \tag{2.6}$$

where the *soft* approximation (gluon momentum $k^\mu \ll k_1^\mu,\, k_2^\mu$) was used and the partons are taken as massless. The factor $C_F = \frac{4}{3}$ is the color factor ("Casimir") associated with gluon emission from a quark and g_s is the QCD coupling constant. An important property of pQCD calculations shows up in Eq. (2.6): the squared matrix element $|\mathcal{M}_{q\bar{q}g}|^2$ *factorizes* into a hard part $|\mathcal{M}_{q\bar{q}}|^2$ and a *soft* gluon emission part. At the level of the phase space, we see that, in the soft limit, *factorization* also applies:

$$d\Phi_{q\bar{q}g} = \left(\prod_{i=1,2} \frac{d^3k_i}{2E_i(2\pi)^3} \right) \frac{d^3k}{2E(2\pi)^3} \approx d\Phi_{q\bar{q}} \frac{d^3k}{2E(2\pi)^3} \tag{2.7}$$

Now, if one calculates the differential cross section for the gluon emission off a $q\bar{q}$ pair, one ends up with the factorized expression:

$$d\sigma_{q\bar{q}g} \approx d\sigma_{q\bar{q}} d\mathcal{S}, \qquad d\mathcal{S} = EdEd(\cos\theta)\frac{d\phi}{2\pi}\frac{\alpha_s C_F}{\pi}\frac{k_1 \cdot k_2}{(k_1 \cdot k)(k_2 \cdot k)} \tag{2.8}$$

There's a couple of consequential remarks one can make about Eq. (2.8), which is calculated in the *soft* and massless limit:

1. The cross section $d\sigma_{q\bar{q}g}$ is dominated by both *soft* and *collinear* kinematic regimes. Just note that both terms on the denominator are singular in either of these cases because, in the massless limit,

   $$k_i \cdot k \approx E_i E(1 - \cos\theta_{ik}) \tag{2.9}$$

   is small for $\vec{k_i} \parallel \vec{k}$ $(\cos\theta_{ik} \approx 1)$ or for $0 \approx E \ll E_i$.

2. An emission's phase space is constrained by that of the previous emission, a property known as *angular ordering*. This can readily be seen if, after some algebraic manipulation, one separates the emission off the quark leg from the emission off the anti-quark leg and integrates each one over the azimuthal angle ϕ of the emitted gluon [20]. The result is

$$d\mathcal{S} = \frac{\alpha_s C_F}{\pi}\frac{dE}{E}\left(\frac{d(\cos\theta_{1k})}{1-\cos\theta_{1k}}\Theta(\theta_{1k} < \theta_{12}) + \frac{d(\cos\theta_{2k})}{1-\cos\theta_{2k}}\Theta(\theta_{2k} < \theta_{12})\right) \tag{2.10}$$

where the first term corresponds to the gluon emission off the quark and the second term to the emission off the anti-quark. The emission angles θ_{1k} and θ_{2k} are measured with respect to the quark and anti-quark direction, respectively. The *angular ordering* property then dictates that emissions with emission angle larger than the previous one are strongly suppressed. Although it may seem from Eq. (2.10) that such emissions are exactly forbidden, one should remember that we're working in the soft and massless limit. This property derives from a color coherence effect and, qualitatively, it can be interpreted as the gluons at large angles not being able to resolve the quark-antiquark pair separately. In our particular example, such gluons will only see the total charge of the pair - zero color charge - and, therefore, won't be emitted [1]

By working in the CM (center-of-mass) frame and manipulating $d\mathcal{S}$ in Eq. (2.8) in the *collinear* limit ($\sin\theta \approx \theta$), one arrives at,

$$d\mathcal{S} = \frac{\alpha_s C_F}{\pi}\frac{dE}{E}\frac{d\theta^2}{\theta^2}\frac{d\phi}{2\pi} \tag{2.11}$$

which highlights the non-integrable *double-log divergences* common to most pQCD calculations, i.e., the singularities at $E \to 0$ and $\theta \to 0$. One can also re-write these divergences in terms of any other kinematic variables linearly related to these two. An example is the pair (z, Q^2), i.e, the energy fraction of the emission and the virtuality of the original $q/\bar{q}$: $d\mathcal{S} \sim \frac{dz}{z}\frac{dQ^2}{Q^2}$.

To make $\sigma_{q\bar{q}g}$ finite, one needs to introduce the virtual corrections of the same order, which, in this case, correspond to the interference between the tree-level diagram of $q\bar{q}$ production and the 1-loop virtual gluon diagram (Fig. 2.4 d)). It can be shown that the cancellation of the singular terms is exact, by, for instance, using dimensional regularization (cf. [29]). Interestingly enough, this *cancellation of infrared (soft) and collinear divergences* can be shown to hold to every order in perturbation theory (cf. Chapter 6 of [29], where the QED analog is treated). However, not every observable benefits from this cancellation of infinities - called IRC (infrared and collinear) safety -, in which case perturbation theory provides us with divergent results. Therefore, any perturbatively calculated QCD observable, in order to be comparable with experiment, must meet the IRC safety requirement. An IRC safe observable can be put as an observable $\mathcal{O}$ which is unchanged under the addition of a zero energy particle or a collinear emission, i.e.,

- Infrared Safety: $\mathcal{O}(p_1, ..., p_N) = \mathcal{O}(p_0, p_1, ..., p_N)$, $p_0 \to 0$
- Collinear Safety: $\mathcal{O}(p_1, p_2, ..., p_N) = \mathcal{O}(\lambda p_1, (1-\lambda)p_1, p_2, ..., p_N)$, $\lambda \in (0, 1)$

As long as we sum over all possible final states in a collision - an *inclusive* observable - our observable is free of IRC divergences. This is established by the KLN (Kinoshita-Lee-Nauenberg) theorem ([30], [31]).

---

[1] In the case of an incoming gluon (instead of a photon), the total charge will not be zero, so this qualitative explanation fails. Nonetheless, angular ordering can still be understood by reinterpreting high angle emissions as coming from the incoming gluon and re-organizing the emission history.

Examples of that are the cross section for jet production and other event shape observables. An important remark to make here is how IRC observables will yield the same result independently of the objects we use to describe the final state - partons or hadrons. This motivates the analysis at parton-level, which this work mainly focus on.

The emission probability calculation of Eq. (2.11) can be done without the soft limit, although only factorizing at the cross-section level, and for the remaining QCD vertices. Notice that the fact the gluon self-interacts implies that when a gluon is emitted it will, in turn, emit further radiation. This is in constrast with QED, where photons can only create e^+e^- pairs after being emitted. In fact, in QCD, soft and collinear gluon emission off a gluon is the dominant process. The generalization of Eq. (2.11) to an arbitrary QCD vertex, in the massless, collinear limit and in NLO, is:

$$d\mathcal{S}_{i \leftarrow j} = \frac{\alpha_s}{2\pi} \hat{P}^{(0)}_{i \leftarrow j}(z) dz \frac{dQ^2}{Q^2} \frac{d\phi}{2\pi} \tag{2.12}$$

where the functions $\hat{P}^{(0)}_{i \leftarrow j}(z)$ are the unregularized, leading-order (LO) splitting functions. These functions can easily be derived from the QCD vertices and have the simple interpretation of the probability of finding a parton of type i in a parton of type j with a fraction z of its momentum. The following question would be ”how do we generalize a single collinear emission to multiple collinear emissions?”.

A possible answer would be that one can calculate the Feynman diagrams to the desired α_s order. However, not only is such a calculation intractable for more than a handful of collinear parton emissions, it also lacks a clear picture of the internal structure of such an event. The usual way to tackle this problem is by seeing it as a continuous evolution process between two resolution scales.

To search for motivation for this approach, consider the final state of the SLAC-MIT DIS experiments [21] mentioned in Section 2.1 - a collection of hadrons. If the initial state is composed of a single hadron, why does one detect several hadrons in the collision’s aftermath? Clearly, the parton which was hard scattered by the photon must have gone through a *fragmentation* process, producing multiple final partons which then recombine to form the hadronic states. Such a process can be put in a more generalized framework, whereby one calculates the probability that a given parton i, being produced from a hard scattering at short distances $1/Q$, forms an inclusive state containing a specific hadron h with a fraction x of its momentum. These probability densities are encoded in the *Fragmentation Functions* (FFs) $D_{h/i}(x, Q^2)$ and their Q^2 dependence provides a way of resolving parton i at different scales, e.g., different values of virtuality. The evolution picture of multiple collinear emissions emerges when one considers the resolution of parton i at successive infinitesimal steps of the scale. If one goes from Q^2 to $Q^2 - \Delta Q^2$, then there’s a probability that a parton j is collinearly emitted off parton i and carries away a fraction z of its momentum. This probability results from considering Eq. (2.12) multiplied by the density of the resulting parton j, integrating for all possible z momentum fractions. Hence, imposing momentum conservation and cancelling soft divergences by including the virtual loop contributions, i.e., an emission which is readily reabsorbed, the change in the FFs is given by (cf. [29] or [28]):

$$
\begin{aligned}
D_{h/i}(x,Q^2-\Delta Q^2) = D_{h/i}(x,Q^2) - \sum_j \int_0^1 dx' \int_0^1 dz \left(\frac{\alpha_s}{2\pi} \hat{P}^{(0)}_{j\leftarrow i}(z) \frac{\Delta Q^2}{Q^2} \right) D_{h/j}(x',Q^2)\delta(x-zx') + \\
+ \sum_j \int_0^1 dz \left(\frac{\alpha_s}{2\pi} \hat{P}^{(0)}_{j\leftarrow i}(z) \frac{\Delta Q^2}{Q^2} \right) D_{h/j}(x,Q^2)
\end{aligned}
\tag{2.13}
$$

where the sum in j accounts for all over possible parton species ($j = q_k,\ \overline{q}_k,\ g$, where k runs over all quark flavours) and the azimuthal angle ϕ was integrated out. One can absorb the virtual part into a regularization of the splitting functions,

$$
P_{j\leftarrow i}(z): \quad \int_x^1 dz P_{j\leftarrow i}(z) f(z) \equiv \int_x^1 dz \hat{P}_{j\leftarrow i}(z) f(z) - \int_0^1 dz \hat{P}_{j\leftarrow i}(z) f(1) \tag{2.14}
$$

By using this prescription, generalizing the splitting functions to include higher order α_s terms and taking the continuous limit, one arrives at the all-order DGLAP equations for the final-state fragmentation of a parton:

$$
\frac{\partial D_{h/i}(x,Q^2)}{\partial \ln Q^2} = \sum_j \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{j\leftarrow i}(z,\alpha_s(Q^2)) D_{h/j}\left(\frac{x}{z},Q^2\right) \tag{2.15}
$$

Succinctly, these coupled evolution equations are a resummation of the leading logarithmic (LL) terms, i.e., an analytical expression for the sum of a subset of the perturbative terms containing the collinear enhanced logarithms $\sim \left(\alpha_s \ln(Q^2)\right)^n$, with $\alpha_s \ln(Q^2) \lesssim 1$. The function $P_{j\leftarrow i}(z,\alpha_s)$ is the all-order regularized *splitting function* and can be calculated via perturbative series $P_{j\leftarrow i}(z,\alpha_s) = P^{(0)}_{j\leftarrow i}(z) + \frac{\alpha_s}{2\pi} P^{(1)}_{j\leftarrow i}(z) + ...$. The exact same type of equations predicts the Q^2 scaling violations of the *Parton Distribution Functions* PDFs $f_{i/h}(x,Q^2)$ describing an hadron's inner structure, the main study focus of the DIS experiments. Such functions are defined such that $f_{i/h}(x,Q^2)dx$ is the probability of resolving a parton i carrying a fraction of its parent hadron h momentum in the vicinity of x, usually called the *Bjorken* x.

2.3 Parton Showers and Jet Reconstruction

To see more clearly how Eq. (2.15) provides us with a *parton shower* picture, let's first introduce a key calculation in QCD. Assuming factorization holds, the parton shower interpretation arises from the calculation of the probability for any number of collinear parton emissions. However, one usually frames the problem in terms of the calculation of the *no-emission* probability, from which the previous the probability follows by unitarity. Consider a parton which is initially resolved at scale Q^2 and introduce a cut-off $Q_0^2 \gtrsim \Lambda^2_{QCD}$ on the resolution scale. Assuming that consecutive emissions are independent, one can multiply the no-emission probabilities in n small subdivisions $\Delta Q_i^2 = Q_{i+1}^2 - Q_i^2 = (Q^2 - Q_0^2)/n$ of the resolution scale and obtain the probability $\overline{P}_j(Q^2, Q_0^2)$ of not emitting partons off parton species j in (Q^2,

Q_0^2). This, together with the emission probability one calculated in Eq. (2.12), implies that

$$\begin{aligned}\overline{P}_j(Q^2,Q_0^2) &= \lim_{n\to+\infty}\prod_i^n \overline{P}_j(Q_{i+1}^2,Q_i^2) = \lim_{n\to+\infty}\prod_i^n\left(1-\int_{z_i}\int_{\phi_i}\frac{\Delta\mathcal{S}_j(Q_i^2,z_i,\phi_i)}{\Delta Q_i^2}\Delta Q_i^2\right)\\ &= \lim_{n\to+\infty}\exp\left(-\sum_i^n\int_{z_i}\int_{\phi_i}\frac{\Delta\mathcal{S}_j(Q_i^2,z_i,\phi_i)}{\Delta Q_i^2}\Delta Q_i^2\right) = \exp\left(-\int_{Q_0^2}^{Q^2}\int_z\int_\phi\frac{d\mathcal{S}_j(Q'^2,z,\phi)}{dQ'^2}dQ'^2\right)\end{aligned} \tag{2.16}$$

where one abreviated $\sum_k d\mathcal{S}_{k\leftarrow j}\equiv d\mathcal{S}_j$ and the two integral signs represent integrations in the azimuthal angle ϕ and the energy fraction z of the parton emission. The explicit result for the no-emission probability off a parton resolved at scale Q^2 is

$$\Delta_i(Q^2,Q_0^2) = \exp\left[-\sum_j\int_{Q_0^2}^{Q^2}\frac{dQ'^2}{Q'^2}\int_{z_-}^{z_+}dz\frac{\alpha_s(Q'^2)}{2\pi}\hat{P}^{(0)}_{j\leftarrow i}(z)\right] \tag{2.17}$$

where $d\mathcal{S}_{j\leftarrow i}$ was written explicitly, the azimuthal angle ϕ was integrated out and the z integral limits $z_- = z_-(Q_0^2,Q'^2)$ and $z_+ = z_+(Q_0^2,Q'^2)$ assure four-momentum conservation and that the emission is resolvable. This is usually called the *Sudakov* form factor and it satisfies $0<\Delta_i(Q^2,Q_0^2)\leq 1$, supporting its interpretation as a probability. Formally, the Sudakov factor in Eq. (2.17) resums, at all orders of α_s, enhanced logarithms coming from hard-collinear and soft-collinear kinematic regimes. The reason why the result is rendered finite and free of infrared divergences is the cut-off Q_0^2 introduction. This cut-off makes Eq. (2.17) a sum of both virtual and unresolvable real contributions, i.e., emissions below the Q_0^2 scale. Hence, the Sudakov factor $\Delta_i(Q^2,Q_0^2)$ gives the probability that a parton i evolves from Q^2 to Q_0^2 without any resolvable branching.

The link between the Sudakov factor and the DGLAP equations in Eq. (2.15) emerges when one manipulates them with LO splittings functions and with both real and virtual contributions made explicit [20]:

$$D_{h/i}(x,Q^2) = \Delta_i(Q^2)D_i(x,Q_0^2) + \int_{Q_0}^{Q}\frac{dQ'}{Q'}\frac{\Delta_i(Q^2)}{\Delta_i(Q'^2)}\sum_j\frac{\alpha_s(Q'^2)}{2\pi}\int_x^1\frac{dz}{z}\hat{P}^{(0)}_{j\leftarrow i}(z)D_{h/j}\left(\frac{x}{z},Q'^2\right) \tag{2.18}$$

where, for readability, the cut-off Q_0^2 was omitted in $\Delta_i(Q^2)$. This first term of the right-hand side makes it clear that $\Delta_i(Q^2)$ gives us the no emission probability between Q and Q_0 - resulting in an unchanged partonic FF $D_i(x,Q_0^2)$ - and the second term sums over all possible paths of emission, with $\Delta_i(Q^2)/\Delta_i(Q'^2)$ quantifying the probability of not branching between scales Q and Q'. It is now clear that the DGLAP equations provide a picture of multiple collinear emissions as an evolution process, whether it is the evolution of the PDFs - *Initial State Radiation* (ISR) -, where spacelike virtualities ($m^2<0$) become increasingly more negative, or the evolution of a fragmenting parton - *Final State Radiation* (FSR) -, where timelike virtualities ($m^2>0$) are gradually relaxed until reaching the hadronic scales.

Although it is the DGLAP equations which provide the evolution picture, it is actually the Sudakov form factor which is suitable for the implementation of a parton shower algorithm via Monte Carlo methods. Such implementation is simply a probabilistic interpretation of a jet's evolution as a sequence of $1\to 2$ collinear emissions. Note that this algorithm, which implements a LL calculation, is appropriate as long as emissions are *strongly ordered* in the shower's *ordering variable* ($Q_1^2 \ll Q_2^2 \ll ... \ll Q_n^2$). This not only allows for factorization, hence, the recursive application of the single emission calculation, but it also assures that the LL contributions are maximal with respect to the subleading ones. An ordering variable is valid at a given accuracy as long as it reproduces the singularity structure of multiple emissions at that accuracy [32]. Examples of ordering variables commonly used are the PYTHIA 8.2.35 [33] (the event generator used in this work and is described in Chapter 3) transverse momentum given by $p_{\perp \text{evol}}^2 = z(1-z)Q^2$ (FSR) or $(1-z)Q^2$ (ISR), virtuality Q, which is an option in PYTHIA 6 [34], and emission angle θ, as in HERWIG [35]. These variables are LL equivalent, meaning that the observables one computes with each parton shower will differ only by sub-leading terms, i.e., next-to-leading logarithmic (NLL) terms. Along with the choice of ordering variable, one usually needs to employ a momentum *reshuffling* strategy. This is done in order to conserve four-momentum when reconstructing the momenta of initial partons from the final ones, which is not trivial since, while the hard process and subsequent emissions are generated with on mass-shell partons, any parton emitting radiation is off mass-shell.

A Monte Carlo implementation of a simple parton shower algorithm for FSR can be stated as follows: given a parton exiting a vertex with hardness Q_2^2, (taken to be of order the hard process Q^2) one seeks a solution of the equation $r = \Delta_i(Q_2^2)/\Delta_i(Q_1^2)$, with $r \in [0,1]$ a uniform random number, and solves it for the hardness of the next branching Q_1^2. If $Q_1^2 \leq Q_0^2$, Q_0 usually taken as $\sim \Lambda_{QCD}$, no splitting is generated and the line is interpreted as a final parton. If $Q_1^2 > Q_0^2$, a branching is generated at the scale Q_1^2. Its z value and the final parton species are generated with a probability proportional to $\hat{P}_{i \leftarrow j}^{(0)}(z)$. The azimuth $\phi \in [0, 2\pi]$ is generated uniformly. This procedure is started with each of the primary process partons, and is applied recursively to all generated partons. It generates an arbitrary number of partons, and it stops when no final-state partons undergo further splitting, i.e., when they reach the hadronization scale Q_0. This type of evolution is called *forward* evolution. Contrasting, ISR generation is done by *backwards* evolution, i.e., starting at the lowest resolution scale, in this case the hard scattering scale. Additionally, the Sudakov factors ratio is weighted by the ratio of the PDFs at the corresponding scales $\left[f(x,Q_1^2)/f(x,Q_2^2)\right] \Delta_i(Q_2^2)/\Delta_i(Q_1^2)$.

From a full event (real or simulated), how does one exactly identify a QCD jet? Usually, by taking the kinematic information of the final state particles in an event, one can make a reconstruction using a *jet algorithm* - a set of rules to cluster the particles' four-momenta into jets. The first-ever jet algorithm was developed by Sterman and Weinberg in the 1970's [26]. It was intended for e^+e^- collisions and classified an event as having two jets if at least a fraction $1-\epsilon$ of the event's energy was contained within two cones of opening half-angle δ (and hence is known as a "cone" algorithm). This definition made it possible to have a fully consistent pQCD calculation of the probability of having two jets in an event. Today, we have a variety of jet algorithms that fall into two categories:

- Cone Algorithms: They identify stable cones of a given angular size. These cones are defined in such a way that if one or two nearby particles are added to or removed from the jet cone, that it won't drastically change the cone location and energy.

- Sequential Recombination Algorithms: They look pairwise at the four-momenta of all particles in an event and combine them according to a certain distance metric. Most often, the obtained jets have boundaries that are not circular, which is the case for the cone algorithms, by construction.

In addition to the jet algorithm and its free parameters one needs to specify a recombination scheme, which dictates how to add the 4-momentum of two particles when they are combined. Together, these two form a *jet definition*, i.e., a connection between what can be calculated and what is measured. For a general jet algorithm to be reliable, it is often imposed that it satisfies a couple of criteria [36]. First, it should be IRC safe and computationally feasible, i.e., it performs reasonably well in real event calculations, where particle multiplicities are $\mathcal{O}(100)$ in pp and $\mathcal{O}(1000)$ in PbPb, and it works both with whatever signature hadrons leave in a detector (experiment) and with partonic calculations (theory). The IRC safety requirement is qualitatively formulated just like mentioned in Section 2.2 - if one modifies an event by the addition of a soft or collinear emission, the set of hard jets found in the event, i.e., the ones formed by sucessive branching of the hard scattered partons, should remain unchanged. More precisely, the observables calculated from the reconstructed jet should yield finite results. Lastly, the reconstructed jet observables ought to be as insensitive as possible to non-perturbative effects, such as *hadronization* and the *underlying event*, and to experimental features, such as detector imperfections and *pile-up*. These hadronic collision features are skimmed through in Section 2.4.

In this work, we focus on sequential recombination algorithms together with the *E-scheme*, which simply sums the children 4-momenta to give the resultant parent 4-momentum. This type of algorithm tries to mimic the sucessive soft and collinear parton branchings from a pQCD viewpoint. Therefore, it decides to recombine a pair of particles based on a distance measure which is small whenever the QCD branching process is kinematically enhanced. Interestingly enough, this process, which is usually called a *clustering process*, may be taken outside the context of jet finding and be used to assign a clustering sequence to a jet. The *branching history* picture one takes from this assignment is introduced in Section 2.5 and it is of the utmost importance for this work.

The most often used algorithms for jet reconstruction in hadronic collisions are the IRC safe *generalized-k_t algorithms* [37]. Their clustering process is applied to the list of all particles in an event in the following manner:

1. For each particle compute all pairwise *inter-particle distances* d_{ij} and the *beam-distance* d_{iB}:

$$d_{ij} = \min(p_{t,i}^{2p}, p_{t,j}^{2p})\Delta R_{ij}^2, \qquad d_{iB} = p_{t,i}^{2p} R^2 \tag{2.19}$$

   where p is the parameter fixing the algorithm in this family, $p_{t,i}$ is the transverse momentum of particle i, relative to the beam direction, and $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the *angular distance* between two particles, measured as a geometrical distance on the rapidity(y)-azimuth(ϕ) plane. R is another free parameter of the algorithm, usually called the *jet radius* and taken as ~ 1 in this work. These distance metrics are all defined in terms of longitudinal boost invariant quantities (Section 2.4).

2. Find the minimum distance of all d_{ij} and d_{iB}:

- If it is a d_{ij} then cluster the particles i and j together, i.e., form a new pseudo-particle k which is the result of the recombination of i and j. Put pseudo-particle k as the replacement of i and j on the list.
- If it is a d_{iB} then i is identified as a jet and taken out of the list.

Go back to step 1 if the particle list has any particles left in it.

The result is a set of jets, i.e., a set of four-momenta, and a cluster sequence assigned to each one of them. Notice that these algorithms, for any value of p or R, favour the clustering of particles which are angularly close to each other - collinear branching. Also, if two particles satisfy $\Delta R_{ij} > R$, then they won't be recombined since the beam distance is most certainly smaller. Hence, R is typically seen as a measure of a jet's size. In hadronic collisions, one often focuses on a few specific values of p:

- The k_t-*algorithm* [38] ($p = 1$), which favours the clustering of low p_t particles first by associating a small *inter-particle distance* to a soft branching. However, the fact that it is so sensitive to soft radiation, makes it underperform in the presence of extra soft radiation from background effects.
- The *Cambridge/Aachen algorithm* (C/A) [39] ($p = 0$), which has a purely geometrical distance metric, making it less sensitive to soft radiation than the k_t-*algorithm*. This algorithm is more often used for the clustering hierarchy it produces, something further detailed in Section 2.5.
- The *anti-*k_t *algorithm* [40] ($p = -1$), which is currently the standard algorithm used at the LHC experiments for jet finding and the one used for that purpose in this work. The main reason for this is that it is practically insensitive to extra soft radiation, reconstructing the jet by starting from the hardest particle and clustering collinear radiation around it until the geometrical distance goes over R. Therefore, not only is the algorithm robust against background radiation, e.g., *underlying event* and *pile-up*, but it also results in circle shaped jets in the $y - \phi$ plane (Fig. 2.5). Note that, in contrast with the previous two algorithms, the clustering sequence derived from this algorithm lacks direct interpretation as a branching history [41].

