Restoring string fragmentation self-similiarity in PYTHIA hadronisation

Jade Abidi

Student ID: 31461964

A thesis submitted for the degree of **Bachelor of Science (Honours)**

November 2025

School of Physics and Astronomy
Monash University

Supervisor: Peter Skands

Abstract

Monte Carlo event generators like PYTHIA are used to simulate high-energy particle collision events involving non-perturbative physics. In PYTHIA, the strong field between a $q\bar{q}$ pair is modelled as a classical string with constant tension, implying that the string fragmentation process should be self-similar at all points along the string. This property is currently violated at the step where string ends are joined, causing a dip in the rapidity plateau and anomalous hadronic chemistry. We introduce an additional tunable parameter and a new algorithm for string fragmentation that improve or resolve these issues, albeit with some limitations.

Contents

Chapter 1

Introduction

The field of particle physics is dedicated to investigating the most fundamental particles and interactions in nature. It naturally evolved from nuclear and atomic physics in the early 20th century as technological and scientific knowledge allowed scientists to probe matter at higher energies and smaller length scales. The physical theory underpinning particle physics developed out of quantum field theory, culminating in the Standard Model of particle physics which was formalised in the 1980s. The Standard Model unifies three of the four fundamental forces of nature (electromagnetism, the weak force, and the strong force) into a single theory, and predicted the existence of the Higgs boson well before its historic discovery at the Large Hadron Collider (LHC) in 2012.

Particle physics experiments typically require particle colliders like the LHC. Such colliders accelerate charged particles, typically electrons or protons, to speeds up to 99.99999% of the speed of light. Modern particle colliders are usually circular, and use high voltages and strong magnetic fields to accelerate particles and keep them within a thin beamline. These beamlines are then made to collide at interaction points, around which bespoke detector systems are able to collect data on the final state.

Chapter 2

QCD, the Lund Model, and PYTHIA

2.1 Quantum Chromodynamics and Collider Physics

1. Explain the history of quantum, atomic, and nuclear physics.
2. Explain the development of quantum field theory and lay out the QFT basics.
3. Explain the development of QCD and the Standard Model.
4. Figure: Running coupling of QED.
5. Figure: Running coupling of QCD.
6. Explain quantum chromodynamics, including SU(3) gauge symmetry, asymptotic freedom, gluon-gluon interactions, and confinement.
7. Pivot to experiment. Explain why particle colliders are necessary. Explain what they do and how they work.
8. Explain how detector systems work in colliders, and what these detectors output.
9. Figure: The ALICE detector at the LHC.
10. Explain concepts like luminosity and cross sections. Provide an outline of important experimental discoveries, such as the discovery of gluons in the JADE experiment, or the top quark and Higgs boson in the LHC.
11. Figure: The discovery of gluons, including the use of JETSET.

2.2 Monte Carlo Event Generators and PYTHIA

1. Explain what a Monte Carlo event generator is and what role it plays in particle physics.
2. Explain why Monte Carlo simulation is necessary to generate events, including the difficulties with non-perturbative QCD and numerical calculations like lattice QCD.

3. Sketch the history of Monte Carlo event generators, including JETSET, PYTHIA, HERWIG, and more. Outline differences between these generators.
4. Elaborate on the history and features of PYTHIA. Include a few potential uses.
5. Explain how an event is generated in PYTHIA. Include a high-level mathematical formulation of the hard scattering, parton shower, hadronisation, and decays. Introduce and explain why factorisation allows us to consider different energy scales of event generation independently of each other.
6. Elaborate on the hadronisation process. Emphasise the necessity of phenomenological models due to the non-perturbative physics involved. Signpost the Lund model.
7. Explain the necessity of tuning and how this process works.

2.3 The Lund String Model

1. Introduce the history and development of the Lund string model. Outline a high level of how it models hadronisation using string breaks.
2. Figure: Hadronisation according to the Lund model.
3. Explain how the Lund model describes the strong colour field as a flux tube with linear potential.
4. Explain the experimental and phenomenological justification for the Lund model. Include lattice QCD simulations of the string tension, and the existence of ggg and gggg interaction vertices in QCD as a reason for the flux tube behaviour.
5. Figure: The electric field vs. the colour field
6. Introduce the yo-yo mode and explain how it is a model for hadrons in the Lund model. Introduce diquarks.
7. Figure: Spacetime diagram of the yo-yo mode.
8. Introduce rapidity and lightcone momenta.
9. Figure: Rapidity vs velocity.
10. Explain how string fragmentation works in the Lund model. Establish that the fragmentation process is fully specified by z fractions. Introduce the fragmentation functions and lightcone scaling, as well as the Schwinger mechanism.
11. Figure: Spacetime diagram of string fragmentation.
12. Briefly introduce gluon kinks and more complex string topologies.
13. Figure: String topologies in the Lund model.
14. Emphasise how the string fragmentation properties are asymptotic in the limit where energy-momentum conservation is not a consideration.

Chapter 3

String Fragmentation in PYTHIA

3.1 The PYTHIA Fragmentation Algorithm

1. Outline what a successful implementation of the Lund model would achieve. Note how the problem is somewhat undefined in terms of energy-momentum conservation.

2. Explain the current hadronisation algorithm in PYTHIA.

3. Outline the Eden paper and how the PYTHIA manuals describe the joining step and energy-momentum conservation in PYTHIA and the Lund model.

Having established the theoretical basis of the Lund string model for hadronisation, we can now describe how the hadronisation process is algorithmically implemented in PYTHIA. We begin by outlining what a "successful" implementation of the Lund string model would achieve. As mentioned in section 2.3, the 1+1-dimensional kinematics (((TODO: Maybe rephrase this?))