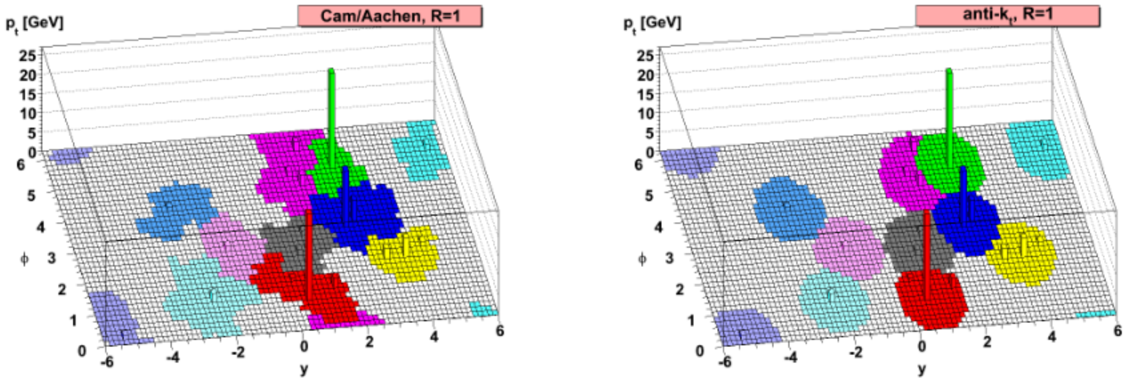


Figure 2.5: The $y - \phi$ plane for jets obtained with the C/A (left) and anti-k_t (right) algorithms, for jet radius $R = 1$. Taken from [27], originally from [40].

2.4 Hadron Collisions and Collider Variables

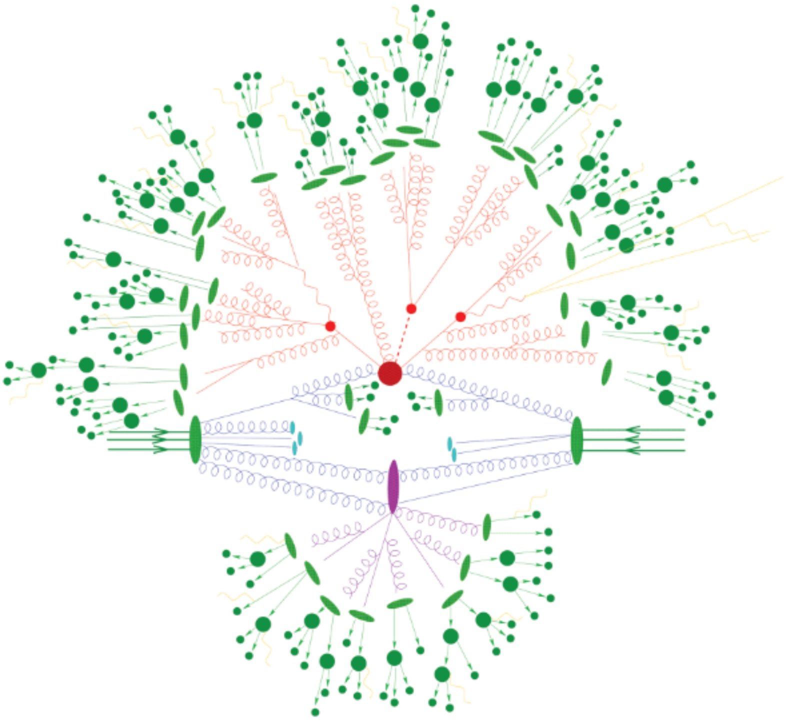


Figure 2.6: Sketch of a typical hadronic collision. The focus is the red radiation - the final state shower - coming off the partons produced in the hard scattering (red blob). Taken from [42].

Hadron collisions provide an extraordinary opportunity to study QCD at disparate energy scales - from assymptotic freedom of partons to their confinement inside hadrons (Section 2.1). In this work, we'll be focusing on proton-proton collisions, which, together with the concepts introduced in Sections 2.5 and 2.6, provide a great framework for a detailed study of the space-time structure of QCD jets.

These events are extremely complex in that they produce a very high amount of information, i.e., the kinematics of all the produced $\mathcal{O}(100)$ final particles. These particles - hadrons, leptons, photons, neutrinos - can be originated in a plethora of physical processes allowed by the Standard Model of Particle Physics. Of these, the average number of final hadronic particles which reach detectors is $\mathcal{O}(10)$. A picture of a typical hadron collision is shown in Fig. 2.6, with colors identifying each part. A multitude of processes ranging disparate energy scales happen, all of which must be simulated by Monte-Carlo event generators in order to reflect the full complexity of the event. Below, the components of a full event are introduced on a decreasing energy scale (hard $\rightarrow$ soft).

The hardest process - the red blob in Fig. 2.6 - is the *hard scattering*, already mentioned in the beginning of Section 2.2 when introducing jets as emerging from the hard scattered primary partons. This process can be calculated through Feynman diagrams, i.e, in the fixed-order pQCD framework, and most calculations are standard $2 \rightarrow 2$ processes. The emission of secondary radiation - ISR and FSR, blue and red, respectively, in Fig. 2.6 -, was introduced by the end of Section 2.2 and is described in the framework of resummation of the kinematically enhanced contributions. Although their resolution scale evolution is done in opposite directions, both radiation modules are implemented through parton shower algorithms based on the Sudakov factor (Section 2.3). Note that, it is only after ISR generation that the PDFs are convoluted with the matrix element calculations to give the final hard cross section. This is because ISR is usually factorized into the PDFs at the hard scattering scale, i.e., one evolves PDFs from the hadronic scale to the hard scattering scale through ISR generation. Eventually, at scales $\mathcal{O}(\Lambda_{QCD})$, the full partonic final state is converted into a set of hadrons - *hadronization*, the green blobs in Fig. 2.6. This process is separated from the FSR module and is usually treated with phenomenological models. One such model, which is the one used in PYTHIA 8.2.35 [33], is the *Lund string fragmentation* [43] - a model standing on the idea of a linear confining potential. The idea is that, for instance, a pair $q\overline{q}$ is connected through a *color flux tube*, which stores potential energy and gets more energetic the further apart the partons are. At some separation point, it is energetically more favourable to break the ”string” and to create an additional pair between them. This happens successively and the end result is a smooth distribution of color charges between the two leading ones, which then combine to form a collection of detectable hadrons.

The *underlying event* - the purple blob and lines in Fig. 2.6 -, which is a source of background for jets and is described through models, accounts for additional parton interactions coming from the remaining partons in the colliding hadrons - multi-parton interactions. This background also results from *beam-remnant* interactions, stemming from soft interactions between the leftover hadronized parton ensembles, which may acquire little transverse momenta and reach detectors. Finally, one should mention *pile-up*. This is a characteristic of full real events, where bunches of protons are collided and multiple proton-proton collisions are registered by the detectors at the same time, making it difficult to discriminate a single collision’s signature.

Each section of this factorized event is plenty of study material on its own, but our interest lies on the final result - either the final partons (parton-level) or the color-neutral hadrons which they recombine into (hadron-level). We will mostly work at parton-level throughout the thesis.

From two initial energetic protons travelling towards each other along a well-defined line - the beam direction -, one gets a multitude of final-state, color-neutral hadrons. For the sake of clarity in the calculations encountered throughout this thesis, it is necessary to point out a feature of hadronic collisions. The actual CM of the partonic collision is not the same as that of the pp collision, which coincides with the laboratory frame, i.e., where detection is taking place. This should be evident when one observes that the hard colliding partons, one from each hadron, don’t necessarily carry the same longitudinal momentum. In fact, it can be shown that this CM has the non-zero rapidity $y^{coll} = \frac{1}{2}\ln\left(\frac{x_1}{x_2}\right)$, where x_1 and x_2 are the Bjorken x’s introduced in Section 2.1. For this reason, it is mandatory to use kinematic variables which are invariant under longitudinal boosts, for it would be cumbersome to compare calculations made in the partonic CM with measurements made in the laboratory frame. Not only this, but non-transverse quantities are usually not conserved, since a part of the energy goes along the beam direction, that is often experimentally inaccessible. Hence, defining the transverse plane relative to the beam line, one writes the four-momentum of a particle as,

$$p^\mu = (E, p_x, p_y, p_z) = (m_t \cosh y, p_t \cos\phi, p_t \sin\phi, m_t \sinh y) \tag{2.20}$$

where $m_t = \sqrt{p_t^2 + m^2}$ is called the transverse mass, $m = \sqrt{p^\mu p_\mu} = \sqrt{E^2 - p_t^2 - p_L^2}$ is the parton's virtuality/invariant mass, ϕ is the azimuthal angle, p_t is the transverse momentum and y is the rapidity. These last two variables are defined by,

$$p_t = \sqrt{p_x^2 + p_y^2}, \qquad y = \frac{1}{2}\ln\left(\frac{E + p_L}{E - p_L}\right) \tag{2.21}$$

where $p_L = p_z$ is the longitudinal momentum relative to the beam line. The energy and transverse momentum are then related by $E = m_t \cosh y$. Note that, in Eq. (2.20), the four-momentum is written in terms of 3 longitudinal boost invariant quantities and the rapidity, which transforms trivially as $y \to y - y_b$, where y_b is the boost's rapidity. This trivial transformation implies that rapidity differences are also invariant under longitudinal boosts.

Another relevant quantitiy which is constantly used in Chapter 4 and which we already mentioned in Section 2.3 when introducing jet algorithms is the *rapidity-azimuth* distance between two partons:

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \tag{2.22}$$

Additionally, in an experimental setting, it is often preferable to use the *pseudo-rapidity*, which is directly related to the angle to the beam line θ:

$$\eta = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_L}{|\vec{p}| - p_L}\right) = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.23}$$

This quantity η doesn't behave like y under longitudinal boosts, but it is identical to it in the massless limit. An important observation to make is that there is a threshold for angular resolution ΔR_{ij} in most experiment's hadronic calorimeters, such as ATLAS and CMS at the LHC. These detectors, usually cylindrical detectors, are maps of a coordinate space spanning $\phi \in [0, 2\pi]$ and $\eta \in (-\eta_{max}, \eta_{max})$, where η_{max} is the maximum pseudo-rapidity coverage (set by the detector's finite length), and are discretized into multiple cells of typical dimensions 0.1×0.1 in (η, ϕ). Therefore, two particles of a jet need to be sufficiently distanced so as to be resolved as two different energy depositions.

2.5 Jet Fragmentation Histories From Clustering Algorithms

Over the last years, jet substructure - the study of observables calculated from the internal structure of a jet - has found plenty of practical applications (cf. [27]) and it is particularly useful because it provides complementary insights to the usual inclusive full jet observables. This field of study was initially developed for discriminating jets that originated from the decay products of boosted (high energy to virtuality ratio) electroweak resonances, such as the W/Z bosons, or by a hard QCD parton ([44], [45]. In current days, jet tagging is still the main task of jet substructure studies, although with a much more widespread definition of *signal* and *background* jets. Not only this, but more sophisticated strategies and observables, e.g. *grooming* algorithms ([46],[47]) and the *Lund jet plane* [48], have been developed and the problems to which they can apply are beyond jet tagging, e.g., mitigation of non-perturbative or extra soft radiation effects and the study of jet-medium interactions in heavy-ion collisions [41].

An important stage of these studies is finding a representation of a jet's internal evolution - its *fragmentation history*. One could take advantage of current parton shower generators and extract this structure directly from the Monte Carlo simulation. However, in an experimental context, one does not have direct access to this internal information. Hence, this task boils down to building a hierarchy of $1 \to 2$ emissions, calculated in a direction opposite to that of a parton shower evolution - from a set of final-state particles which belong to a reconstructed jet, sequentially *cluster* them until there is only one resultant particle. This is exactly what one would do in jet reconstruction, except that one is now interested in the *clustering sequence*. The merging decision is done through a jet algorithm and the resulting merged cluster is calculated with a recombination scheme. Because this procedure is applied after a jet is found, it is often called a *reclustering procedure*, where each merging step is seen as a *splitting*, or a branching, and each cluster's 4-momentum is identified with a virtual particle. The result, in a computational sense, is a binary tree structure where each particle is associated with two *children* particles, its 4-momentum being given by the recombination of the children's 4-momenta. From the pQCD perspective, this structure aims at resolving the branching history of a jet in the parton shower interpretation, i.e., in the collinear limit of factorized multiple emissions. It should be clear that this splitting sequence is an oversimplified view of a parton shower which, in itself, is a LL order probabilistic implementation of the evolution of a hard parton's quantum wave function into a final multi-parton state.

Because there's freedom in choosing what pair of particles to cluster at each step of the procedure, one usually has a large combinatorial number of possible fragmentation histories for a given list of jet particles. From a quantum mechanical point of view, it makes sense that there isn't a unique possibility. Nevertheless, since the branching pattern ought to reproduce at least some of the basic QCD features, such as angular ordering or the collinear divergences, this number can be brought down. Good candidates for jet algorithms assigning clustering sequences which agree, to a certain extent, with this basic restriction are the generalized-k_t algorithms for $p \geq 0$, introduced in Section 2.3. The distance metrics (the *inter-particle* distances) of these algorithms follow the QCD singularity structure because they are small, i.e., favour the clustering decision, in phase space regions where emissions are logarithmically enhanced.

The other reason why this family of algorithms, for non-negative values of p, is suitable for building a jet's fragmentation pattern is how they order the splittings - the clustering sequence is such that consecutive branchings are ordered according to valid QCD ordering variables (see Section 2.3). If one wants to give any physical meaning to this internal representation by associating it to a parton shower, then it should correctly reproduce QCD emissions at least to leading-logarithmic order.

In this work, one will draw brief comparisons between four different reclustering algorithms - three algorithms from the generalized-k_t family (for values of $p = 0,\, 1/2,\, 1$) and the *Jade algorithm*:

- The k_t-*algorithm* $(p = 1)$:

  $$d_{ij} = \text{min}(p_{t,i}^2, p_{t,j}^2)\Delta R_{ij}^2 \tag{2.24}$$

  Just like any generalized-k_t algorithm for $p > 0$, this algorithm's distance metric follows the double QCD kinematic enhancements. The first splitting of the resulting internal structure will be the hardest one since the algorithm favours the clustering of soft particles first. This is particularly useful when one intends to divide the jet into a number of hard subjets, e.g., the two hardest subjets are the products of the first splitting, i.e., the last reclustering step.

- The *C/A algorithm* $(p = 0)$:

  $$d_{ij} = \Delta R_{ij}^2 \tag{2.25}$$

  This algorithm resolves the jet based only on its angular structure. In the resulting branching history, the last reclustering step will correspond to the largest emission angle branching. One should then expect the tree structure to be an *angular-ordered* shower.

- The τ *algorithm* $(p = 1/2)$ (see, e.g., [15]):

  $$d_{ij} = \text{min}(p_{t,i}, p_{t,j})\Delta R_{ij}^2 \equiv d_{ij}^{\tau} \tag{2.26}$$

  This distance metric, as we shall in Section 4.4, is proportional to the inverse of an approximation of the parton formation time introduced in Section 2.6. As such, one expects this algorithm to produce a branching tree which is ordered in formation time.

- The *Jade algorithm* [49]:

  $$d_{ij} = 2E_iE_j(1 - \cos\theta_{ij}) \tag{2.27}$$

  where E_i and E_j are the particle energies and θ_{ij} the emission angle between them. It was first introduced in the 1980's by the JADE collaboration in the context of e^+e^- annihilation [49] and is not used jet finding in hadronic collisions. It can, however, serve as a reclustering algorithm. The distance metric is the mass m_{ij}^2 of the pair (i, j), i.e., the virtuality of their parent particle, in the massless limit. However, this is only true for massless on-shell particles, i.e., massless final-state particles. One would expect the fragmentation pattern assigned by the *Jade* algorithm to be virtuality-ordered if each of each of the i and j virtualities are small enough (usually if $m_i/p_{t,i} \ll 1$).

The four reclustering procedures are said to be LL equivalent, i.e., the observables calculated with information from each of the clustering hierarchies agree up to differences in terms of NLL order. Although a comparison between them is drawn in Section 4.3, one will focus mainly on the τ algorithm through the remaining of the analysis. This decision is based on the formation time ordering one gets from the resulting clustering sequence, which, in principle, is a good representation of a jet's space-time structure.

2.6 From Parton Formation Time to Jet Space-Time Structure

It can be said that a virtual parton, i.e. an off mass-shell parton emerging from some process, lives for a certain amount of time before it radiates another parton, redistributing energy and momentum. One can think of this quantity as how long it takes for a parton to become an independent colored object, i.e., a source of further colored particles [50].

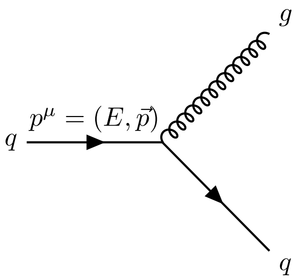


Figure 2.7: Example of a parton branching diagram $q \to qg$. We're set to calculate the formation time of the quark with momentum p^μ, which is assumed to come from some previous interaction, i.e, it is a virtual quark.

Throughout this work, one shall refer to this lifetime as the *formation time* of the subsequent pair of states a virtual state evolves into. A naive estimate of it can be made by resorting to the Heisenberg's uncertainty principle [50],

$$t_{rest}\Delta E \sim 1 \tag{2.28}$$

where t_{rest} is defined as the formation time in the parton's rest frame and one is working in natural units, i.e, $\hbar = c = 1$. Note that, throughout this work, one will often convert formation times from GeV^{-1} back to time units fm/c by multiplying the result by $\hbar c \sim 0.197$ GeV·fm. The quantity ΔE is the amount of energy the virtual parton would need to become on mass-shell and is given by $\Delta E = m = \sqrt{p^\mu p_\mu}$, the parton's virtuality (Fig. 2.7). In this calculation, one is assuming that, for quarks, their virtualities are much larger than their on mass-shells. Otherwise, ΔE would have to take them into account. If one wants to calculate this time in a reference frame where the parton is moving in a given direction, then one should Lorentz boost it according to,

$$t' = \gamma(t_{rest} - \vec{\beta} \cdot \Delta\vec{x}_{rest}) = \gamma t_{rest}, \qquad \vec{\beta} = \vec{p}'/E' \text{ and } \gamma = (1-\beta^2)^{-1/2} = E'/m \tag{2.29}$$

where $\Delta\vec{x}_{rest}$ is the parton's displacement in time t, which vanishes in the rest frame, and $\vec{p}'/E'$ is the velocity of the parton as measured in the desired reference frame, with $\vec{p}'$ its three-momentum and E' its energy. In what follows, we consider 3 distinct reference frames: the laboratory frame, coinciding with the hadron collision's CM, the transverse frame, where the parton has no longitudinal momentum, and the hard partonic collision's CM.

In the first case, the boost is simply in the direction of the parton's movement as measured in the laboratory frame, i.e., $\vec{\beta}^{rest\to lab} = \vec{p}/E$ and, therefore, $\gamma^{rest\to lab} = E/m$. Quantities measured in the laboratory frame shall be written without subscript. Since $\Delta x_{rest} = 0$ the formation time in the laboratory frame is

$$\tau \sim \frac{E}{m^2} = \frac{m_t \cosh y}{m^2} \equiv \tau_0 \tag{2.30}$$

where $m_t = \sqrt{m^2 + p_t^2}$ is the transverse mass introduced in Section 2.4 and is the same in any longitudinally boosted reference frame. Throughout the remaining of this thesis, one shall refer to τ_0 as the *exact estimate*. This expression for the parton formation time can also be derived through the more rigorous calculation of the single emission amplitude, assuring the parton is produced off-mass shell in some subprocess (Section A.1).

For the transverse reference frame, one can boost the laboratory frame in the longitudinal direction using the parton's longitudinal velocity, i.e., $\vec{\beta}^{lab\to tv} = \vec{p}_L/E$. The energy E_{tv} in this reference frame can then be calculated by the Lorentz transformation law for the 0th component of the 4-momentum:

$$E_{tv} = \gamma^{lab\to tv}(E - \vec{\beta}^{lab\to tv} \cdot \vec{p}) = \frac{E}{\cosh y} = m_t \tag{2.31}$$

The energy could have been calculated more readily by noting that, in this frame, there's only transverse movement: $m^2 = E_{tv}^2 - p_t^2$. The formation time in the transverse frame is given by:

$$t_{tv} \sim \frac{m_t}{m^2} \tag{2.32}$$

Note that the transverse and laboratory frame formation times are related by

$$\tau_0 = \cosh y \cdot t_{tv} \tag{2.33}$$

where $\cosh y$ is the relative boost $\gamma^{lab\to tv}$ between the two frames, which is a simple longitudinal boost. Note that, although relation (2.33) is different for each parton because it depends on each one's rapidity, in the collinear approximation the radiation is contained inside a very narrow cone and, therefore, the parton rapidities have negligible differences among them. This means that, for the narrowest jets, one can simply write $\tau_0 = \cosh(y^{jet}) \cdot t_{tv}$ for every constituent. Furthermore, by taking the jet rapidity to be $\ll 1$, which is approximately true for jets with large angles to the beam line, i.e., $p_L/p_t \ll 1$, then we get the lab and the transverse reference frames as equivalent: $\tau_0 = t_{tv}$.

The third and final reference frame is unique to hadron collisions, where the coordinates of CM of the actual partonic collision differ from the lab coordinates, as explained in Section 2.4. The residual longitudinal momentum sum $\vec{p}_L^{coll}$ (measured in the laboratory frame) is to be used if one wants to Lorentz boost the laboratory frame to the hard partonic collision's CM, i.e., $\vec{\beta}^{lab\to hCM} = \vec{p}_L^{coll}/E^{coll}$, with E_{coll} the collision energy. By using a similar argument as for the previous calculation and noting that $\gamma^{lab\to hpCM} = \cosh y^{coll}$:

$$E_{hpCM} = \cosh y^{coll}\left(E - \frac{p_L^{coll}}{E^{coll}} p_L\right) = E\frac{\cosh\left(y - y^{coll}\right)}{\cosh y} = m_t \cosh\left(y - y^{coll}\right) \tag{2.34}$$

where $y^{coll} = \frac{1}{2}\ln(x_1/x_2)$ as mentioned in Section 2.4. The parton formation time in the hard partonic collision CM is then given by:

$$t_{hpCM} \sim \cosh\left(y - y^{coll}\right)\frac{m_t}{m^2} = \cosh\left(y - y^{coll}\right)\cdot t_{tv} \tag{2.35}$$

Looking at relations (2.33) and (2.35), if one considers both the collision and the parton rapidites to be $\ll 1$, the former being verified if the colliding hard partons have similar longitudinal momenta, i.e., similar Bjorken x's, then the 3 reference frames are equivalent. Hence, in this approximation, the formation time is equal to t_{tv} in all of them.

In the result analysis of Chapter 4, we will be focusing on the parton formation time as measured in the laboratory frame. In Section 4.4, one will expand formula (2.30) in two different ways - solely considering the kinematics of the 2 subsequent partons (local) and going all the way through to the final real particles (global). As we shall see, the former is mainly instructive, while the latter seems to be experimentally more useful.

Another important timescale in jet evolution is the hadronization time t_h, which imposes an upper limit on the formation time. Essentially, it is the time during which a hard scattered parton behaves as if it were a free coloured particle. It can be shown [51] that the condition $\tau_0 < t_h$, i.e., the parton is formed as an independent coloured radiation source before it hadronizes, amounts to the pQCD regime.

With a branching history assigned to a jet and a formation time associated with each of the splittings in it, one can further calculate a *space-time* structure. Although we will be dealing with formation times throughout this thesis and not do any mention of actual positions, the conversion from time intervals to space intervals is simply done by a velocity, which is approximately that of light, and by the emission angles of partons. Hence, we will continue to refer to the resulting structure as *space-time* structure.

In this work, two ways of *ordering splitting* are taken into account - the reclustering order and the *absolute time* order. The first order is exactly the inverse sequence by which partons were recombined, with the *inter-particle* distances decreasing from the last step to the first step. For the τ algorithm, it is expected that this splitting order is approximately the same as ordering the splittings in increasing formation time. The second order tries to impose an absolute time scale (denoted by $\Sigma\tau$ in Section 4.5.1) to the fragmentation history by recursively summing formation times along the tree. The calculation is simply

$$(\Sigma\tau)^i = \sum_j^i \tau^j \tag{2.36}$$

where the j index is a symbol for every parton that comes before parton i in the branch it belongs to. Algorithmically, starting from the *seed* splitting, for which the absolute time is equal to the formation time, one travels along the binary tree's branches and sums the previous absolute time $(\Sigma\tau)^{i-1}$ with the current parton formation time τ^i to get the current absolute time $(\Sigma\tau)^i$. Afterwards, one can order the splittings by increasing order in absolute time. An interesting remark to make is that these two splitting orders, even for reclustering applied with the τ algorithm, are not necessarily the same. The reason for this may lie in the spacing between sucessive formation times or in the fact that the orders may alternate between branches - this is further explored in Section 4.5.1. Additionally, one should note that the splitting orders used in this work and the parton shower's usual orderings mentioned in Section 2.3 are separate concepts.

By having an order whereby one refers to splittings, one can have a representation of how formation times are distributed along the jet splittings, together with other possible observables and correlations which one presents in Section 4.

Chapter 3

Simulation and Reconstruction Setup

In this work, we chose pp collisions as our framework to study the space-time structure of QCD *vacuum* jets, those without any background medium like the QGP, and generate such events with the general purpose Monte Carlo PYTHIA 8.2.35 [33], with the option `Tune:pp=5`, corresponding to ”Tune 4C”. We simulate 10^6 events at a CM energy of 5 TeV and use the final particle kinematic information for further analysis. The final-state kinematics are stored into and read from, event-by-event, a `TTree`, a class which is part of the data analysis framework ROOT 6.18.00 [52]. The analysis consists on finding the jets in each event and then calculating the space-time structure by assigning them a fragmentation history and calculating both formation times and absolute times for each one, in a manner described in Section 2.6. The sequential recombination algorithms used in both clustering and reclustering procedures, along with any extracted jet kinematical information, are dealt with by resorting to FASTJET 3.3.2 [37], the standard tool for such a task. Before diving into this work's results, one should detail the setup and parameters of the event generation and subsequent analysis.

The event generator PYTHIA 8.2.35 uses a $p_\perp$-ordered shower [53] and the hadronization model described in Section 2.4 - the Lund string fragmentation [43]. The hard process is generated from matrix element calculations at LO and we'll be focusing on events of the type Z+jet, e.g., $qg \to Zq$ (Fig. 3.1 on the left). The Z+jet process is generated with pure Z contributons, i.e., there is no interference term between photon and Z production. The options corresponding to this setting are `WeakZ0:gmZmode=2`, `WeakBosonAndParton:qqbar2gmZg` and `WeakBosonAndParton:qg2gmZq`. Experimentally, Z+jet is typically a much cleaner event than Dijet events (e.g., $gg \to gg$ (Fig. 3.1 on the right)), simply because there's a single hard scattered parton. Not only that, but the possibility of retrieving the Z boson's kinematic information, which is done by studying its decay products, gives one additional information about the jet's kinematics. In this work, the Z boson's decay is turned off, because we are not interested in finding it and so that there isn't the possibility of a jet arising from it. In terms of kinematic cuts, the products of the hard scattering are forced to have a transverse momenta above 90 GeV (the `ptHatmin` parameter of the setup) and the events are only accepted for Z bosons with transverse momenta $p_t^Z \geq 100$ GeV and rapidities $|y^Z| \leq 2$. In order to avoid the statistical error degradation caused by the rapid decay of the p_t

spectrum, one reweights events by oversampling higher p_t^Z through `PhaseSpace:bias2SelectionPow = 6.3`.

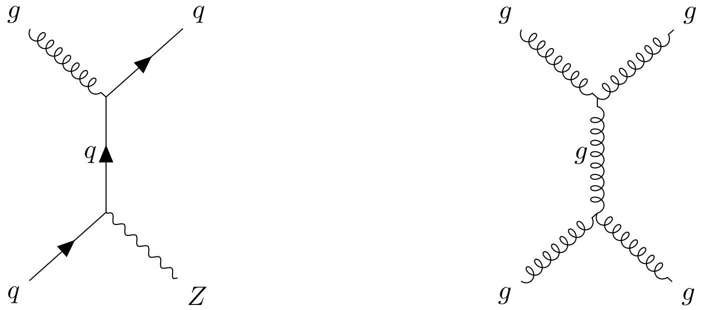


Figure 3.1: Examples of hard interaction diagrams for Z+jet, $qg \to Zq$ on the left, and Dijet, $gg \to gg$ on the right. For Z+jet there is a single hard scattered parton (the final quark) and for Dijet there are two hard scattered partons (the two final quarks).

As we have seen in Section 2.4, the pp collision products not only go beyond what we are interested in (FSR), but are also experimentally detected at hadron-level. The final state population is then composed of hadronic jets and will include additional soft hadrons. This will interfere with the analysis of QCD jets at the partonic level. It is then imperative that, at least qualitatively, one knows their impact on the quantities we are studying. One is specifically concerned with the effects of ISR and hadronization.

Initial state radiation populates the final state with further particles which may end up in the direction of the jet's axis. By adding particles to the reconstructed jet whose content, in an idealized situation, would only consist of the radiation emitted off the hard scattered parton, ISR is said to ”contaminate” our jet, whether we're working at parton - increased parton multiplicity - or hadron-level - increased hadron multiplicity. In some cases, ISR might even produce the largest p_t jet of the event.

Hadronization is a non-perturbative phenomenon which is unavoidable in an experimental context, as already explained in previous sections. In simulations, one can stop the generation right before the hadronization model is applied, leaving one with the final state partons, i.e., a parton-level simulation. The question of whether parton-level results can be translated to hadron-level, which is where measurements take place, then arises. A possible answer lies on the idea of Local Parton Hadron Duality (LPHD [50], [54], [55]), which essentially suggests that, after accounting for parton emissions down to scales $\sim \Lambda_{QCD}$, the transition from parton to hadrons is local in phase-space with corrections accounted by multiplicative factors. This indicates that the momenta and multiplicities of hadrons will be closely related to those of the partons. Although the model PYTHIA 8.2.35 [33] uses for hadronization - the Lund string model [43] - is more complex than this, it seems to be in agreement with the LPHD idea [55]. No further quantitative analysis of hadronization corrections is done in this work.

Although we will be focusing on parton-level FSR showers in the later part of Section 4, we will still try to do a simple exploration of the sensitivity of a few suitable observables to the impact of ISR and hadronization effects. To this end, in the beginning of Section 4, one will vary the reconstruction radius and the reclustering algorithm and see how we can suppress the influence of these effects on the space-time structure of jets (the main focus of this work), while trying to reconstruct most of the emitted final state radiation. With regards to background processes, we include the beam-remnants but shut off multi-parton interactions. None of what concerns detector-level effects or pile-up is considered in this work.

The jet algorithms described in Section 2.3 are implemented with the FASTJET package [37]. For jet reconstruction, one uses the anti-k_t algorithm due to its extra soft radiation robustness (Section 2.3) and the jet radius $R \sim 1$ is set to a value large enough because, as already mentioned, we want to capture most of the FSR inside a single jet. With regard to kinematical acceptance criteria, the final partons are filtered according to $p_t^{\text{parton}} \geq 0.5$ GeV and $|y^{\text{parton}}| \leq 3.5$ and we'll be working with jets with transverse momenta $p_t^{\text{jet}} \geq 20$ GeV and rapidities $|y^{\text{jet}}| \leq 2.5$. The reconstructed jet and the Z boson should be separated azimuthally by $\Delta\phi = 7\pi/8$. After a large jet is reconstructed, one applies the reclustering procedure with the algorithms described in Section 2.5, together with the E-scheme. In this procedure, the jet radius is larger than the one used in reconstruction, so as to assure every single jet particle is taken into account in the fragmentation history. The class `PseudoJet` from FASTJET is repeatedly used in the `C++` code developed for this work, for its usefulness in storing 4-momenta and in extracting various kinematic quantities from them. Most graphical results displayed throughout this thesis were done by resorting to the ROOT framework.

Chapter 4

Results Analysis

In this chapter, the relevant results obtained in this work are presented and discussed in a complete manner. In the first instance, the compromise between jet reconstruction efficiency and ISR contamination is analysed, followed by a study of the sensitivity of the first splitting's formation time to ISR and hadronization on its distribution and its correlation with the jet mass. Afterwards, one compares space-time structures for branching histories obtained with different algorithms, followed by the exploration of two approaches to the calculation of parton formation time - local and global - and an analysis of various kinematic limits. Finally, a comprehensive investigation of the space-time structure of a partonic FSR jet, with a branching history fixed by the τ algorithm, is made. One explores the possible orderings of formation time inside a jet and what orders of magnitude its distribution spans.

4.1 Reconstruction efficiency and ISR contamination

One of the parameters of a jet definition is the jet radius R, which is a measure for the typical size of a jet (Section 2.3), allowing one to tune how much radiation goes inside it. In our work, the aim is that the reconstructed jet contains as much FSR as possible, i.e., that one can reconstruct, as best as possible, the 4-momentum of the hard scattered parton which is produced alongside the Z boson in the hard scattering. By doing so, from the final partons (or hadrons), one would then apply the reclustering process detailed in Section 2.5 and make an attempt at representing the full branching history without any information from the Monte Carlo generator (as was done in, e.g., [12]). However, from a given collection of final particles in a collision, one doesn't necessarily reconstruct a single jet, even if the hard scattering process produces a single parton, as is our case. There can be multiple reasons for this, but one is mainly concerned with the possibility of ”fat” jets, i.e., first emissions which are energetically symmetric and have a wide opening angle (originating, typically, two smaller jets) and the existence of ISR which, as mentioned in Chapter 3, necessarily contaminates the reconstructed jet. The solution for the former problem would be to increase the jet radius until the reconstruction is satisfactory. Unfortunately, because the direction of the radiation coming off the initial colliding partons isn't necessarily correlated with the beams' direction, one should not increase the jet radius carelessly, for that would cause more ISR to be included in the reconstructed jet. Hence, there is a need for a trade-off between finding out as much FSR as possible and supressing the impact of ISR in relevant jet observables (Section 4.2).

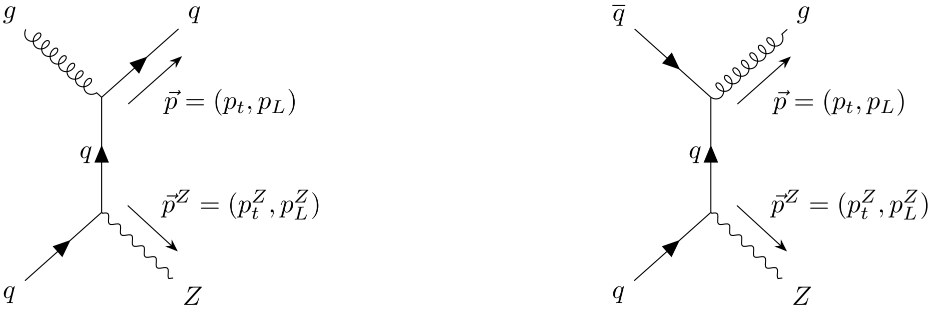


Figure 4.1: Hard interaction diagrams for Z+jet event type, along with kinematics of the hard scattered particles. Because we assume the initial partonic collision carries no transverse momentum, we necessarily have $p_t = p_t^Z$. Hence, in this work, an idealized jet reconstruction would have $p_t^{jet}/p_t^Z = 1$.

Firstly, one will analyse the efficiency of the reconstruction by resorting to the Z boson's transverse momentum p_t^Z. Here, as described in Chapter 3, the Z boson's decay is turned off and we take its p_t^Z directly from the Monte Carlo generator. Note, however, that, in an experimental context, one could find the Z boson's p_t^Z equally well by studying its decay products. If we consider the initial colliding partons to have negligible transverse momenta, which is true for this work's simulation setup, then the hard scattered parton and the Z boson are produced with exactly the same transverse momentum (Fig. 4.1). Therefore, one can quantify our reconstruction's efficiency, i.e., how much of the FSR we include, by evaluating the ratio of the jet's transverse momentum p_t^{jet} to the Z's transverse momentum p_t^Z.

In Fig. 4.2, one shows how this ratio changes when we open up our jet's radius and the influence of that when we take ISR into account. At a first glance, it is clear that, in comparison with the FSR reconstruction for the remaining radii (Fig. 4.2c and 4.2e), reconstruction with $R = 0.5$ (Fig. 4.2a) performs poorly, missing out on between $1\% - 5\%$ (bins between 0.95 and 0.99) of the FSR shower's content for a relevant portion of events. This is most visible when looking at the lower tails and peak values of all FSR plots. Then, as one would expect, $R = 1.75$ reconstructs FSR better than $R = 1$ - an average ratio closer to unity and a halved standard deviation. In spite of this, when looking at the results with the addition of ISR in each event, there's a significant increase of the distribution's tail for ratios above 1 when going from $R = 1$ to $R = 1.75$ - Fig. 4.2d) and 4.2f, respectively. This suggests that using the larger radius ($R = 1.75$) causes ISR to take over the jet's radiation content. Hence, reconstruction radii in the range $R \in (1, 1.75)$ seem to be the best compromise.

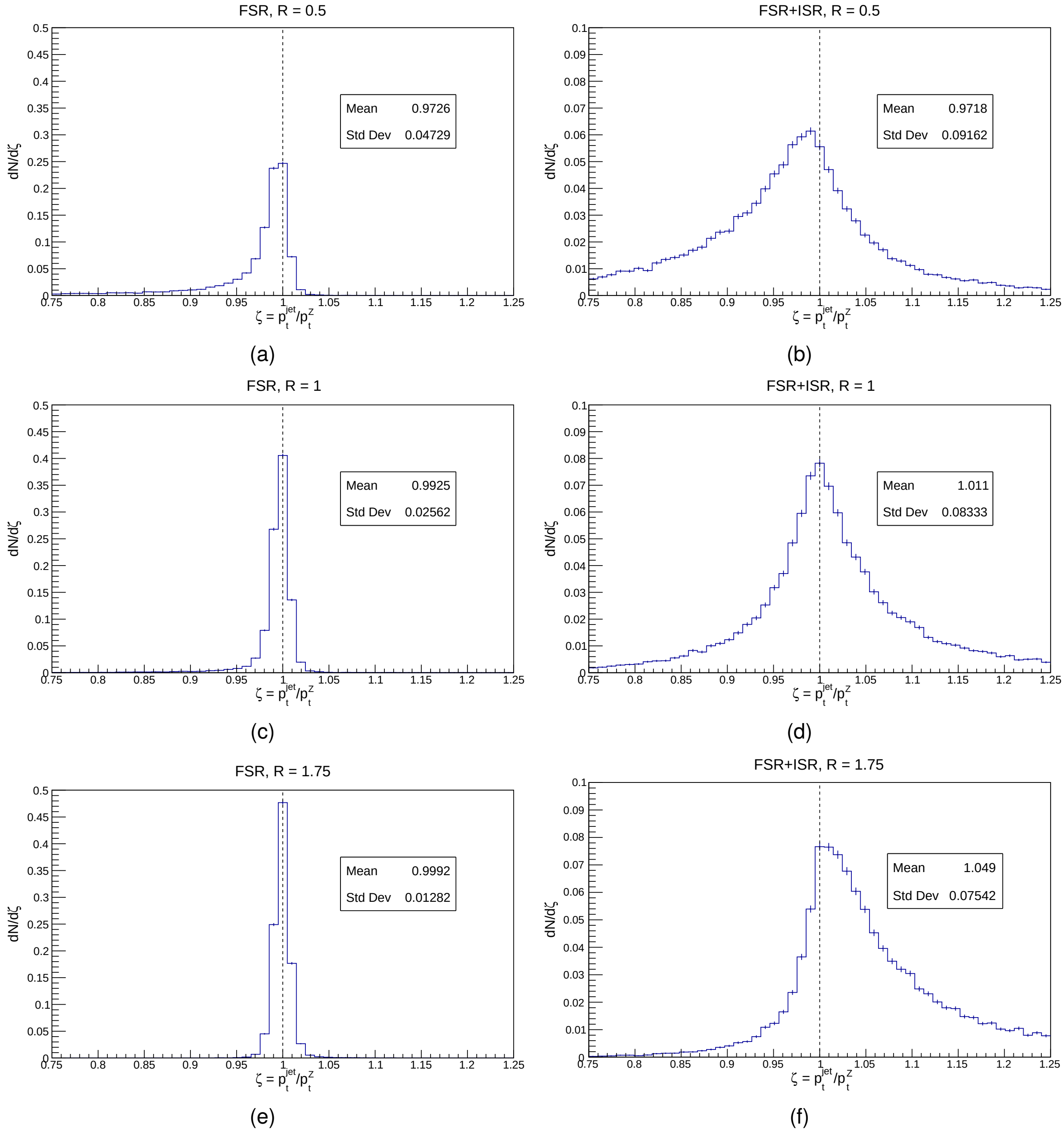


Figure 4.2: Ratios p_t^{jet}/p_t^Z for various reconstruction radii, with and without Initial State Radiation. All distribution are normalized to unit integrals.

4.2 First splitting's formation time

The parton formation time derived in Section 2.6 (Eq. (2.30)) only depends on the 4-momentum of the parton we are considering. This implies, in particular, that the formation time of the parton leading to the first splitting (the "seed" parton), only depends on the recombination scheme and not on the specific branching history one calculates, i.e, it is independent of the reclustering algorithm. However, one should keep in mind that, as we shall see next, this quantity still depends on what one captures inside the jet. In fact, as mentioned in Section 4.1, an optimal reconstruction would return the hard scattered parton's 4-momentum, so that one would be calculating its formation time.

Because we are using the E-scheme - the sum of 4-momenta - the seed parton's 4-momentum is exactly equal to the reconstructed jet's, making the first splitting's formation time uniquely dependent on the jet's kinematic variables:

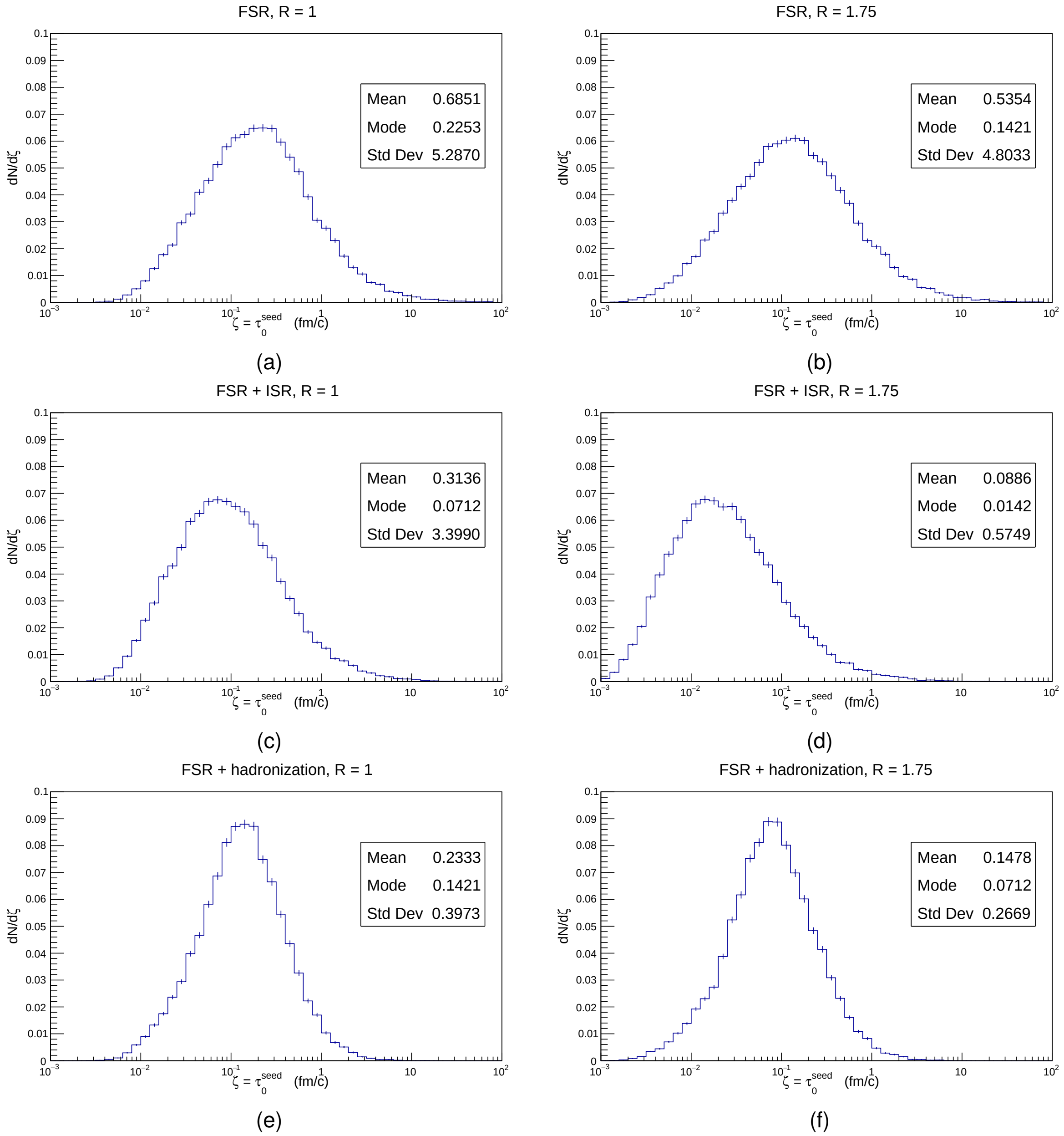
$$\tau_0^{\text{seed}} = \frac{E^{\text{jet}}}{(m^{\text{jet}})^2} = \frac{p_t^{\text{jet}}\sqrt{1+(m^{\text{jet}}/p_t^{\text{jet}})^2}\cosh y^{\text{jet}}}{(m^{\text{jet}})^2} \tag{4.1}$$

The relevance of this conclusion is that this quantity, as one shall see in Section 4.5.1, is a proxy for the shortest splitting timescale inside a jet whose branching history is calculated with the τ algorithm [15], the reclustering algorithm we will focus on for the most part of this thesis. Not only this, but this calculation can be made directly through experimentally measurable kinematic properties of a jet. For this reason, its estimation is of significant importance for studies involving the initial instants of QGP production, which happens concurrently with jet production.

4.2.1 Sensitivity to ISR and hadronization

Now that the interplay between effectively reconstructing FSR and supressing the inclusion of ISR is quantitatively understood, one ought to measure the effect of ISR and hadronization on relevant observables. In this work, because the main task is to calculate a space-time structure for a jet, we will be focusing on the sensitivity of formation times. In particular, we will analyse the first splitting's formation time or, as used equivalently throughout the text, the "seed" formation time.

In what follows, one varies the reconstruction radius and analyses the changes in the distribution of the first splitting's formation time when simulating events with (1) FSR, parton-level, (2) FSR + ISR, parton-level and (3) FSR, hadron-level, i.e., adding hadronization. The aim is to understand if one can minimize the difference in formation time when including either one of these generator modules, while trying to gain knowledge about where it comes from. In Fig. 4.3, the statistical distribution of seed formation times is shown for two jet radii, $R = 1$ and $R = 1.75$ (left and right columns, respectively), for the three mentioned setups, each corresponding to a line of plots. Only the portion of each distribution in the range $(10^{-3}, 10^{2})$ (fm/c) is shown, for the total range spans multiple orders of magnitude (further explored in Section 4.5.2). This also explains why the mode and the mean are so disparate for every distribution. With respect to increasing the reconstruction radius, the mean and the mode of the seed formation time decreases for whatever setup we are considering. Including ISR results in an overall shift of the first splitting's formation time to lower values, as one can observe by the decrease in the mean and mode values. This is especially true when adding ISR to FSR for a radius $R = 1.75$ - the mode is reduced by a factor of 1/10 (Fig. 4.3b and 4.3d). One can explain this at the light of the results of Section 4.1. The fact that we use a wider jet definition opens up the possibility for more ISR to be reconstructed with the FSR jet. Not only this, but if we consider that this radiation is mainly soft, then, considering Eq. (4.1), an increase in jet mass would be enough for the formation time to decrease as drastically as we see in Fig. 4.3d. The hadronization has the main effect of squeezing the bulk of the distribution into a smaller range of values, while slightly shifting the mode value. If one assumes that it keeps approximately all the energy inside the jet reconstructed at parton-level, then only the jet's mass can vary. This can happen since the the hadronization model rearranges the final (mostly massless) partons into final (massive) hadrons which, in turn, have further decays. Overall, both modules seem to cause a suppression of larger formation times, while either shifting (ISR) or maintaining (hadronization) the lower tail of the distribution. Moreover, one confirms once again that the jet definition cannot be too wide - $R = 1.75$ reveals a greater sensitivity of the seed formation time to both ISR and hadronization. Future work can include the usage of substructure techniques to mitigate the effect of ISR (e.g. grooming algorithms [46]), reducing its impact on formation time calculations and allowing for a more efficient reconstruction of FSR. For the remaining sections of this work, one will focus on jets reconstructed with radius $R = 1$.



(a) (b)

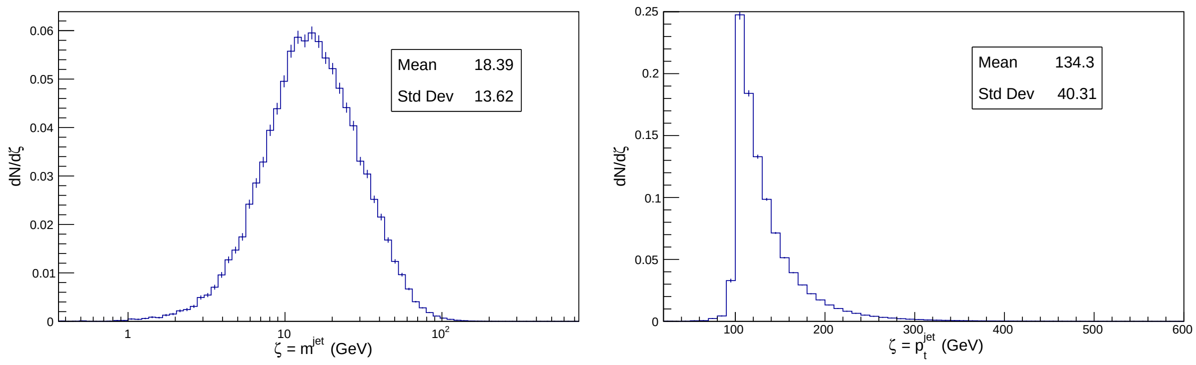
(c) (d)

(e) (f)

Figure 4.3: First splitting's formation time distributions for the 3 setups mentioned in the text and with reconstruction radii $R = 1$ and $R = 1.75$. As for most distributions encountered throughout the thesis and which cover more than 1 order of magnitude, the bins of each histogram are logarithmic, i.e., equally spaced in logarithmic space. All plots are normalized to unit integrals.

4.2.2 Correlation with Jet Mass

In the previous section we were led to the conclusion that the reconstruction of the jet's 4-momentum is all we need to calculate the first splitting's formation time τ_0^{seed}. An interesting follow-up question is whether one can reduce this dependence to a single experimentally measurable quantity, taking the remaining dependences into account through a dispersion measure. To do that, one should optimally choose a quantity that correlates well with τ_0^{seed}. By looking at Eq. (4.1), the most straightforward choice is the jet's mass m^{jet}. The reason for this choice of variable, rather than E or p_t, is that the dispersion of the mass spectrum, in relative terms, is much greater than the dispersion of either one of those spectra, as can be seen in Fig. 4.4. Not only this, but m^{jet} covers a wider range of orders of magnitude. Alternatively choosing either one of these two kinematic variables would result in a highly dispersed τ_0^{seed} for each fixed bin, with a span of values corresponding to all possible masses for a specific value of p_t^{jet} or E^{jet}.



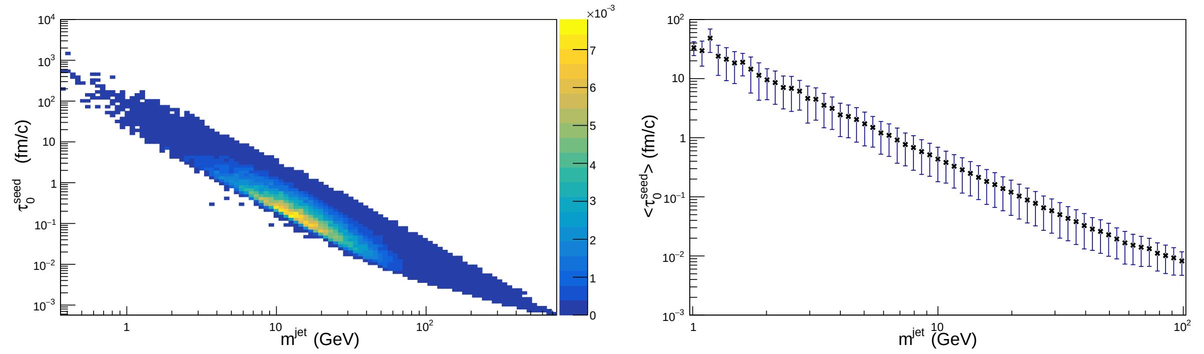
(a) (b)

Figure 4.4: a) Jet mass m^{jet} spectrum. b) Jet transverse momentum p_t^{jet}spectrum. Both results are parton-level and are normalized to unit integrals.

Thus, in Fig. 4.5a, the scattering plot of the first splitting's formation time as a function of the jet mass is shown and one immediately verifies that, indeed, these two variables are strongly correlated. Through inspection of Eq. (4.1), the mean $\langle\tau_0^{\text{seed}}\rangle$ and dispersion σ_τ in each bin of jet mass are trivially given by, respectively, the mean value and the dispersion of the energy spectrum at that specific jet mass:

$$\langle\tau_0^{\text{seed}}\rangle = \frac{\langle E\rangle}{m^2} \qquad\qquad \sigma_\tau(m) = \frac{\sigma_E(m)}{m^2} \tag{4.2}$$

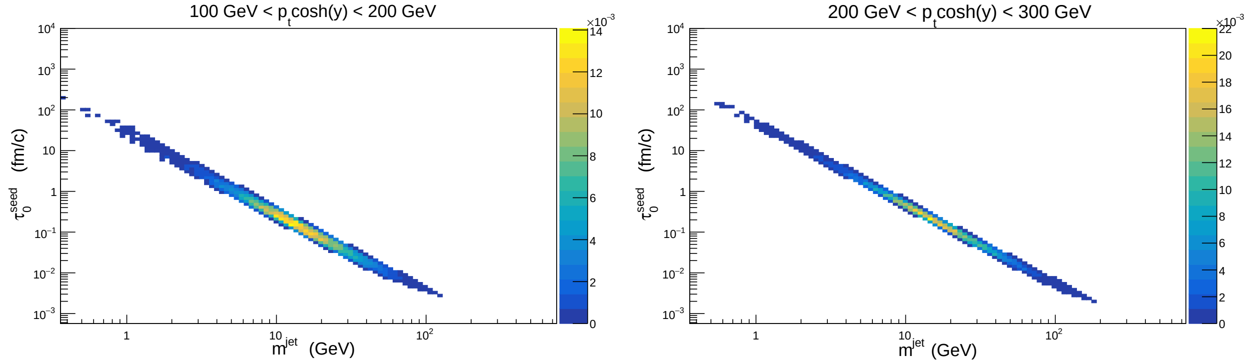
where one dropped the "jet" superscript for clarity's sake and will do so for the remaining of the section. Furthermore, for all jet masses, there is a significant probability that the seed formation time is between 50% below and above the mean value we see in Fig. 4.5b, i.e., $\sigma_\tau/\langle\tau\rangle \sim 0.5$. This is a direct consequence of the independence of the jet energy from its mass - for all jet masses one has $\sigma_E(m) = \sigma_E$ - and of the the properties of the energy spectrum - the standard deviation $\sigma_E \approx 136$ GeV and the mean value $\langle E\rangle \approx 233$ GeV.



(a)

(b)

Figure 4.5: a) Scattering plot of first splitting formation time τ_0^{seed} and jet mass m^{jet}. b) Average first splitting formation time for each jet mass in the range 1 GeV $< m^{\text{jet}} < 200$ GeV, along with the standard deviation calculated for each mass bin. Both results are parton-level.



(a)

(b)

Figure 4.6: Scattering plot of first splitting formation time τ_0^{seed} and jet mass m^{jet} for phase space cut a) 100 GeV $< p_t \cosh y < 200$ b) 200 GeV $< p_t \cosh y < 300$ GeV. Both results are parton-level

A more accurate prediction of the seed formation time demands a cut in the jet's phase space. Restricting the jet's energy is not suitable, for it is not longitudinal boost invariant. In Section 2.4 we introduced collider variables and mentioned their importance in an experimental context, so it is natural that one plots the correlation for bins of, for instance, $p_t \cosh y$. Because $E \approx p_t \cosh y$ (which is valid since m/p_t is, on average, of order 10^{-1} and always smaller than unity), it is evident that using $p_t \cosh y$ as kinematic cut variable will reduce the dispersion in Fig. 4.5. In Fig. 4.6, one shows the correlation plot between the seed formation time and jet's mass for two different $p_t \cosh y$ bins with a width of 100 GeV. Then, in Fig. 4.7, the average formation times and corresponding standard deviations are calculated for each jet mass bin, for 4 distinct $p_t \cosh y$ bins with the same 100 GeV width. In comparison with Fig. 4.5b, the dispersion decreases significantly, allowing for a more accurate prediction of the first splitting's formation time. For the $p_t \cosh y$ bin that results in the largest dispersion (100-200 GeV, the red markers in Fig. 4.7), one can expect the seed formation time to be within a relative interval of no more than $\sim 20\%$ around the mean value. Additionally, as expected, increasing the value of $p_t \cosh y$ results in an increase of formation times. In fact, each window of $p_t \cosh y$ corresponds to a well defined line of average formation times with the same slope, just different offsets. Each offset could be extracted by fitting, for each kinematic window, a function of the form $\langle\tau_0^{\text{seed}}\rangle(m) = a/m^2$, with a the fitting parameter. This parameter, after dividing by $\hbar c$ to adjust units, would represent the average jet energy in each $p_t \cosh y$ bin.

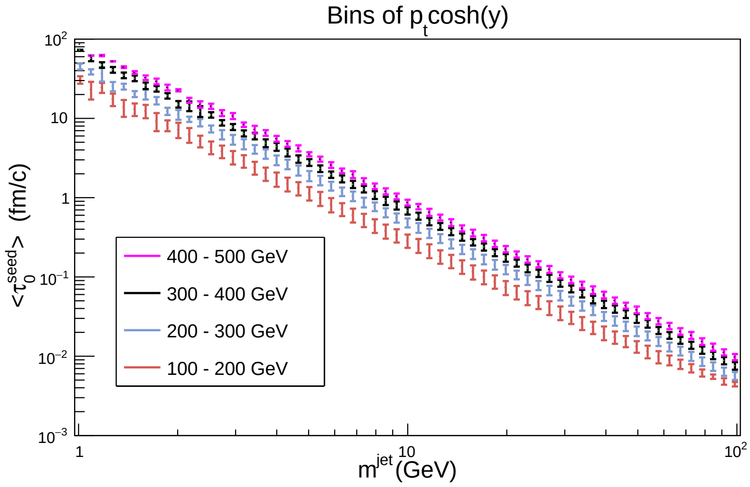


Figure 4.7: Average first splitting formation time for each jet mass (roughy the center of each marker) in the range 1 GeV $< m^{\text{jet}} < 100$ GeV, along with the standard deviation calculated for each mass bin. Results are presented for various $p_t \cosh y$ bins and at parton-level.

4.3 Reclustering algorithm comparison

In the previous sections one thoroughly analysed the seed formation time. The next step is to ask how the remaining formation times are distributed throughout the branching history. One will postpone the detailed answer to this question to later sections. In this section, one will explore how the average formation times of each splitting vary between jet histories calculated with the 4 different algorithms introduced in Section 2.5.

Let us first focus on Fig. 4.8. It shows an example of two different branching trees calculated from the same set of final partons of a single event - in Fig. 4.8a a "τ tree" and in Fig. 4.8b a "C/A tree". After the final parton (in red) indices, in this case numbers $0-6$, the internal parton (in blue) indices are ordered according to the reclustering steps, i.e, 7 is the first reclustering step and 12 the last one. The quantity assigned to each internal line corresponds to that parton's formation time in fm/c. Firstly, we confirm that the first splitting's formation time is the same for both trees. For this particular set of final partons, the first reclustering step is exactly the same for both algorithms - they both merge partons 3 and 6 first. Despite the fact that the two reclustering sequences are different for the remaining steps, there are formation times which are equal in both trees. This is caused by partons which, although formed through a different sequence of steps, give rise to identical subjets - they have the same 4-momentum. That is the case for partons 9 and 10, from which partons $\{0,1,2\}$ and $\{3,4,6\}$ emerge, respectively. These partons do not necessarily have to be formed at the same reclustering step, as is the case. Rather, the main point is that they are the result of merging the exact same final partons. This conclusion is the underlying motivation for the *global* approach to formation time detailed in Section 4.4.

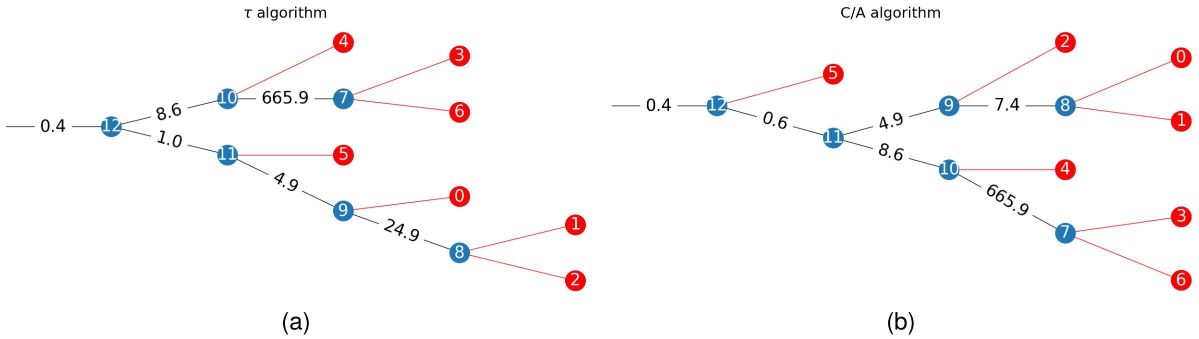
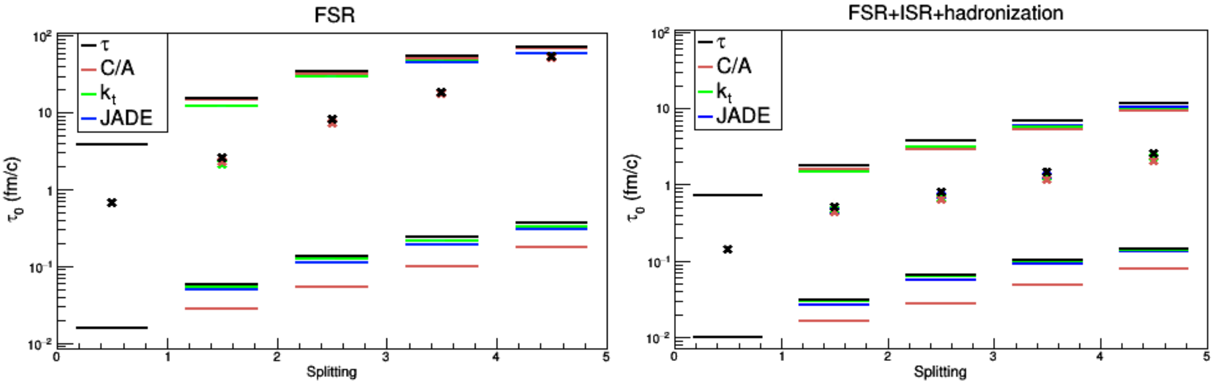


Figure 4.8: Space time structure for a branching history calculated with a) τ algorithm b) C/A algorithm. Red nodes/lines correspond to final partons, while the blue nodes (with corresponding black line) correspond to internal partons. Each formation time assigned to each black line is in fm/c and the 4-momenta of partons are not shown for readability's sake. The results are parton-level FSR. The figures were generated through a Python package for network algorithms [56], with information extracted from a single generated event.

Following this example, one can try to quantify, over multiple events, the differences between the formation time distribution of each algorithm's tree. In Fig. 4.9, we present the mean formation time of each splitting, for the first 5 splittings, ordering them according to the absolute time splitting order one mentioned at the end of Section 2.6. In this order, the first splitting is the one with the smallest absolute time, i.e, the seed splitting, the second splitting the one with the second smallest and so forth. It is worth stating that these results are inclusive in the sense that they represent jets with any particle multiplicity. Also in the same plots, one shows the 5% and 95% quantiles of each splitting's formation time distribution and for each algorithm. These distributions are investigated with more detail in Section 4.5 for the τ algorithm.

Firstly, with respect to adding both ISR and hadronization (Fig. 4.9b) to parton-level FSR (Fig. 4.9a), one concludes that all mean formation times decrease as a consequence of that. This is a generalization of what we saw in Section 4.2.1, where we concluded the same for the first splitting's formation time, examining the impact of ISR and hadronization individually. Additionally, at least for the first 5 splittings, the difference between the two sets of mean formation times increases as one moves deeper into the jet, having a minimum at the first splitting. For our working condtions, the mean seed formation time of a jet produced in an event with FSR, ISR and at hadron-level is roughly 5 times smaller than that of an event with only FSR at parton-level.

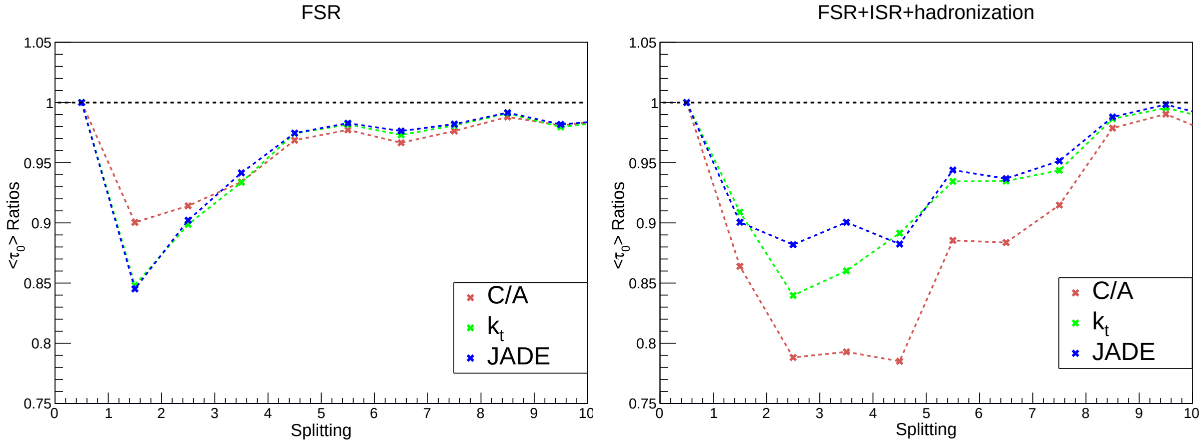
Looking at the quantiles for each algorithm, it seems that all algorithms except for C/A generally agree on the overall dispersion of each splitting's formation time distribution. The lower quantiles of the C/A's space-time structure are significantly lower than the remaining, indicating that the C/A's formation times are slightly more dispersed and take small values ($\sim 10^{-2}$) more often. The relation between the four algorithms' mean formation times is more clearly seen through ratios with respect to a reference - in this case, one chose the τ algorithm. Hence, in Fig. 4.10, for the remaining algorithms, each splitting's ratio corresponds to the mean formation time of that splitting divided by the same quantity calculated for a τ tree. This is shown for the first 10 splittings. Looking at both plots one concludes that, with or without ISR and hadronization, the τ space-time structure attributes, on average, larger formation times than any other algorithm we are considering. One can further conclude that the most striking differences between algorithms are seen in earlier splittings, except for the seed formation time which, as we saw in Section 4.2.1, is independent of the branching history. For jet-medium interactions, these early splittings are most important, so a more detailed analysis of such differences would be relevant in future work. The algorithm's dissimilarities are amplified when one includes ISR and works at hadron-level (Fig. 4.10b), suggesting that the different branching histories agree to a higher degree when working with events of parton-level FSR. In particular, in Fig. 4.10a, the 3 algorithms seem to be well in agreement with each other and, in Fig. 4.10a, the C/A curve jumps to much lower values and distances itself from the k_t and JADE curves, which still reasonably agree. Nonetheless, the mean formation times of each branching tree are of the same order for every splitting, follow the same trend and differ, at most, by $\sim 25\%$.



(a)

(b)

Figure 4.9: Mean formation time for each splitting according to their order in absolute time, along with the 5% and 95% quantiles which are represented by the the lower and upper solid lines, respectively. The results are shown for the 4 different algorithms, for the first 5 splittings and for events of a) FSR and b) FSR+ISR+hadronization. Note that each point is plotted in the center of each bin, e.g., the first splitting corresponds to 0.5.



(a)

(b)

Figure 4.10: Ratios of the mean formation times of each splitting, having the τ algorithm as a reference. Results are presented for the remaining algorithms, for the first 10 splittings and for events of a) FSR and b) FSR+ISR+hadronization.

4.4 Formation Time Estimates and Kinematical Dependences

Let us focus on parton-level FSR showers and branching histories calculated with the τ algorithm from now on. In this section, one will explore the kinematic dependences of the formation time, whether starting from the kinematics of the two subsequent partons (local approach) or from the kinematics of a relevant subset of final partons (global approach). By taking certain kinematic limits, one will arrive at approximated expressions for formation time and evaluate their validity as estimates of formation time.

4.4.1 Local Approach

Consider the calculation of the parton formation time introduced in Eq. (2.30) by taking into account the kinematics of the two subsequent partons (Fig. 4.11).

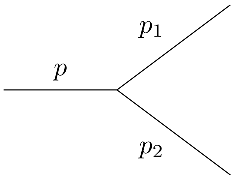


Figure 4.11: Branching history tree with 3 final partons. Each parton's 4-momentum is represented by p_i. The numbering follows no specific order.

Let us calculate $m^2 = p^2 = (p_1 + p_2)^2$:

$$m^2 = m_1^2 + m_2^2 + 2p_1 \cdot p_2 = m_1^2 + m_2^2 + 2(E_1E_2 - |\vec{p}_1||\vec{p}_2|\cos\theta_{12}) \tag{4.3}$$

where θ_{12} is the opening angle between the two partons and $\vec{p}_i$ their respective three-momenta. By rewriting the third term using the collider variables introduced in Section 2.4 and the virtuality of each parton, one ends up with:

$$m^2 = m_1^2 + m_2^2 + 2p_{t,1}p_{t,2}\left(\sqrt{1 + m_1^2/p_{t,1}^2}\sqrt{1 + m_2^2/p_{t,2}^2}\cosh(y_1 - y_2) - \cos(\phi_1 - \phi_2)\right) \tag{4.4}$$

One can conclude from this expression that, even if a particle splits into two almost collinear particles, its formation time is not necessarily large because the subsequent particles may have a non-negligible virtuality.

Expressing the energy of the initial parton as $E = E_1 + E_2$ with $E_i = p_{t,i}\sqrt{1 + m_i^2/p_{t,i}^2}\cosh y_i$, the formation time in Eq. (2.30) can finally be written as:

$$\tau_0 = \frac{E}{m^2} = \frac{p_{t,1}\sqrt{1 + m_1^2/p_{t,1}^2}\cosh y_1 + p_{t,2}\sqrt{1 + m_2^2/p_{t,2}^2}\cosh y_2}{m_1^2 + m_2^2 + 2p_{t,1}p_{t,2}\left(\sqrt{1 + m_1^2/p_{t,1}^2}\sqrt{1 + m_2^2/p_{t,2}^2}\cosh(y_1 - y_2) - \cos(\phi_1 - \phi_2)\right)} \tag{4.5}$$

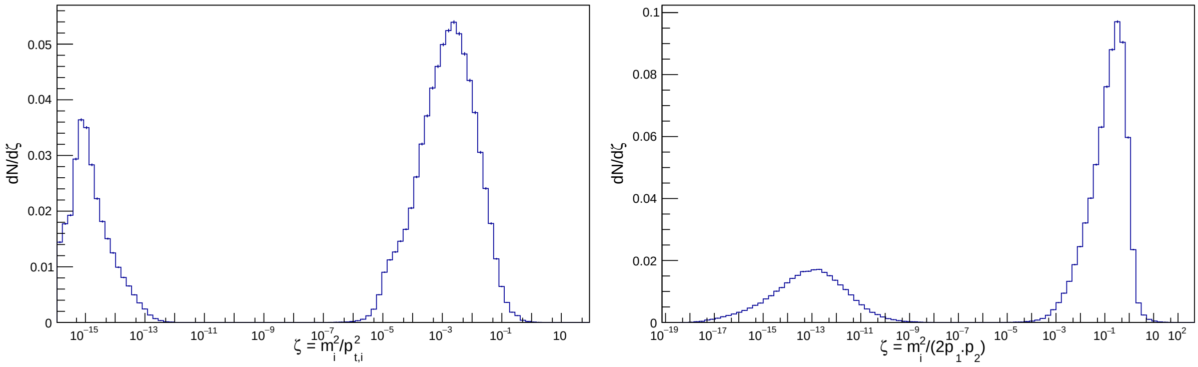
To write the formation time fully in terms of collider variables, one would need to neglect virtualities m_i^2. There are two relevant terms involving virtualities m_i^2 in Eq. (4.5). The terms m_i^2 in the denominator can be ignored if $m_i^2/2p_1 \cdot p_2 \ll 1$. By neglecting them, one gets:

$$\tau_1 \equiv \frac{E}{m^2 - m_1^2 - m_2^2} \tag{4.6}$$

Then, there is an approximation to be made on the term $2p_1 \cdot p_2$ and on the numerator, which is $m_i^2/p_{t,i}^2 \ll 1$. Applying this would result in:

$$\tau_2 \equiv \frac{p_{t,1}\cosh y_1 + p_{t,2}\cosh y_2}{2p_{t,1}p_{t,2}\left(\cosh(y_1 - y_2) - \cos(\phi_1 - \phi_2)\right)} \tag{4.7}$$

The following plots (Fig. 4.12a and 4.12b) show the distributions for the relevant terms involving virtualities, for each parton on the branching history except the seed parton. Notice that there are two peaks in each distribution - presumably, the one positioned at lower values mostly belongs to final-state partons, i.e., almost real $m_i^2 \approx 0$ partons.



(a)

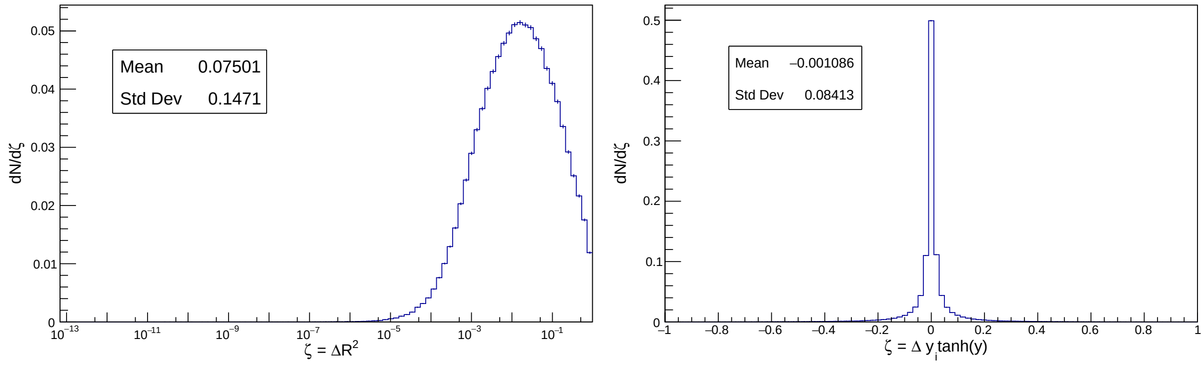
(b)

Figure 4.12: a) Factor controlling the first order corrections of the square-root term in Eq. (4.5), plotted for each final parton. b) Factor controlling the smallness of the m_i^2 terms in Eq. (4.5), plotted for each parton of each children pair. Both plots are normalized to unit integrals.

Further approximations can be made on Eq. (4.7). Specifically, in the case where the product partons are close in angle - the small-angle (collinear) approximation $\Delta y^2, \Delta\phi^2 \ll 1$ -, we end up with

$$\tau_3 \equiv \frac{(p_{t,1} + p_{t,2})\cosh y}{p_{t,1}p_{t,2}\Delta R^2} \tag{4.8}$$

where ΔR is the rapidity-azimuth distance introduced in Section 2.4 and $\cosh y_1 \approx \cosh y_2 \approx \cosh y$, where y is the parent's rapidity. To first order in $(y_i - y) \equiv \Delta y_i$, one has $\cosh y_i \approx \cosh y + \Delta y_i \sinh y$. For Eq. (4.8) to be a reasonable formation time estimate, one has to confirm that the angular distance satisfies $\Delta R^2 \ll 1$ and that $\Delta y_i \tanh y \ll 1$ for a significant percentage of the entries, which is indeed what Fig. 4.13a and 4.13b) show. On average, the first order expansion for $\cosh y_i$ seems to be valid, translating into a relative error of, at most, $\pm 20\%$ in approximating $\cosh y_i \approx \cosh y$.



(a) (b)

Figure 4.13: a) Angular distance ΔR^2 distribution for each pair of children partons. b) First order correction to the factor $\cosh y_i / \cosh y$ for each parton of each children pair. Both plots are normalized to unit integrals.

Finally, one can take the soft limit, which, for $z_i = p_{t,i}/(p_{t,1}+p_{t,2})$, takes the form $\min(z_1, z_2) \equiv z \ll 1$, (Fig. 4.14), and the expression one obtains from applying it to Eq. (4.8) is

$$\tau_4 \equiv \frac{\cosh y}{\min(p_{t,1}, p_{t,2})\Delta R^2} \tag{4.9}$$

which is exactly the inverse of the distance metric of the τ clustering algorithm (Section 2.5) multiplied by $\cosh y$. This result is the motivation behind Section 4.5.1, where one explores, among other orderings and time estimates, if the formation times calculated in this kinematic limit, for a branching history assigned by the τ algorithm, are ordered according to the reclustering steps.

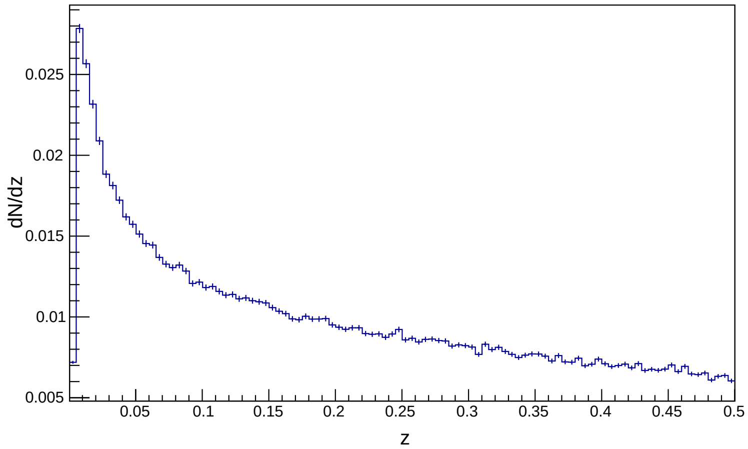


Figure 4.14: Softest transverse momentum fraction z distribution for each pair of children partons. Normalized to unit integral.

Interestingly enough, by comparing Eq. (4.9) with Eq. (2.33) in Section 2.6, one can identify:

$$t_{tv} \sim \frac{1}{\min(p_{t,1}, p_{t,2})\Delta R^2} \tag{4.10}$$

In this limit, the formation time calculated in the transverse reference frame, which is defined in Section 2.6, is exactly equal to the inverse of the τ algorithm's distance metric.

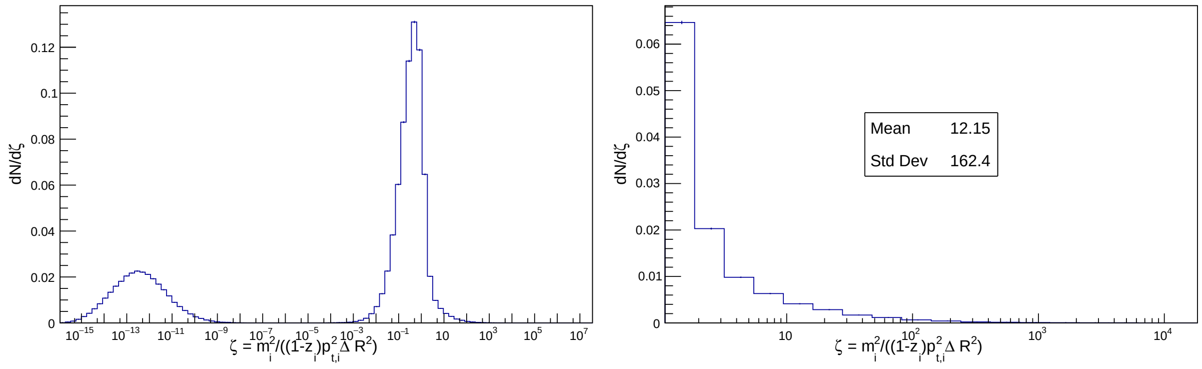
An interesting algebraic manipulation to do now is to check how these limits interfere with each other. The limits of low virtuality $m_i^2/p_{t,i}^2 \ll 1$, collinear partons $\Delta R^2 \ll 1$ and soft emissions $z \ll 1$ were done successively and ignoring any m_i^2 dependence, but one can apply them simultaneously while keeping terms to first order in $m_i^2/p_{t,i}^2$. Applying the two first approximations to the calculation of m^2 and E:

$$\begin{aligned} m^2 &\approx m_1^2 + m_2^2 + \left(\frac{p_{t,2}}{p_{t,1}} m_1^2 + \frac{p_{t,1}}{p_{t,2}} m_2^2 + p_{t,1}p_{t,2}\Delta R^2 + \mathcal{O}\left(\frac{p_{t,j}}{p_{t,i}} m_i^2 \Delta y_{12}^2\right)\right) \\ E &\approx (p_{t,1}+p_{t,2})\cosh y \left(1 + z_1\left[\Delta y_1 \tanh y + \frac{m_1^2}{2p_{t,1}^2}\right] + z_2\left[\Delta y_2 \tanh y + \frac{m_2^2}{2p_{t,2}^2}\right] + \mathcal{O}\left(z_i \frac{m_i^2}{p_{t,i}^2}\Delta y_i \tanh y\right)\right) \end{aligned} \tag{4.11}$$

where one kept the first order term Δy_i from the $\cosh y_i \approx \cosh y$ approximation on E and $\Delta y_{12} = y_1 - y_2$. The expression for m^2 turns explicit that one can't ignore m_i^2 without caring for the first order $m_i^2/p_{t,i}^2$ coming from the term $2p_1 \cdot p_2$, because at least one m_i^2 dominates over the corresponding $m_i^2/p_{t,i}^2$. For instance, in the soft limit, if one ignores m_1^2 in comparison with $(p_{t,2}/p_{t,1})m_1^2$ (meaning $p_{t,2}/p_{t,1} \ll 1$), then one should keep m_2^2 in comparison with $(p_{t,1}/p_{t,2})m_2^2$. These four terms also compete with the angular part of the expression - one can't ignore $m_i^2/p_{t,i}^2$ simply by assuming it to be $\ll 1$, because it competes with the smallness of the rapidity-azimuth distance ΔR^2. If one further manipulates m^2 in Eq. (4.11) and applies the soft limit $z \ll 1$, one can show that the formation time in Eq. (4.5) can be written as:

$$\tau_0 \approx \frac{\cosh y}{p_{t,2}\Delta R^2}\left(\frac{1}{z_1} + \Delta y_1 \tanh y + \frac{m_1^2}{2p_{t,1}^2}\right)\left(1 + \frac{1}{(1-z_1)\Delta R^2}\frac{m_1^2}{p_{t,1}^2}\right)^{-1} \tag{4.12}$$

where we picked $z = \min(z_1, z_2) = z_2 \ll 1$, i.e., $z_1 \approx 1$. Hence, in the low virtuality, collinear and soft limits, the corrections to Eq. (4.9) are controlled by $\frac{1}{(1-z_i)\Delta R^2}(m_i^2/p_{t,i}^2)$. This quantity is plotted in Fig. 4.15a. Note that there is a non-negligible probability (about 10% of the distribution) that the correction factor is above $\mathcal{O}(1)$ and, in those cases, Eq. (4.9) underestimates the formation time in Eq. (4.5) by a factor which can go up to orders of magnitude and is, on average, $\sim$ 10 (Fig. 4.15b).



(a) (b)

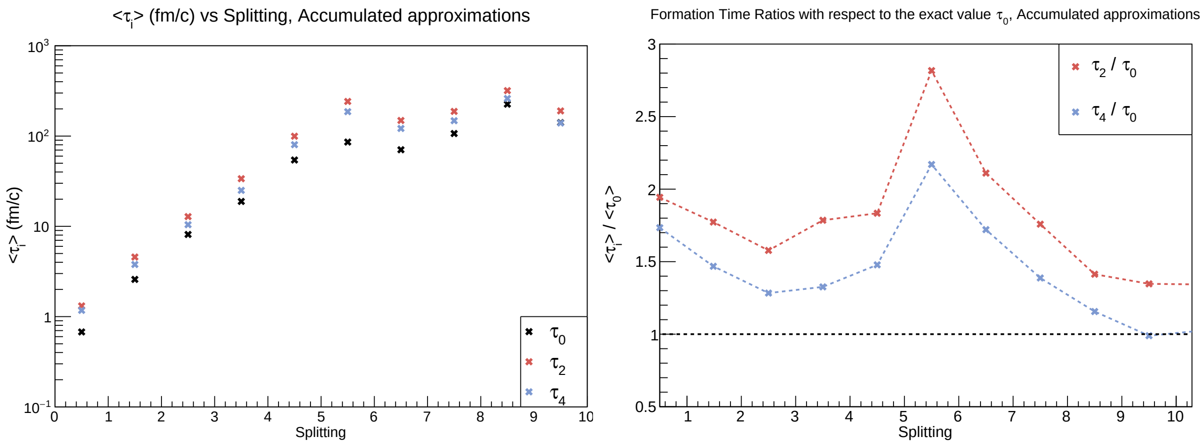
Figure 4.15: a) Factor controlling the major corrections to Eq. (4.9), which takes into account possible enhancements to the small $m_i^2/p_{t,i}^2$ term in the soft and collinear limit. The plot is normalized to unit integral. b) Zoomed distribution in the region $\frac{1}{(1-z_i)\Delta R^2}(m_i^2/p_{t,i}^2) > 1$.

The previous analysis motivates the study of the impact of all approximations when applied cumulatively and individually. In Fig. 4.16, we show the formation time distributions calculated with two of the approximate estimates, in comparison with the exact estimate τ_0 in Eq. (4.5). We present the estimates for low-virtuality (τ_2, Eq. (4.7)) and for the accumulation of this approximation with the collinear and soft limits (τ_4, Eq. (4.9)). In Fig. 4.17, we show the same comparison but for formation time estimates with only a single approximation (τ_1 and the $\overline{\tau}_i$'s). Both plots are shown without error bars for the sake of readability. The local approximations applied individually result in

$$
\begin{aligned}
\overline{\tau}_1 &= \tau_1 = \frac{E}{m^2 - m_1^2 - m_2^2} \\[2ex]
\overline{\tau}_2 &= \frac{p_{t,1}\cosh y_1 + p_{t,2}\cosh y_2}{m_1^2 + m_2^2 + 2p_{t,1}p_{t,2}\left(\cosh(y_1 - y_2) - \cos(\phi_1 - \phi_2)\right)} \\[2ex]
\overline{\tau}_3 &= \frac{\left(p_{t,1}\sqrt{1 + m_1^2/p_{t,1}^2} + p_{t,2}\sqrt{1 + m_2^2/p_{t,2}^2}\right)\cosh y}{m_1^2 + m_2^2 + 2p_{t,1}p_{t,2}\left(\sqrt{1 + m_1^2/p_{t,1}^2}\sqrt{1 + m_2^2/p_{t,2}^2}(1 + \Delta y_{12}^2) - (1 - \Delta\phi_{12}^2))\right)} \\[2ex]
\overline{\tau}_4 &= \frac{p_{t,h}\sqrt{1 + m_h^2/p_{t,h}^2}\cosh y_h}{m^2}
\end{aligned}
\tag{4.13}
$$

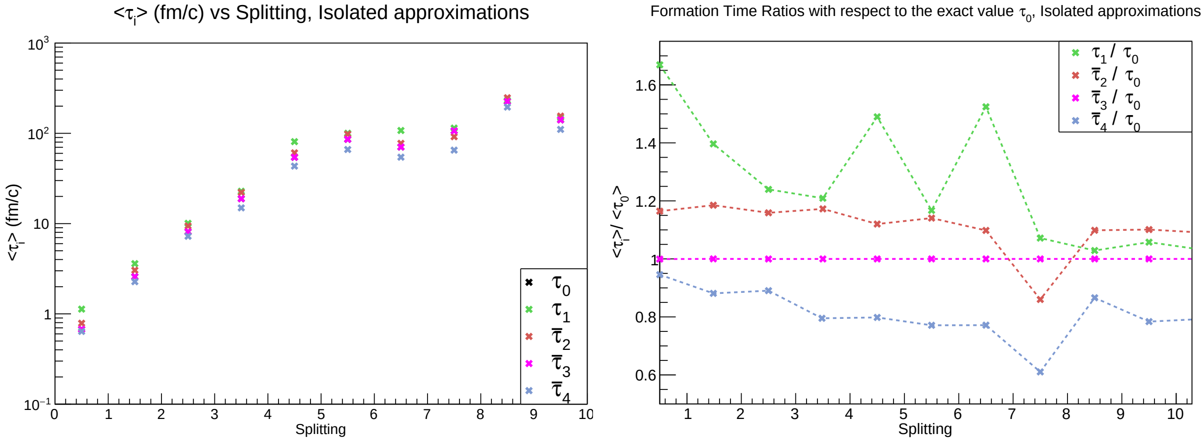
where $\Delta y_{12} = y_1 - y_2$, $\Delta\phi_{12} = \phi_1 - \phi_2$, h is the hardest parton index and we assumed, in $\overline{\tau}_4$, that the $m_i^2/p_{t,i}^2$ are of the same order. The estimates' numbers match those of the cumulative approximations in Eqs. (4.6), (4.7), (4.8) and (4.9).

Firstly, when we look at the ratio of τ_1 in Fig. 4.17b, we confirm that ignoring the m_i^2's altogether is the the most impactful approximation. A reasonable explanation for this is that, as we have seen in Eq. (4.11) in the limit $m_i^2/p_{t,i}^2 \ll 1$, at least one of the virtualities dominates over the corresponding term $p_{t,j}m_i^2/p_{t,i}$. Moreover, in comparing Fig. 4.16b and 4.17b, it is made clear that the approximations are not independent of each other. For instance, observe the difference between the τ_2 and the $\overline{\tau}_2$ ratios. The ratio of $\overline{\tau}_2$ is close to 1 while the ratio of τ_2 is about a factor of 2 higher, i.e., $\tau_2 \sim 2\overline{\tau}_2$. This indicates that ignoring terms $m_i^2/p_{t,i}^2$ when both m_i^2's aren't present isn't reasonable. The reason for this is that at least one term containing a virtuality must compete with the azimuth-rapidity distance ΔR^2, which is what we had already concluded from Eq. (4.12) and Fig. (4.15a). As for the collinear limit, it is clear by Fig. (4.17b) that it has an a negligible impact - the plots in Fig. 4.13 were already hinting at that. Lastly, the soft limit, when applied individually, seems to be accurate for the first 3 splittings, but the deviations become higher as we move to later splittings. When applied cumulatively in τ_4, it produces values consistently smaller than τ_2, while following its trend. This is expected, because if one compares the expressions for τ_3, whose curve would be very close to τ_2's, and for τ_4, one verifies that their ratio τ_4/τ_3 is exactly the hardest parton's transverse momentum fraction. Hence, we are looking at this momentum fraction's distribution when comparing the curves of τ_2 and τ_4 in Fig. 4.16b.



(a) (b)

Figure 4.16: a) Average formation time for the first 10 splittings, for the τ_i defined in Eqs. (4.5), (4.7) and (4.9). The approximations in these expressions were done sequentially. For instance, τ_4 in Eq (4.9) is the result of the low-virtuality, collinear and soft approximations. b) Ratio of the average formation times τ_2 and τ_4 with respect to the exact estimate τ_0.



(a) (b)

Figure 4.17: a) Average formation time for the first 10 splittings, for the τ_0 and τ_1 defined in Eqs. (4.5) and (4.6). The remaining $\overline{\tau}_i$ are defined in Eq. (4.13) The remaining $\overline{\tau}_i$ are defined in Eq. (4.13) where the respective approximations are done individually. For instance, $\overline{\tau}_3$ is obtained by taking the collinear limit $\cosh(y_1 - y_2) \approx 1 + (y_1 - y_2)^2$, $\cos(\phi_1 - \phi_2) = 1 - (\phi_1 - \phi_2)^2$ and $\cosh y_i \approx \cosh y$ in Eq. (4.5). b) Ratio of the average formation times with respect to the exact estimate τ_0.

4.4.2 Global Approach

Now consider that we calculate the parton formation time solely from final-state partons, instead of using the reclustering kinematics. Whatever the recombination scheme or the reclustering algorithm one uses to build the branching history, one can always represent each parton's 4-momentum as the recombination of all the subsequent partons' 4-momenta. Because we are using the E-scheme, this recombination is simply energy and momentum conservation. Hence, a parton's 4-momentum is given by the sum of the 4-momenta of a subset of all the final partons.

Let us ilustrate this simple reasoning by an example. Consider a history with 2 splittings (a shower with 3 final partons), in which the seed parton (parton 1) splits into parton 2, which in turn splits into final partons 4 and 5, and final parton 3 (Fig. 4.18).

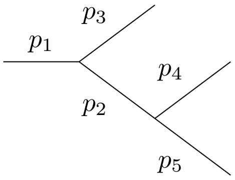


Figure 4.18: Branching history tree with 3 final partons. Each parton's 4-momentum is represented by p_i. The numbering follows no specific order.

If, for instance, one would want to calculate the first splitting's formation time, then one would need to find E_1/m_1^2.

$$m_1^2 = m_2^2 + m_3^2 + 2p_2 \cdot p_3 \tag{4.14}$$

If we use $m_2^2 = m_4^2 + m_5^2 + 2p_4 \cdot p_5$ and $p_2 = p_4 + p_5$, then

$$\tau_0 = \frac{E_3 + E_4 + E_5}{m_3^2 + m_4^2 + m_5^2 + 2(p_3 \cdot p_4 + p_3 \cdot p_5 + p_4 \cdot p_5)} \tag{4.15}$$

and, if one wants this expression written in terms of collider variables and masses, one can use $E_i = p_{t,i} \cosh y_i \sqrt{1 + m_i^2/p_{t,i}^2}$ and $p_i \cdot p_j = p_{t,i} p_{t,j} \left(\sqrt{1 + m_i^2/p_{t,i}^2} \sqrt{1 + m_j^2/p_{t,j}^2} \cosh(y_i - y_j) - \cos(\phi_i - \phi_j) \right)$, just like one did in the previous section. If one were to calculate p_2's formation time, only p_4 and p_5 would matter, as expected. The generalization of Eq. (4.15) for an arbitrary number of final partons and for a given internal parton of the branching history is then:

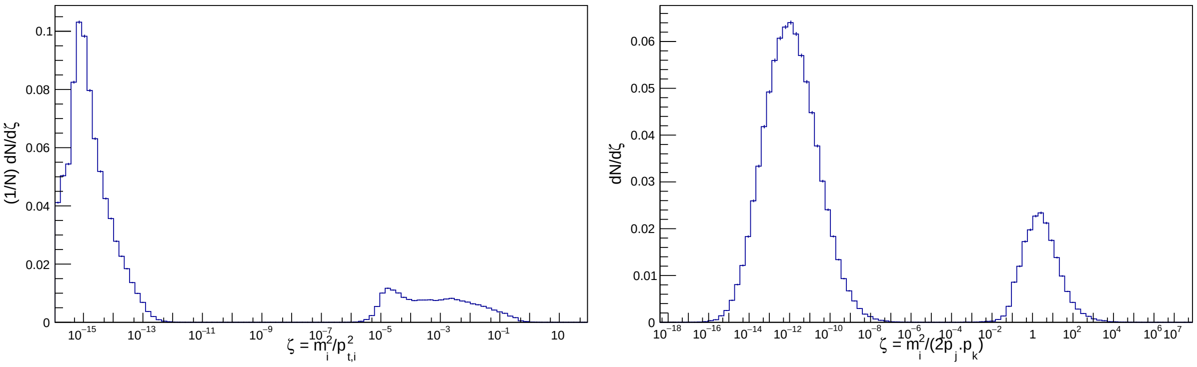
$$\tau_0 = \frac{E}{m^2} = \frac{\sum_i E_i}{\sum_i \sum_j p_i \cdot p_j} = \frac{\sum_i p_{t,i} \sqrt{1 + m_i^2/p_{t,i}^2} \cosh y_i}{\sum_i \sum_j p_{t,i} p_{t,j} \left(\sqrt{1 + m_i^2/p_{t,i}^2} \sqrt{1 + m_j^2/p_{t,j}^2} \cosh(y_i - y_j) - \cos(\phi_i - \phi_j) \right)} \tag{4.16}$$

where the i and j indices run over a subset of all the final partons - those belonging to branches which go through the parton we are considering. Consequently, the formation time of a given parton does not depend on its subsequent branching history but simply on what final partons emerge from it. This confirms analytically that the first splitting's formation time is totally independent of the reclustering algorithm.

The interesting thing about this approach is that if, eventually, one would want to study a fully hadronized jet (hadron-level events), each formation time would be given by measurable quantities of the final detected hadrons, provided that their masses could be identified. Nevertheless, if one wants to work at parton-level, one would only need to assume the final partons to be massless in order for the formation time to be fully expressed in terms of collider variables. This means that, even if we ignore the final partons' masses, we are still considering a non-null contribution from each "internal" parton's virtuality. The result of applying the massless limit to Eq. (4.16) is then:

$$\tau_2^g \equiv \frac{\sum_i p_{t,i} \cosh y_i}{\sum_i \sum_{j \neq i} p_{t,i} p_{t,j} \left(\cosh(y_i - y_j) - \cos(\phi_i - \phi_j) \right)} \tag{4.17}$$

where the superscript g stands for global approximation. By neglecting each final parton's mass, we would still be ignoring the same type of terms $m_i^2/(2p_j \cdot p_k)$ and $m_i^2/p_{t,i}^2$ but, presumably, these are much smaller than for "internal" partons, whose virtuality is non-zero. The dot product terms to consider for the first ratio are dependent on what formation time we are calculating - it can be a single term or all the possible terms coming from every final parton pair, as in the calculation of the first splitting's formation time. Although Fig. 4.19a confirms that $m_i^2/p_{t,i}^2$ is, in the majority of cases, negligible, one can not say the same about ignoring m_i^2 relative to the relevant $p_j \cdot p_k$. In Fig. 4.19b, we plot, for each final parton i, the largest possible ratio $m_i^2/(2p_j \cdot p_k)$ by picking the lowest possible $p_j \cdot p_k$. There is a non-negligible percentage of events with ratios above $\mathcal{O}(1)$, indicating that the approximation should be used with care. Additionally, it is now clear that the peaks at lower values one observed in Fig. 4.12a are mostly due to final-state partons.



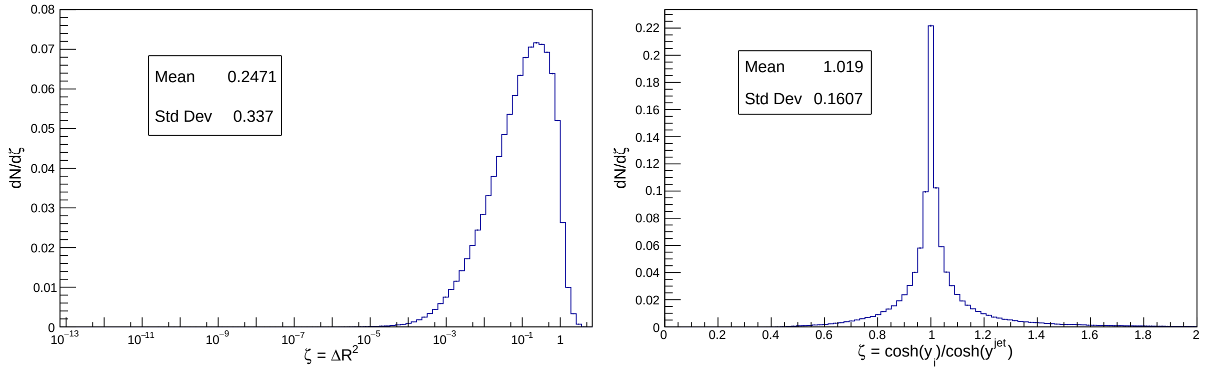
(a) (b)

Figure 4.19: a) Factor controlling the first order corrections of the square-root term in Eq. (4.16), plotted for each final parton. b) Factor controlling the smallness of the m_i^2 terms in Eq. (4.16). Plotted for each parton and for the minimum $p_j \cdot p_k$ possible from all possible jk pairs. Both plots are normalized to unit integrals.

One can also apply the collinear and soft approximations to Eq. (4.17). The collinear approximation is applied by saying every possible pair of the relevant subset of final partons, not only the ones clustered together, has a small ΔR_{ij}^2:

$$\tau_3^g \equiv \frac{2\sum_i p_{t,i} \cosh y^{jet}}{\sum_i \sum_{j \neq i} p_{t,i} p_{t,j} \Delta R_{ij}^2} \tag{4.18}$$

where we took the zeroth order approximation $\cosh y_i \approx \cosh y^{jet}$. For the azimuth-rapidity distance in Fig. 4.20a, the distribution peaks at values $\mathcal{O}(10^{-1})$ and the percentage of pairs with $\Delta R_{ij}^2 > 1$ is much higher than in Fig. 4.13a. The corrections to $\cosh y/\cosh y^{jet} \approx 1$, which come from the finite difference in rapidity between each final parton and the jet, are shown in Fig. 4.20b. Evidently, this approximation is not as accurate as taking each parton's rapidity to be the same as its parent's. Nonetheless, the approximation results in a multiplicative factor of ~ 0.5 or 1.5, at most.



(a) (b)

Figure 4.20: a) Angular distance ΔR_{ij}^2 distribution for each possible pair of final partons. b) Factor $\cosh y_i/\cosh y^{jet}$ for each final parton. Both plots are normalized to unit integrals.

One can further approximate this calculation by saying there is a final parton which is much more energetic than all the other partons in the subset. This approximation would be suitable for jets with a single hard branch from which soft and collinear radiation would emitted. In this limit, we get

$$\tau_4^g \equiv \frac{2p_{t,h}\cosh y^{jet}}{\sum_i \sum_{j\neq i} p_{t,i}p_{t,j}\Delta R_{ij}^2} \tag{4.19}$$

where h is the index of the hardest parton. The validity of this approximation is dependent on what parton we wish to compute the formation time of. To do so, one needs to know the p_t fraction of the hardest parton relative to the remaining partons in the relevant subset. For simplicity, in Fig. 4.21, we show the ratio between the p_t of the second hardest and the p_t of the hardest parton. The resulting distribution seems to hint that the jet configuration where this limit applies is not that frequent. If this was the case, we would see a higher concentration towards lower values of the ratio.

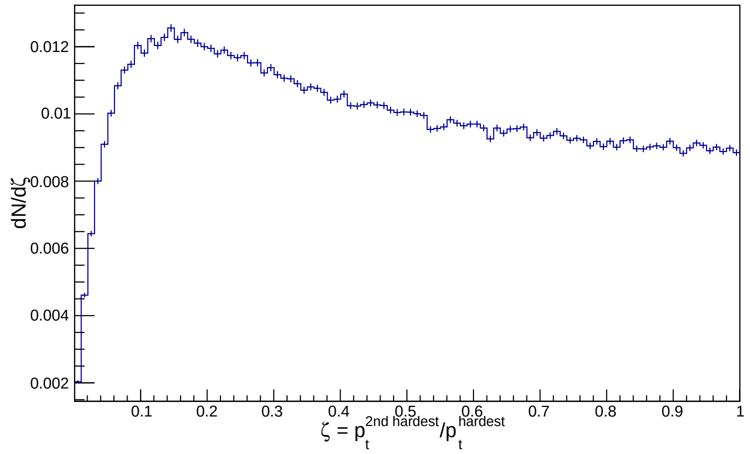


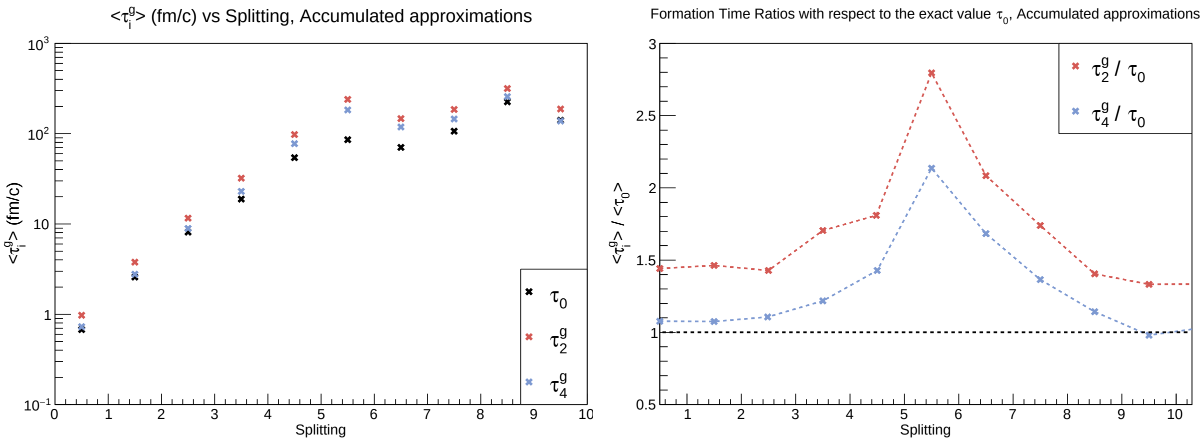
Figure 4.21: Ratio of the transverse momentum of the second hardest parton to that of the hardest parton in a jet. Normalized to unit integral.

Just like for the local approach, we now show the impact of the approximations described above when applied cumulatively and individually. In Fig. 4.22, we show the formation time distributions calculated with two of the approximate estimates, in comparison with the exact estimate τ_0. We present the estimates for low-virtuality (τ_2^g, Eq. (4.17)) and the estimate for the expression where all limits are applied (τ_4^g, Eq. (4.19)). In Fig. 4.23, we show the same comparison but for formation time estimates with only a single approximation (τ_1^g and the $\overline{\tau}_i^g$'s). All plots are shown without error bars for the sake of readability. The global approximations applied individually result in

$$
\begin{aligned}
\overline{\tau}_1^g &= \tau_1^g = \frac{E}{m^2 - \sum_i m_i^2} \\[2ex]
\overline{\tau}_2^g &= \frac{\sum_i p_{t,i}\cosh y_i}{\sum_i m_i^2 + \sum_i \sum_{j\neq i} p_{t,i}p_{t,j}\left(\cosh(y_i - y_j) - \cos(\phi_i - \phi_j)\right)} \\[2ex]
\overline{\tau}_3^g &= \frac{\sum_i p_{t,i}\sqrt{1 + m_i^2/p_{t,i}^2}\cosh y^{jet}}{\sum_i \sum_j p_{t,i}p_{t,j}\left(\sqrt{1 + m_i^2/p_{t,i}^2}\sqrt{1 + m_j^2/p_{t,j}^2}(1 + \Delta y_{ij}^2) - (1 - \Delta\phi_{ij}^2)\right)} \\[2ex]
\overline{\tau}_4^g &= \frac{p_{t,h}\sqrt{1 + m_h^2/p_{t,h}^2}\cosh y_h}{m^2}
\end{aligned}
\tag{4.20}
$$

where $\Delta y_{ij} = y_i - y_j$, $\Delta\phi_{ij} = \phi_i - \phi_j$, h is the hardest parton index and, as in the local approach, we assumed in $\overline{\tau}_4^g$ that the $m_i^2/p_{t,i}^2$ are of the same order. The estimates' numbers match those of the cumulative approximations in Eqs. (4.17), (4.18) and (4.19). Estimate τ_1^g is presented in the first line of Eq. (4.20).

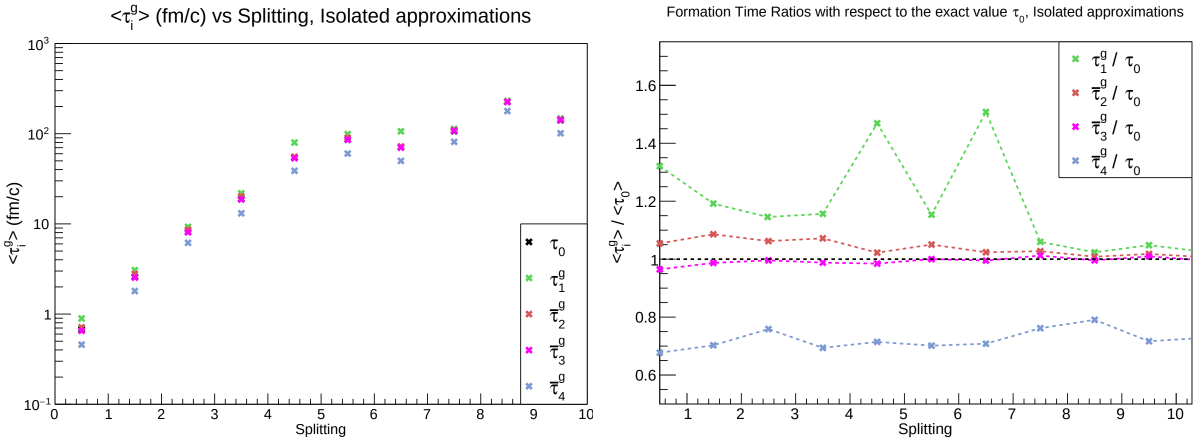
First of all, ignoring all terms containing masses is still not a reasonable approximation (τ_2^g in Fig. 4.22b). Solely ignoring $m_i^2/p_{t,i}^2$ implies only minor deviations ($\overline{\tau}_2^g$ in Fig. 4.23b), smaller than what we observed for the local approach ($\overline{\tau}_2$ in Fig. 4.17b)- this was expected by looking at Fig. 4.19a and 4.12a. Neglecting m_i^2 with respect to $p_{j}\cdot k$ is the most impactful approximation, as was expected from the previous analysis. The collinear limit estimate presents a slight deviation in the first splitting which we didn't see in Fig. 4.17b - it only makes sense, since we're saying every final parton has rapidity equal to the jet's and the azimuth-rapidity distances include contributions from the most distanced final partons in the jet. Nevertheless, applying collinearity still has almost no impact. Finally, the soft limit, applied individually, presents an underestimate of τ_0 by an average factor of ~ 0.7 for the first 10 splittings. In the cumulative plot, we see the same behaviour as for the local approach - the differences between the two curves τ_2^g and τ_4^g in each splitting are essentially given by the distribution of the energy fraction of the hardest parton with respect to the jet.



(a)

(b)

Figure 4.22: a) Average formation time for the first 10 splittings, for the τ_i^g defined in Eqs. (4.5), (4.17) and (4.19). The approximations in these expressions were done sequentially. For instance, τ_4^g in Eq (4.9) is the result of the low-virtuality, collinear and soft approximations. b) Ratio of the average formation times τ_2^g and τ_4^g with respect to the exact estimate τ_0.



(a)

(b)

Figure 4.23: a) Average formation time for the first 10 splittings. The τ_0 is defined in Eq. (4.5) and the remaining are defined in Eq. (4.20). The $\overline{\tau}_i^g$'s are the formation time estimates one obtains by doing the respective approximation directly in τ_0. For instance, $\overline{\tau}_3^g$ is obtained by taking the collinear limit $\cosh(y_i - y_j) \approx 1 + (y_i - y_j)^2$, $\cos(\phi_i - \phi_j) = 1 - (\phi_i - \phi_j)^2$ and $\cosh y_i \approx \cosh y^{jet}$ in Eq. (4.5). b) Ratio of the average formation times with respect to the exact estimate τ_0.

4.5 Space-Time Structure Analysis

The assignment of a space-time structure to a jet allows one to explore how splittings are distributed in time, prompting basic questions such as "what are the orders of magnitude of the timescales involved in a jet's development?", "are the splittings of each branch well separated in time?" or "are there defined time orderings along the tree?". We can also ask how we can leverage a jet's space-time structure so that we can be sensitive to the time scales involved in the evolution of the QGP. This section is an attempt at answering such questions.

4.5.1 Formation Time Orderings

Formation Times Ordering along each Branch

In Section 2.5, one mentioned that a branching history calculated with the τ algorithm, i.e., the generalized k_t algorithm with $p = 1/2$, was expected to be ordered in formation times. The motivation for this assertion is that, as we saw in Section 4.4 with Eq. (4.9), in the soft, massless and collinear limit the parton formation time is proportional to the inverse of the distance metric of the algorithm:

$$\tau_4 = \frac{\cosh y}{\min(p_{t,i}, p_{t,j})\Delta R_{ij}^2} = \frac{\cosh y}{d_{ij}^{\tau}} \tag{4.21}$$

In the reclustering algorithm each step has an associated value of the distance metric, which increases from step to step until one gets to the final recombination, the seed splitting, which will have the largest d_{ij}^{τ}. An immediate implication of this is that the splittings along each branch are ordered according to d_{ij}^{τ}, i.e., it decreases when we go from the seed splitting to the end of the branch. Hence, one would expect that the quantity τ_4 would increase when travelling along each branch in that direction - a formation time ordering. It is not only interesting to find out if the $\cosh y$ spoils this ordering, but it is also worth verifying whether the remaining formation time estimates, including the exact τ_0, are ordered along each branch or not.

In Fig. 4.25, we present the ratios of the formation times τ_k^{i+1}/τ_k^i of consecutive partons in a branch, starting at the seed splitting. An example of the plotted quantities is shown in Fig. 4.24.

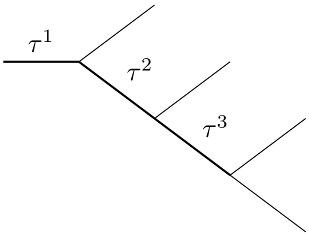
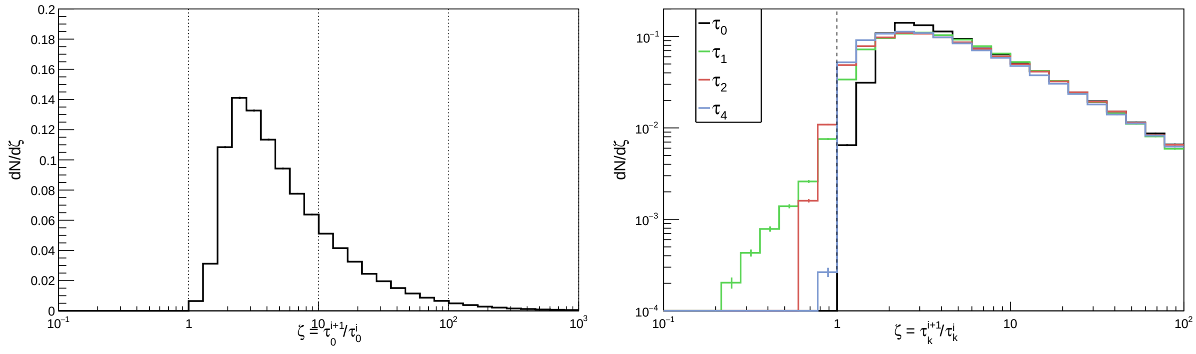


Figure 4.24: Branching history tree with 4 final partons and a single branch. Each internal parton (thicker lines) contributes with a formation time. The numbering follows no specific order. In this case, one would plot ratios τ^3/τ^2 and τ^2/τ^1. For trees with more than a single branch, this is done for each branch.

The aim of the plots is to see if each formation time estimate τ_k monotonically increases when travelling through a branch and we present the results for estimates τ_0, τ_1, τ_2 and τ_4, respectively in Eqs. (4.5), (4.6), (4.7) and (4.9). In Fig. 4.25a we see the consecutive splittings time ratios for the exact estimate $\tau_0 = E/m^2$ and we conclude that it never decreases from splitting to splitting. As for the remaining estimates in Fig. 4.25b, the one which most often violates the ordering is τ_1 (the green curve), for which the terms $m_k^2/2p_i \cdot p_j$ were ignored. For this estimate, around 1% of the splittings have a formation time smaller than their parents'. For the estimate τ_4, which accumulates every approximation, there are very few cases where the ordering is spoiled by the term $\cosh y$, i.e., the rapidity decreases in such a way from one splitting to the next that it compensates the decrease in the value of the distance metric (the increase in $1/d^{\tau}_{ij}$). For τ_4, about 0.02% of the splittings violate the branch ordering.

Thus, one can conclude that a tree calculated with the τ algorithm will be exactly ordered, in each branch, in $\tau_0 = E/m^2$ and, for all practical purposes, it can be said the same about an ordering in τ_4. For the remaining formation time estimates, the ordering is true on average and most often, only with a small percentage of violations. Additionally, the formation times are logarithmically separated inside each branch, with an average multiplicative factor of order 10 separating them.



(a) (b)

Figure 4.25: a) Ratio of the formation times, calculated with τ_0, of consecutive partons i and $i+1$ in each branch. b) Same plot with logarithmic y-axis with superimposed curves for τ_0, τ_1, τ_2 and τ_4 as given by Eqs. (4.5), (4.6), (4.7) and (4.9). All plots are normalized to unit integrals and resulted from data of all generated events.

Formation Times Ordering along Reclustering Steps

As detailed at the end of Section 2.5, splitting orders can serve to list emissions in the direction of monotonical evolution of a given quantity. The reclustering algorithm provides us with a splitting order which is exactly the inverse of the sequence of steps it took. In Fig. 4.27, one tests if each of the formation time estimates τ_k always decreases when going from reclustering step i to $i-1$ (following the exact reclustering sequence), so that one would have $\tau_k^i/\tau_k^{i-1} > 1$. The last step, the seed splitting, corresponds to $i = 1$ and the first step to $i = N$. For instance, for the tree presented in the following Fig. (4.26), one plots the values of τ^3/τ^2 and τ^2/τ^1.

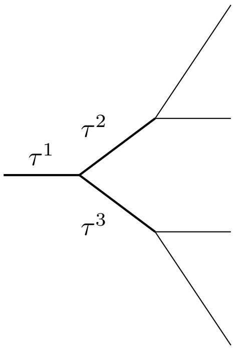
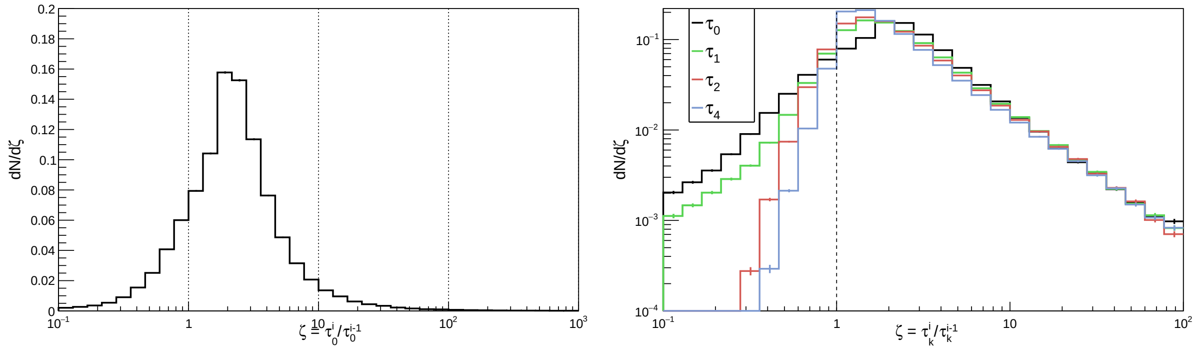


Figure 4.26: Branching history tree with 4 final partons and two branches. Each internal parton (thicker lines) contributes with a formation time. The numbering follows the specific sequence of reclustering, i.e., τ_3 is the first reclustering step, τ_2 is the second and so on.

If the ordering in τ_0 was exact when travelling along a branch, the same cannot be said when analysing formation time ratios along the reclustering sequence, as can be seen in Fig. 4.27a. Such happens because, most often, the sequence is not done full branch by full branch but rather by alternating between branches. In fact, by observing Fig. 4.27b, we conclude that the exact estimate (the black curve) is the one for which consecutive splittings are more likely to violate the ordering - about 17% of the distribution below 1. The remaining τ_k seem to spoil the reclustering order less as one accumulates more approximations. Specifically, going from τ_1 to τ_2, i.e., ignoring terms $m_i^2/p_{t,i}^2$, seems to have the greatest impact. The formation time which, *a priori*, would be the one for which the ordering would be most valid - $\tau_4 = \cosh y/d_{ij}^{\tau}$, the blue curve - is exactly the one with the smallest percentage of ratios below unity ($\sim 6\%$).

Hence, on average and most often, one can say that the splitting order dictated by the τ reclustering produces a sequence of increasing formation times. The evolution in such direction is most likely when applying the massless, collinear and soft limit to the formation times, i.e., when calculating a tree's formation times with τ_4. Note that the spoiling of the formation time ordering along the reclustering sequence has to be a consequence of its branch alternation - successive reclustering steps belonging to the same branch follow the ratios in Fig. 4.25. This is in agreement with usual parton shower generators where, although each branching is ordered according to a given quantity, distinct branches have their own logarithmic separation between consecutive splittings. In terms of formation times and looking at Fig. 4.26, although one has strong ordering branch-wise $\tau^3, \tau^2 \gg \tau^1$, one doesn't necessarily have $\tau^3 \gg \tau^2$.

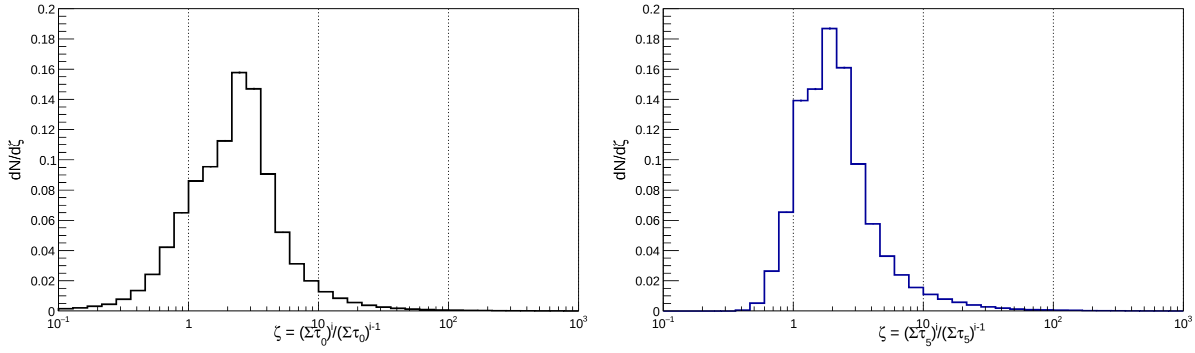


(a) (b)

Figure 4.27: a) Ratio of the formation times, calculated with τ_0, of consecutive reclustering steps i and $i-1$. b) Same plot with logarithmic y-axis with superimposed curves for τ_0, τ_1, τ_2 and τ_4, as as given by Eqs. (4.5), (4.6), (4.7) and (4.9). All plots are normalized to unit integrals and resulted from data of all generated events.

Absolute Times Ordering along Reclustering Steps

A question one could then ask is if the splitting order resulting from calculating absolute times ($\Sigma\tau$) is exactly the same as the one resulting from the reclustering procedure. The answer boils down to evaluating the ratio of absolute times for consecutive reclustering steps - the same reasoning for the plots in Fig. 4.27, but with absolute times rather than with formation times. In Fig. 4.28a, the distribution of these ratios for τ_0 is presented and it is clear that the answer is negative - there are plenty of cases, about 17%, where absolute times decrease from one merging step to the previous. One can say the splitting orders are, on average and most often (the peak is at $(\Sigma\tau_0)^i/(\Sigma\tau_0)^{i-1} \sim 2.5$), the same, but there is a significant probability of it not being true for at least one splitting of a jet.



(a)

(b)

Figure 4.28: a) Ratio of the absolute times, calculated with τ_0, of consecutive reclustering steps i and $i-1$. b) Ratio of the absolute times, calculated with $\tau_5 = 1/d_\tau$, of consecutive reclustering steps i and $i-1$. All plots are normalized to unit integrals and resulted from data of all generated events.

The reason for this high number of ordering violations in absolute time can be traced back, once more, to the fact that the reclustering sequence forms branches simultaneously, not one at a time. In fact, the equality between the splitting order coming from the sum of consecutive formation times and the inverse reclustering sequence is not even exact for formation times which are precisely given by the inverse of the distance metric, i.e., $\tau_5 = 1/d^\tau$. This is proven by Fig. 4.28b, where one plots the ratios of absolute times for this time estimate - around 10% of the distribution is below unity. One can see how such violations may arise by considering the simple example in Fig. (4.29), a tree with 5 partons and, therefore, 4 reclustering steps.

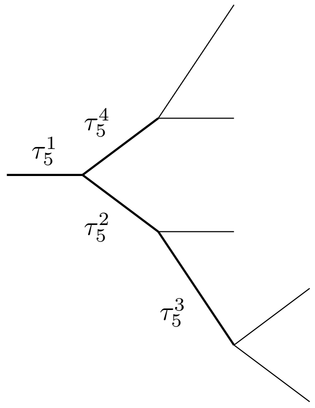


Figure 4.29: Branching history tree with 5 final partons and two branches. Each internal parton (thicker lines) has a corresponding inter-particle distance d_i^τ and contributes with a formation time $\tau_5^i = 1/d_i^\tau$. The numbering follows the specific sequence of reclustering, e.g., τ_5^4 is the first reclustering step's formation time.

With this sequence of steps one has

$$d_1^\tau > d_2^\tau > d_3^\tau > d_4^\tau \implies \tau_5^1 < \tau_5^2 < \tau_5^3 < \tau_5^4 \tag{4.22}$$

but the question is if this order is the same when considering absolute times. For this tree, the absolute times are given by

$$\begin{aligned}(\Sigma\tau_5)^2 &= \tau_5^2 + \tau_5^1\\ (\Sigma\tau_5)^3 &= \tau_5^3 + \tau_5^2 + \tau_5^1\\ (\Sigma\tau_5)^4 &= \tau_5^4 + \tau_5^1\end{aligned} \tag{4.23}$$

and $(\Sigma\tau_5)^1 = \tau_5^1$. If the splitting order imposed by absolute times were equal to the one imposed by the reclustering sequence, we would have:

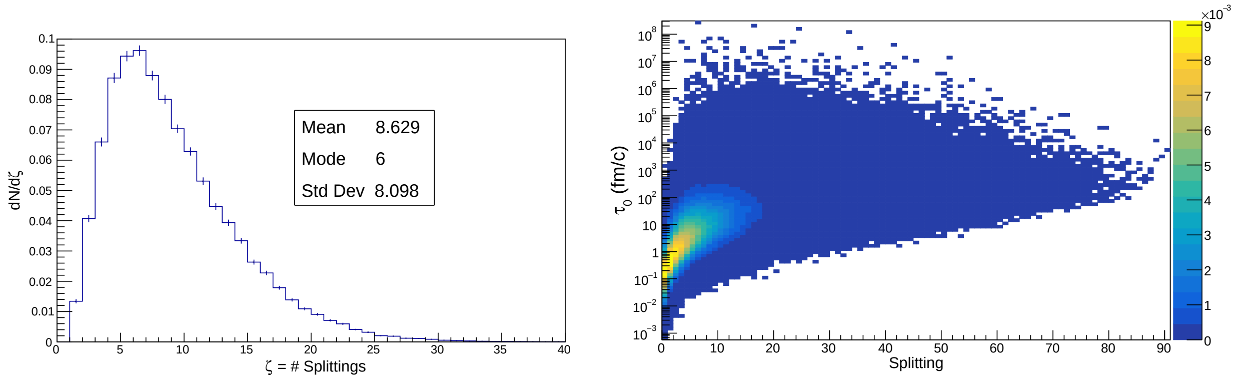
$$(\Sigma\tau_5)^1 < (\Sigma\tau_5)^2 < (\Sigma\tau_5)^3 < (\Sigma\tau_5)^4 \tag{4.24}$$

However, a problem arises in the last inequality - it doesn't hold if $\tau_5^2 + \tau_5^3 > \tau_5^4$. This setup is possible in a reclustering sequence, even if $\tau_5^4 > \tau_5^2, \tau_5^3$. This explains why, even for formation times calculated with the inverse of the metric, one does not necessarily obtain an equivalence between the inverse reclustering sequence and the absolute time splitting order (Fig. 4.28b).

4.5.2 Formation Time Distribution

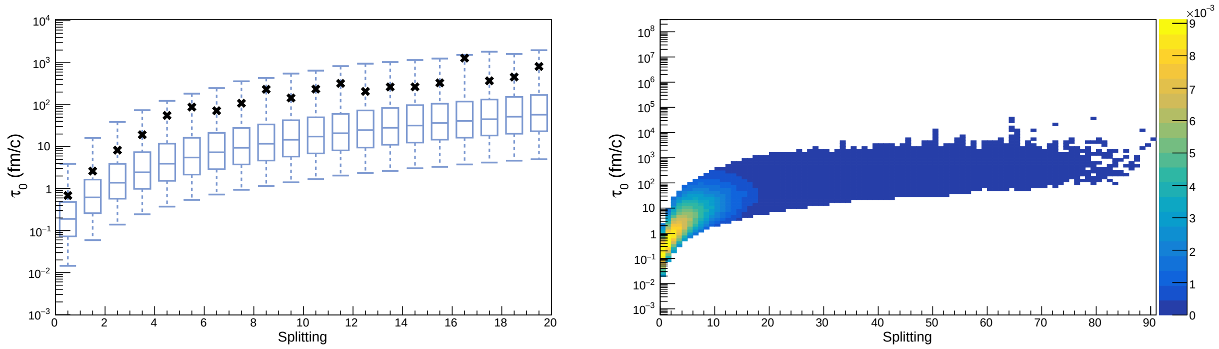
In this section, we investigate how formation times are distributed along the splittings of a jet. The space-time structure one obtains from both splitting orders takes into account different branches, giving one the ability to know how the jet's internal structure is distributed over time, even for jets with multiple branches. With this information, one can tell, for instance, the average number of partons at a given moment in time. This results in different "representations" of the same jet at different frames of its temporal development. The representations in absolute time are particularly useful when studying the QGP. For instance, the fact that the nth splitting of a jet occurs at a given absolute time below the medium's lifetime (~ 10 fm/c) implies that, at that point in time, the jet behaves as a system of n colored charges. Whether the plasma actually resolves each parton as an independent color source or if it interacts with the jet through a number of *effective emitters* [57] is a question to be answered by studying its resolution length ([57], [58]).

In the scattering plot of Fig. 4.30b, we present the formation time of each splitting, ordering them according to absolute time. We face ourselves with a startling result: formation times inside a jet span approximately 12 orders of magnitude in fm/c. This is reflected on very wide tails for the formation time distributions of each splitting, with the upper tails being the widest. Nonetheless, peak formation times are well defined at least for the first ~ 20 splittings, where the jet multiplicity spectrum is relevant (Fig. 4.30a).



(a) (b)

Figure 4.30: a) Scattering plot for the formation time of each splitting according to their order in absolute time. b) Jet particle multiplicity spectrum. All results are parton-level and the multiplicity spectrum is normalized to unit integral.

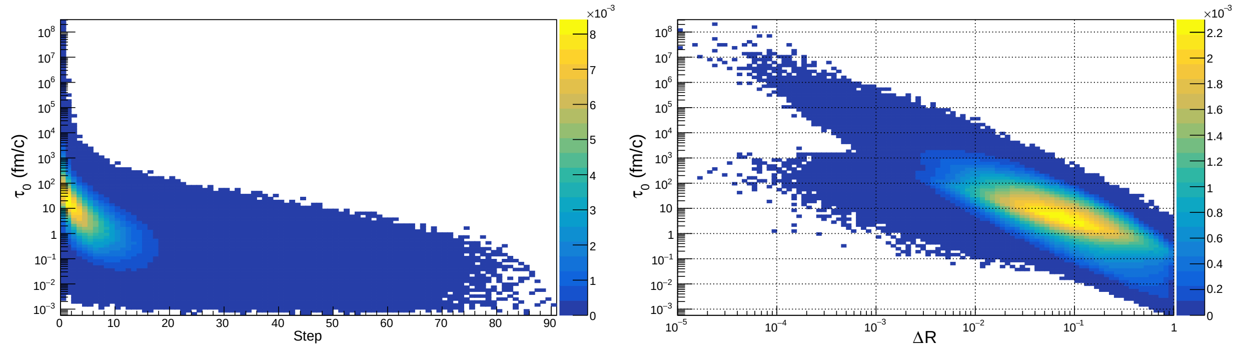


(a) (b)

Figure 4.31: a) Plot of the relevant quantiles (blue lines) and the mean values (black markers) for each each bin. The inner box represents the 25%, 50% (the median formation time) and 75% quantiles. The lower and upper bounds represent the 5% and the 95% quantiles, respectively. This is plotted for the first 20 splittings. b) Scattering plot for the formation time of each splitting according to their order in absolute time. Only the portion of each bin's distribution between the corresponding 5% and 95% quantiles is ploted. All results are parton-level.

One can clearly see by Fig. 4.30b that, although the wide upper tails exist, most of the distribution is contained in a smaller span of orders of magnitude. This is most clearly represented in Fig. 4.31a where each splitting's inter-quantile range is shown (inner blue box), along with its mean value (black marker) and the quantiles corresponding to 5% and 95% probability - the lower and upper blue ticks, respectively. The inter-quantile range is very small in comparison with the total span of values one sees in Fig. 4.30b. In fact, the upper tail's influence is reflected on the large differences between median and mean values. Additionally, the median formation times of each splitting follow a distinct trend, while the mean formation times are heavily influenced by the upper tails and, except for the first 4/5 splittings, do not follow a clear trend. In Fig. 4.31b, one shows the distribution in Fig. 4.30b with cuts at the 5% and 95%, the tails below and above these quantiles being discarded. This shows, in a more plain manner, that the bulk of the distribution is within fewer orders of magnitude. What remains of the dispersion is probably caused by the inclusive nature of the plot - it contains jets with all possible particle multiplicities and kinematic configurations.

A more interesting investigation to carry on would be to pinpoint the cause of such a wide distribution of formation times. In fact, the span of multiple orders of magnitude is not so puzzling if we examine the dependences on $\tau_0 = E/m^2$ for internal partons of the branching history - the transverse momentum ($p_t \approx E/\cosh y$) covers about 3 orders of magnitudes, from the seed parton's $\sim 10^2$ GeV to the final partons' $\sim 10^{-1}$ GeV, and the virtualities range from about $\sim 10^2$ GeV to almost real partons ($m \approx 0$ GeV). These partons with negligible virtuality result from recombining a pair of massless and almost collinear final partons, thus resulting in disparately large formation times (this was made clear in Section 4.4). This assertion is confirmed in Fig. 4.32a, where one shows the evolution of formation times according to the sequence of reclustering steps. In this plot, the first bin corresponds to the first reclustering step. The central conclusion is that the first couple of reclustering steps, where one expects ΔR to be the smallest, are the main source of large, outlier formation times. This is supported by the right hand side plot in Fig. 4.32b, where one can clearly sees that formation times above 10^4 are only found for partons separated by $\Delta R < 10^{-2}$, which represents about $\sim 0.01\%$ of the distribution.



(a)

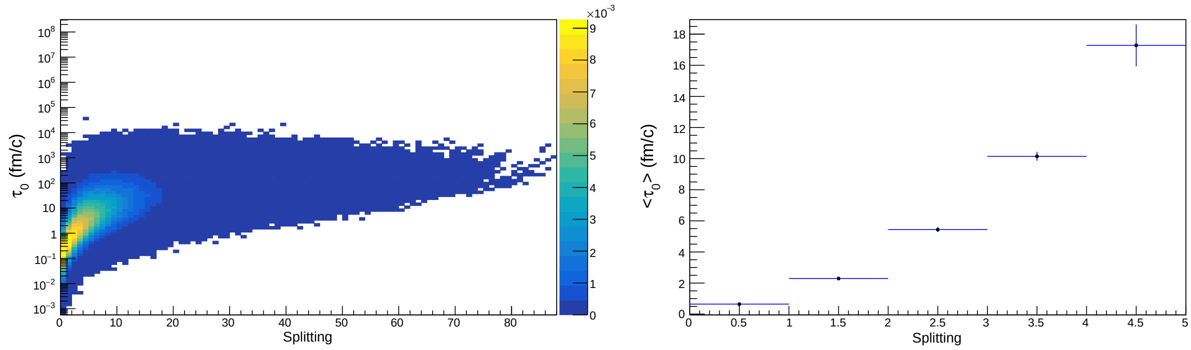
(b)

Figure 4.32: a) Scattering plot for the formation time of each splitting according to the reclustering sequence b) Scattering plot of the formation time of each splitting and corresponding ΔR of the subsequent partons. All results are parton-level.

In practice, for QGP studies, any intra-jet activity occuring after 10 fm/c does not influence the plasma-jet interaction. Not only this, but very large formation times, which correlate with small azimuth-rapidity distances ΔR (Fig. 4.32b), might not be detectable. As mentioned in Section 2.4, calorimeters have a finite angular resolution - two almost collinear particles are detected as a single energy deposition. In Fig. 4.33a, one shows the same plot as in Fig. 4.30b, but ignoring the formation time of partons that split into two collinear partons which wouldn't be resolved by a calorimeter. We take this resolution cut to be $\Delta R < 10^{-2}$, as an estimate of an experimental cut. Evidently, one should rather be talking about collinear hadrons, since these are the detected particles. However, following the LPHD idea mentioned in Chapter 3, two collinear partons will, in principle, originate collinear hadrons. Moreover, for a splitting which doesn't pass this cut, its formation time is only ignored if both of its resulting partons are either final partons or have not passed the distance cut themselves. This strategy would not be needed were the branching history calculated with C/A, where each branch would be strictly ordered in ΔR.

The main conclusion is that most of the dispersion coming from the upper tail of each distribution wouldn't be resolved by usual calorimeters. In Fig. 4.33b, the mean formation times for each splitting are plotted along with their respective standard errors of the mean. They are shown to follow an approximate exponential trend, being well separated between consecutive splittings. The first splitting has an average formation time of, approximately, 0.65 fm/c with a weighted standard error of the mean of about 0.017 fm/c. One should also note that there is only about 13% of the jets with at least one ΔR cut and, for such jets, on average, a single cut is made.

It may seem, at first sight, that the ΔR cut generates a problem. In fact, this cut is also neglecting splittings which would happen inside the medium, as can be seen in Fig. 4.32b. These cases occur when a parton splits into two collinear parton which have non-negligible virtualities, something which is made explicit by Eq. (4.5) of Section 4.4. However, even if this parton is splitting at an absolute time below 10 fm/c, the subsequent collinear partons wouldn't be resolved as separate color charges because the medium has a mimimum resolution length ([57], [58]). Hence, a cut based on the detector's finite resolution does not neglect meaningful information about the jet-medium interaction.



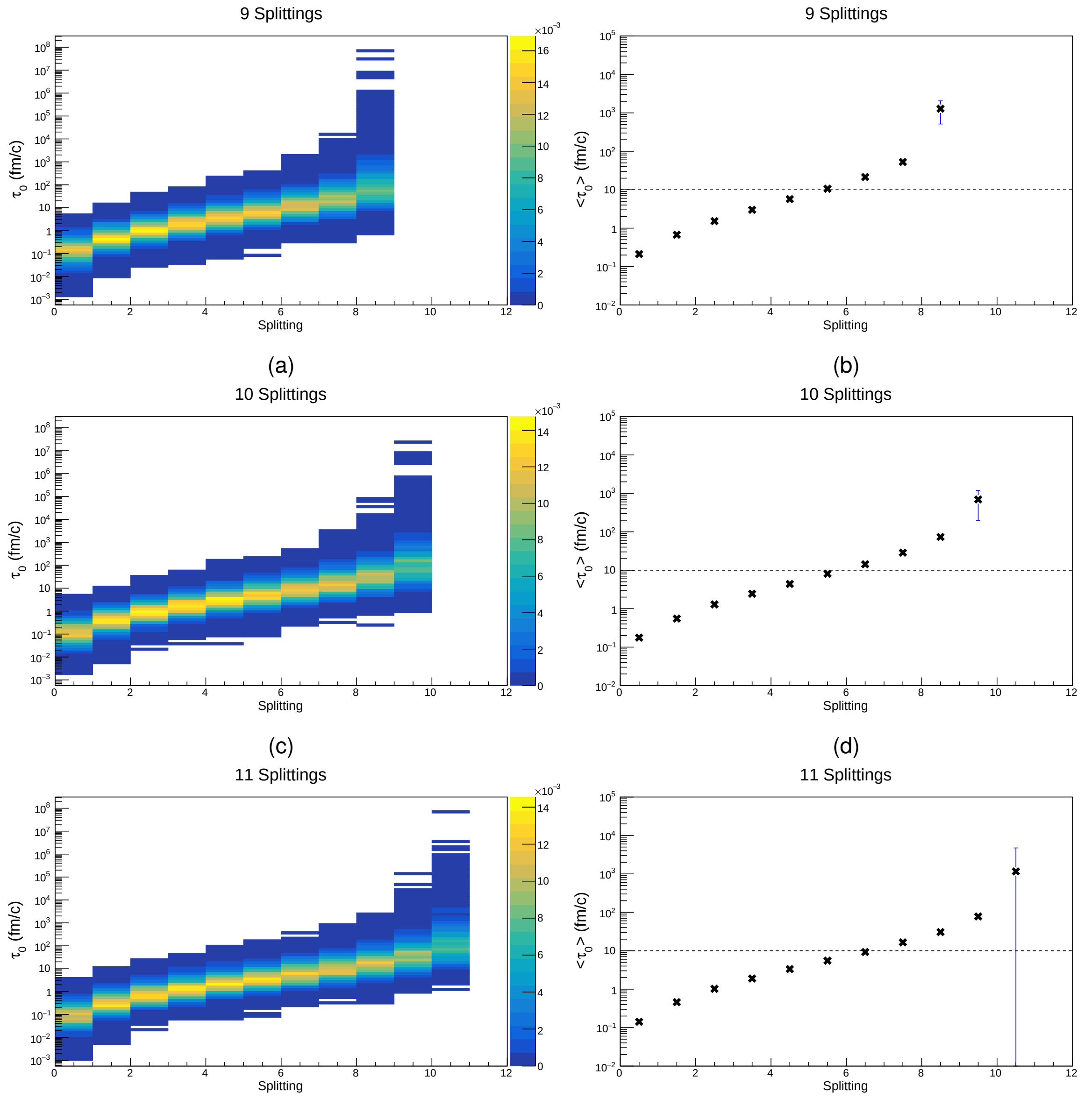
(a) (b)

Figure 4.33: a) Scattering plot for the formation time of each splitting according to their absolute time and only for splittings which result in partons separated by, at least, $\Delta R = 10^{-2}$. b) Average of formation times for each splitting, with errors given by the standard error of the mean (SEM), for the first 5 splittings of the history. The SEMs are almost unperceptible, only being relevant for the last two splittings on the plot. All results are parton-level.

Although the reclustering and absolute time splitting orders are not fully equivalent, as was shown previously in Section 4.5.1, one can say that the splittings corresponding to the first reclustering steps are generally the ones with the largest absolute times. This is expected since the τ algorithm produces a branching history which is ordered in increasing formation time - also shown in Section 4.5.1. Not only that, but the first reclustering step, in about 90% of the events, corresponds to the splitting with the largest absolute time - hereinafter refered to as the last splitting of the jet. In cases where this is not strictly true, the corresponding absolute times are, on average, of the same order.

If one fixes the jet multiplicity (Fig. 4.34), which is equivalent to fixing the number of splittings, one observes that the later splittings of each jet are exactly the ones causing most of the dispersion. One can conclude, by extrapolation, that the long upper tail of the nth splitting's formation time distribution is mainly due to the last splitting of a jet with n splittings. A most likely explanation for this is that it is an artifact introduced by the phase space sampling in the parton shower generation, which possibly renders the last emissions in each branch unphysical. In spite of the decrease in the range of formation times for each splitting, there is still a span of about 3 or 4 orders of magnitude which, presumably, is due to the many possible kinematic configurations a set of n final partons can have. This is a consequence of the way each set was obtained - through parton shower generation, where energy fractions and opening angles (or any LL equivalent variable) are generated through a probabilistic implementation. Hence, calculations which depend on parton formation times will always have this associated dispersion. Because there are the splittings whose distributions have relevant tails at values of order $10 fm/c$, the dispersion is important, in particular, for simple energy loss models depending on the path length travelled by each parton (e.g. [12] and [59]). This is motivation for future work involving the search for quantities which correlate with a significant decrease in the overall kinematic dispersion of a jet's formation time evolution.

A very important observation to make from the plots in Fig. 4.34 is how the whole line of average formation times decreases when one increases the jet particle multiplicity. This conclusion is of special relevance for the exploration of the timescales involved in jet-quenching, which can then hint on the space-time development of the QGP. The reason for this is that, because the medium has a lifetime of about 10 fm/c, jets which are more sensitive to its time evolution are the ones which have a greater portion of their activity happening at earlier time [12]. Hence, jets with higher multiplicities seem to be those with a higher sensitivity to shorter timescales. This fact is in agreement with an earlier conclusion of Section 4.2.2 - jets with greater mass have shorter first splitting formation times.



(a) (b) (c) (d) (e) (f)

Figure 4.34: Left column: Scattering plot for the formation time of each splitting according to their absolute time. Right column: Average of formation time for each splitting, with errors given by the standard error of the mean (SEM). The jet multiplicity is fixed in each plot at 9, 10 and 11 splittings. Results are shown for the first 12 splittings and are all parton-level.

Chapter 5

Conclusions

5.1 Main Achievements and Future Work

In this work, we set out to understand the relevant timescales involved in the space-time evolution of a jet. The result was a detailed analysis of a number of interesting features of that evolution, which can be reproduced by generating events in the working conditions described in Chapter 3. It is perhaps useful to emphasize a few key points made throughout the thesis, while suggesting further work directions.

First, one understood that there must be a compromise between efficiently capturing most of the FSR and avoiding to reconstruct ISR in a jet. Consequently, the jet reconstruction radius should not be increased recklessly. This is particularly important because the formation times are sensitive to the presence of this additional, mostly soft radiation. Not only this, but working at hadron-level rather than parton-level also has an impact on the distribution of each splitting's formation time. In essence, both ISR and hadronization cause an overall decrease in each splitting's mean formation time. For the first splitting, for which a more detailed analysis was made, the whole distribution slightly shifts to lower formation times. A deeper understanding of the interplay between their influence on a jet's kinematics and the change in a jet's space-time structure would be of relevance in future work, especially for the first splittings. In particular, it might be interesting to mitigate ISR's influence by resorting to the usual grooming techniques (e.g. [46]), allowing for a higher efficiency in FSR reconstruction and, as a consequence, more accurate predictions of formation times.

Following this, the formation time of the first splitting of a jet was studied in-depth. We concluded that it is the shortest splitting timescale for a jet whose branching history is assigned by the τ algorithm. This is a consequence of the exact ordering in E/m^2 along the branches of the corresponding tree, which we verified in a later section. Moreover, this quantity's value is independent of the algorithm chosen for reclustering. Its only dependence is the jet's 4-momentum, so its accuracy is dependent on the efficiency which one captures FSR with and its statistical dispersion is reflected by the allowed phase space of the jet. We saw, with parton-level FSR, that by selecting a narrower region of the jet's $p_t \cosh y$ while making a single experimental measurement - the jet's mass - one can determine the first splitting's formation time with a relative dispersion of, at most, 20% in our working conditions. Future relevant work may include studying how this correlation with the jet's mass changes when introducing ISR and working at hadron-level and how this estimate of the first splitting's formation time relates to the findings in [15]. Additionally, one can look for kinematic quantities which correlate better with the first splitting's formation time. The motivation for these studies is the conclusion that, in our working conditions and for the majority of cases, the first splitting will happen inside the QGP, for a jet produced in an heavy-ions collision. Hence, more accurate predictions of the first splitting's formation time are of the utmost importance for the investigation of the timescales in jet-medium interactions.

One then moved to an analysis of the remaining splittings' formation times. First of all, we verified that different reclustering algorithms generate different space-time structures for the same jet. The dissimilarities are most significant in early splittings. Hence, it is relevant what algorithm one chooses to assign a branching history, especially when implementing, for instance, an energy loss model which depends on the path length each parton traverses inside the medium [59]. It would be pertinent to investigate where, in each jet's phase space, the space-time structures differ the most. This task can be performed by, for instance, analysing specific regions of Lund jet planes [48]. Following that, one could try to suppress such regions and make the structure algorithm-invariant. In spite of the differences, the resulting mean formation times have relative errors of, at most, 25%. Not only this, but, sometimes, different branching histories may result in similar subjets. This means that, although the formation times inside each subjet may be unequal, the formation time of the partons originating them are the same irrespective of the algorithm. This was later verified analytically.

An attempt at understanding the kinematic dependences of the parton formation time was subsequently made. For that, one considered the calculation made with two approaches - local and global. The main takeaway point from the local approach is that the virtualities of the subsequent partons should not be neglected without analysing each case. For roughly 10% of the jets in our working conditions, neglecting parton virtualities results in a considerable overestimation of the corresponding formation times. This error can go up to orders of magnitude for emissions which are soft and collinear enough and produce partons with large enough virtualities. An interesting future task could be to attempt to suppress the highly soft and collinear splittings, such that virtualities could be ignored altogether and the calculation could be made simpler. With respect to the global approach, one turned explicit that the calculation of the space-time structure of a jet's branching history can be made solely through final-state information which can, in principle, be measured. Additionally, one concluded that it only matters what final partons belong to the subjet emerging from an "internal" parton. Nonetheless, the structure is still dependent on the reclustering algorithm. Furthermore, one verified that the factors $m_i^2/p_{t,i}^2$ are much smaller for final partons than for "internal" partons. Despite that, ignoring the final masses altogether (in this case, the quark masses) is still not viable since, in the formation time's denominator, they compete with every possible $p_i \cdot p_j$ between two partons of the relevant subset.

Following this, one made a thorough analysis of the spatio-temporal structure of τ branching histories of parton-level FSR. We started by exploring if the formation times had any well defined ordering along the spittings of the branching history. The central achievement was the verification that not only are formation times ($\tau_0 = E/m^2$) exactly ordered along the branches of the jet's tree, but the consecutive formation times are logarithmically separated - an average factor ~ 10 between them. Additionally, one concluded that, much like in parton shower generation, different branches have different logarithmic orderings. This implies that the inverse reclustering sequence does not necessarily present splittings in increasing formation time. Lastly, one verified that the absolute time splitting order is, for a substantial portion of jets, not the same as the one derived from the reclustering sequence.

In the last section, we understood that, because of the large available phase space for the substructure constituents, formation times inside a jet span about 12 orders of magnitude. Not only this, but one shows that much of the large dispersion is caused by the last splittings in a jet. For a portion of the events, these splittings contribute with disparately large formation times. In such cases, the corresponding subsequent partons are very close in azimuth-rapidity plane, thus are not even resolved by ordinary detectors. Hence, large outlier formation times are, most of the time, not detected. The bulk (90%) of the distribution lies in, approximately, 5 orders of magnitude. In particular, in our working conditions, one goes from $\sim 10^{-2}$ for the first splitting of a jet to $\sim 10^3$ for the last splittings of jets with $\gtrsim 20$ splittings. This range is still caused by the inclusive nature of the results, so one makes an attempt at reducing it by fixing a kinematical property of the jet - its multiplicity. This revealed a very well defined trend of the mean formation times along each splitting ordered in absolute time, with each splitting's formation time distribution being confined to 3 or 4 orders of magnitude. Further work on this topic could focus on searching for additional observable quantities which one can fix experimentally in order to reduce the dispersion of the overall trend of a jet's space-time structure. Most importantly, one found, by looking at jets with 3 different multiplicities, that the line of mean formation times suffers an overall decrease when increasing the multiplicity. This suggests that jets with higher multiplicity have a greater sensitivity to the medium's initial timescales. Future directions for this investigation might include exploring additional kinematic regions which correlate with overall shorter formation times inside a jet.

Future challenges include coupling the space-time structure of a vacuum jet to a model of the QGP. A trivial example of such a model would be that of a static "brick" with a fixed length. This model can take formation times into account by having a dependence on the distance each parton travels inside the region delimited by the brick's length. However, in order to study different space-time regions of the QGP's evolution, which is what one desires, one would need a dynamic model. By doing so, one could gain a better understanding of the relation between jet modifications and the interplay between the jet and the medium's temporal evolution. Finally, one could leverage this information to maximize jet sensitivity to medium interactions, allowing for a more detailed study of the QGP itself.

Bibliography

[1] Kjell Johnsen. The ISR in the time of Jentschke. *CERN Courier*, 2003.

[2] S. Ozaki. RHIC commissioning and first results. *PACS2001*, 2001.

[3] Busza, Wit and Rajagopal, Krishna and van der Schee, Wilke. Heavy Ion Collisions: The Big Picture and the Big Questions. *Annual Review of Nuclear and Particle Science*, 68(1):339–376, 2018.

[4] M. van Leeuwen. Results of the ALICE experiment. *54 Int. Winter Meet. on Nuc. Phys.*, 2016.

[5] (J. Adams et al.), STAR Collaboration. *Nucl. Phys.*, A757:102, 2005.

[6] B. Muller, J. Schukraft and B. Wyslouch. First Results from Pb+Pb Collisions at the LHC. *Ann. Rev. Nucl. Part. Sci.*, 62:361, 2012.

[7] Collins, J. C. and Perry, M. J. Superdense Matter: Neutrons or Asymptotically Free Quarks? *Phys. Rev. Lett.*, 34:1353–1356, 1975.

[8] Heinz, Ulrich and Snellings, Raimond. Collective Flow and Viscosity in Relativistic Heavy-Ion Collisions. *Annual Review of Nuclear and Particle Science*, 63(1):123–151, 2013.

[9] Gale, Charles and Jeon, Sangyong and Schenke, Björn. Hydrodinamy Modeling of Heavy-Ion Collisions. *International Journal of Modern Physics A*, 28(11):1340011, Apr 2013.

[10] Mehtar-Tani, Yacine and Milhano, J. Guilherme and Tywoniuk, Konrad. Jet Physics in Heavy-Ion Collisions. *International J. of Modern Physics A*, 28(11):1340013, 2013.

[11] Müller, Berndt and Schukraft, Jürgen and Wysłouch, Bolesław. First Results from Pb+Pb Collisions at the LHC. *Annual Review of Nuclear and Particle Science*, 62(1):361–386, Nov 2012.

[12] Casalderrey-Solana, Jorge and Milhano, José Guilherme and Quiroga-Arias, Paloma. Out of medium fragmentation from long-lived jet showers. *Physics Letters B*, 710(1):175–181, Mar 2012.

[13] Liliana Apolinário and José Guilherme Milhano and G. P. Salam and Carlos Salgado. Probing the Time Structure of the Quark-Gluon Plasma with Top Quarks. *Physical review letters*, 120 23: 232301 , 2018.

[14] Liliana Apolinário. Time Evolution of a medium-modified jet. *PoS (EPS-HEP 2019)*, 296, 2019.

[15] Liliana Apolinário and André Cordeiro and Korinna Zapp. Time reclustering for jet quenching studies, 2020.

[16] W. MissMJ. PBS NOVA, Fermilab, Office of Science, United States Department of Energy. Particle Data Group.

[17] Han, M. Y. and Nambu, Y. Three-Triplet Model with Double $\mathrm{SU}(3)$ Symmetry. *Phys. Rev.*, 139:B1006–B1010, Aug 1965.

[18] Gross, David J. and Wilczek, Frank. Ultraviolet Behavior of Non-Abelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.

[19] R.Michael Barnett et al. Review of particle physics. Particle Data Group. *Phys. Rev. D*, 54:1–720, 1996.

[20] Ellis, R. K. and Stirling, W. J. and Webber, B. R. *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

[21] Bjorken, J. D. and Paschos, E. A. Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon. *Phys. Rev.*, 185:1975–1982, Sep 1969.

[22] Bloom, E. D. and Coward, D. H. and DeStaebler, H. and Drees, J. and Miller, G. and Mo, L. W. and Taylor, R. E. and Breidenbach, M. and Friedman, J. I. and Hartmann, G. C. and Kendall, H. W. High-Energy Inelastic $e^{-}p$ Scattering at 6° and 10°. *Phys. Rev. Lett.*, 23:930–934, Oct 1969.

[23] R.P. Feynman. The behavior of hadron collisions at extreme energies. *Conf. Proc. C*, 690905:237–258, 1969.

[24] http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/qevid.html.

[25] Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 8\ \mathrm{TeV}$ with the ATLAS detector and determination of the strong coupling. *Phys. Rev. D*, 98:092004, Nov 2018.

[26] George F. Sterman and Steven Weinberg. Jets from Quantum Chromodynamics. *Phys. Rev. Lett.*, 39:1436, 1977.

[27] Simone Marzani, Gregory Soyez, and Michael Spannowsky. *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, volume 958. Springer, 2019.

[28] Gavin P. Salam. Elements of QCD for hadron colliders. In *2009 European School of High-Energy Physics*, 11 2010.

[29] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[30] Kinoshita,Toichiro . Mass Singularities of Feynman Amplitudes. *Journal of Mathematical Physics*, 3(4):650–677, 1962.

[31] Lee, T. D. and Nauenberg, M. Degenerate Systems and Mass Singularities. *Phys. Rev.*, 133:B1549–B1562, Mar 1964.

[32] Dasgupta, Mrinal and Dreyer, Frédéric A. and Hamilton, Keith and Monni, Pier Francesco and Salam, Gavin P. Logarithmic accuracy of parton showers: a fixed-order study. *Journal of High Energy Physics*, 2018(9), Sep 2018.

[33] Sjöstrand, Torbjörn et al. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[34] Sjöstrand, Torbjörn and Mrenna, Stephen and Skands, Peter. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026–026, May 2006.

[35] Bähr, Manuel and Gieseke, Stefan and Gigg, Martyn A. and Grellscheid, David and Hamilton, Keith and Latunde-Dada, Oluseyi and Plätzer, Simon and Richardson, Peter and Seymour, Michael H. and Sherstnev, Alexander and et al. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, Nov 2008.

[36] John E. Huth et al. Toward a standardization of jet definitions. In *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90)*, pages 0134–136, 12 1990.

[37] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012.

[38] S. Catani, Yu.L. Dokshitzer, M.H. Seymour, and B.R. Webber. Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions. *Nuclear Physics B*, 406(1):187 – 224, 1993.

[39] Dokshitzer, Yu.L and Leder, G.D and Moretti, S and Webber, B.R. Better jet clustering algorithms. *Journal of High Energy Physics*, 1997(08):001–001, Aug 1997.

[40] Cacciari, Matteo and Salam, Gavin P and Soyez, Gregory. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.

[41] Andrews et al. Novel tools and observables for jet physics in heavy-ion collisions, 2020.

[42] John Campbell, Joey Huston, and Frank Krauss. *The black book of quantum chromodynamics: a primer for the LHC era*. Oxford University Press, Oxford, 2018.

[43] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[44] M.H. Seymour. Tagging a heavy Higgs boson. In *ECFA Large Hadron Collider (LHC) Workshop: Physics and Instrumentation*, pages 557–569, 1 1991.

[45] Butterworth, J. M. and Cox, B. E. and Forshaw, J. R. WWscattering at the CERN LHC. *Physical Review D*, 65(9), May 2002.

[46] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[47] Butterworth, Jonathan M. and Davison, Adam R. and Rubin, Mathieu and Salam, Gavin P. Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider. *Phys. Rev. Lett.*, 100:242001, Jun 2008.

[48] Dreyer, Frédéric A. and Salam, Gavin P. and Soyez, Grégory. The Lund jet plane. *Journal of High Energy Physics*, 2018(12), Dec 2018.

[49] W. Bartel et al. Experimental Studies on MultiJet Production in e^+e^- Annihilation at PETRA Energies.

[50] Yuri L. Dokshitzer, A. Khoze, Alfred H. Mueller, and S.I. Troian. *Basics of perturbative QCD*. 1991.

[51] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in QCD. *Sov. Phys. JETP*, 46:641–653, 1977.

[52] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework. *Nucl. Inst. Meth. in Phys. Res. A*, 389:81–86, 1997.

[53] Sjöstrand, T. and Skands, P. Z. Transverse-momentum-ordered showers and interleaved multiple interactions. *The European Physical Journal C*, 39(2):129–154, Feb 2005.

[54] Dokshitzer, Yu. L. and Khoze, V. A. and Troyan, S. I. and Mueller, A. H. QCD coherence in high-energy reactions. *Rev. Mod. Phys.*, 60:373–388, Apr 1988.

[55] Andersson B. Dahlqvist P., Gustafson G. On local parton-hadron duality. *Zeitschrift für Physik C Particles and Fields*, 44:461–466, 1989.

[56] Aric A. Hagberg and Daniel A. Schult and Pieter J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux and Travis Vaught and Jarrod Millman, editor, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[57] Jorge Casalderrey-Solana, Yacine Mehtar-Tani, Carlos A. Salgado, and Konrad Tywoniuk. New picture of jet quenching dictated by color coherence. *Physics Letters B*, 725(4):357 – 360, 2013.

[58] Casalderrey-Solana, J. and Milhano, G. and Pablos, D. and Rajagopal, K. Modification of jet substructure in heavy ion collisions as a probe of the resolution length of quark-gluon plasma. *Journal of High Energy Physics*, 2020(1), Jan 2020.

[59] Casalderrey-Solana, Jorge and Gulhan, Doga Can and Milhano, José Guilherme and Pablos, Daniel and Rajagopal, Krishna. A hybrid strong/weak coupling approach to jet quenching. *Journal of High Energy Physics*, 2014(10), Oct 2014.

Appendix A

Formation time calculation for a single emission

A.1 Formation time calculation for a single emission

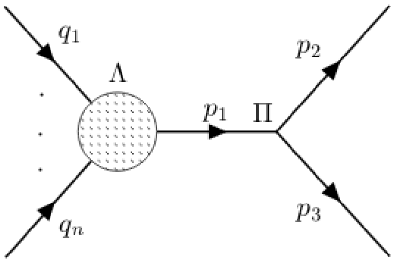


Figure A.1: Generic Feynman diagram for a single parton emission off a virtual parton.

The general form of the the Lorentz invariant amplitude $\mathcal{M}$ for the process in Fig. A.1 is

$$i\mathcal{M} = \Lambda(\{q_i\})\frac{1}{(p_2+p_3)^2 - m^2 + i\epsilon}\Pi(p_2,p_3) \tag{A.1}$$

where m is the on-shell mass of the internal line and the possible structures of Λ and Π, e.g., Dirac, Lorentz, flavour and color structure, were omitted. The $\Lambda(\{q_i\}))$ subprocess depends on the n ingoing momenta q_i and represents an arbitrary process from which a virtual parton (a quark or a gluon) with momentum $p_1 = \sum_i q_i$ can be produced. The $\Pi(p_2,p_3)$ depends on the outgoing momenta and contains the remaining terms of the QCD process, which can be $g \to gg$, $q \to qg$ or $g \to q\overline{q}$. It includes either the quark-gluon vertex or the 3 gluon vertex and the corresponding numerator of the propagator of parton p_1. These terms will contract with the necessary fermion spinors $\overline{u}(p_i)$ and gluon polarization vectors $\epsilon_\mu(p_i)$. Any dependence on p_1 in both Λ and Π amplitudes is already taken into account by momentum conservation ($\sum_i q_i = p_1 = p_2 + p_3$).

Let us focus on the relevant term for this calculation - the denominator of the propagator of the internal line. By resorting to the Dirac delta distribution one can re-write the propagator as

$$\frac{1}{(p_2+p_3)^2-m^2+i\epsilon}=\int d^4p_1\frac{\delta^{(4)}(p_1-p_2-p_3)}{p_1^2-m^2+i\epsilon}=\int d^4x\int\frac{d^4p_1}{(2\pi)^4}\frac{e^{-ix\cdot(p_1-p_2-p_3)}}{p_1^2-m^2+i\epsilon} \tag{A.2}$$

and the full amplitude in Eq. (A.1) is then re-written as:

$$i\mathcal{M}=\Lambda(\{q_i\})\Pi(p_2,p_3)\int d^4x\int\frac{d^4p_1}{(2\pi)^4}\frac{e^{-ix\cdot(p_1-p_2-p_3)}}{p_1^2-m^2+i\epsilon} \tag{A.3}$$

This manipulation of the amplitude allowed us to attach a space-time interval x to the momentum p_1 of the virtual particle. We will later use this interval as a measure of the parton formation time. To make the following calculations cleaner, let us re-write every 4-vector in light-cone coordinates, i.e.,

$$\begin{gathered}p=(p^0,\vec{p})\to p=\left(p^+,p^-,\vec{p}_t\right)\\ \text{where}\quad p^\pm=\frac{p^0\pm p^3}{\sqrt{2}},\quad \vec{p}_t=(p^1,p^2)\end{gathered} \tag{A.4}$$

with the dot product between two 4-vectors given by $x\cdot p=x^+p^-+x^-p^+-\vec{x}_t\cdot\vec{p}_t$ and volume element $d^4p=dp^-dp^+d^2p_t$. Notice that we are working in natural units $c=1$. In these coordinates, the Lorentz invariant amplitude can be written as

$$i\mathcal{M}=\Lambda(\{q_i\})\Pi(p_2,p_3)\int d^4x e^{ix\cdot(p_2+p_3)}\int\frac{dp_1^+d^2p_{1,t}}{(2\pi)^4}e^{-i(x^-p_1^+-\vec{x}_t\cdot\vec{p}_{1,t})}\frac{1}{2p_1^+}\int dp_1^-\frac{e^{-ix^+p_1^-}}{p_1^--\left(\frac{p_{1,t}^2+m^2}{2p_1^+}-i\epsilon\right)} \tag{A.5}$$

where one purposely isolated the integral in p_1^- to get an exponential in x^+. Alternatively, one could have isolated the integral in p_1^+. What matters is that we want to extract a relevant timescale, so it is natural that we focus on either x^+ or x^-.

The rightmost integral has a first order pole at $p_1^-=\frac{p_{1,t}^2}{2p_1^+}-i\epsilon$. It can be integrated out by using complex integration and defining a countour which runs clockwise and extends to the lower half of the complex plane. The result can be extracted by means of the residue theorem

$$\int dp_1^-\frac{e^{-ix^+p_1^-}}{p_1^--\left(\frac{p_{1,t}^2+m^2}{2p_1^+}-i\epsilon\right)}=-\theta(x^+)\exp\left\{-i\left(\frac{p_{1,t}^2+m^2}{2p_1^+}\right)x^+\right\} \tag{A.6}$$

where $\theta(x^+)$ is the usual unit step function. For $x^+<0$ one would close the contour in the upper half of the complex plane, in which case the result would vanish given that there is no complex pole there.

After integrating in x^- and $\vec{x}_t$ to get momentum conserving Dirac delta functions we end up with:

$$\begin{aligned}i\mathcal{M}=-\Lambda(\{q_i\})\Pi(p_2,p_3)&\int\frac{dp_1^+d^2p_{1,t}}{2p_1^+}\delta(p_1^+-p_2^+-p_3^+)\delta^{(2)}(\vec{p}_{1,t}-\vec{p}_{2,t}-\vec{p}_{3,t})\times\\ &\times\int dx^+\theta(x^+)\exp\left\{-i\left(\frac{p_{1,t}^2+m^2}{2p_1^+}-(p_2^-+p_3^-)\right)x^+\right\}\end{aligned} \tag{A.7}$$

Calculating the cross section for the process in Fig. A.1 would demand that we square the Lorentz invariant amplitude in Eq. (A.7). Because the term multiplying x^+ in the exponential is fixed by the Dirac delta functions, once we square the amplitude we get

$$|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^\dagger \propto \int\int dx^+ d\overline{x}^+ \theta(x^+)\theta(\overline{x}^+) \exp\left\{-i\left(\frac{p_{1,t}^2 + m^2}{2p_1^+} - (p_2^- + p_3^-)\right)(x^+ - \overline{x}^+)\right\} \tag{A.8}$$

where $\vec{p}_{1,t} = \vec{p}_{2,t} + \vec{p}_{3,t}$ and $p_1^+ = p_2^+ + p_3^+$. The rightmost term on the exponential's argument $\Delta x^+ = x^+ - \overline{x}^+$ can be identified as a time difference between the emission in $\mathcal{M}$ and the emission in its conjugate amplitude $\mathcal{M}^\dagger$. In turn, this time difference can be interpreted as the lifetime of the virtual state p_1, i.e., the parton formation time. Of course that, for the purpose of calculating the cross section of a process, it is meaningless to ask what the actual value of Δx^+ is, since it is an integration variable. Nevertheless, the information we extract from this calculation is meaningful - we identify the relevant space-time region for the integration in Eq. (A.8). This region is identified in the exponential's argument through the following timescale

$$\Delta x^+ \sim \frac{2p_1^+}{p_1^2 - m^2} \tag{A.9}$$

where $p_1^- = p_2^- + p_3^-$ was used. The light-cone coordinates written with respect to the direction of parton p_1 ($p_1^3 = |\vec{p}_1|$) result in:

$$\Delta x^+ = (\tau + x^3)/\sqrt{2} = (1 + E_1/|\vec{p}_1|)\tau/\sqrt{2} = (p_1^+/|\vec{p}_1|)\tau \tag{A.10}$$

For quarks with virtualities much greater than their on-shell masses, i.e., $p_1^2/m^2 \ll 1$, one can then write the parton formation time as:

$$\tau \sim \frac{2|\vec{p}_1|}{p_1^2} \tag{A.11}$$

If one further assumes the low virtuality limit, i.e., $m_1/|\vec{p}_1| \sim m_1/E_1 \ll 1$, then $E_1 \approx |\vec{p}_1|$ and we end up with

$$\tau \sim \frac{2E_1}{p_1^2} \tag{A.12}$$

which is, up to a factor of 2, the parton formation time one derives from simple arguments based on the Heisenberg's uncertainty principle and special relativity kinematics (Section 2.6).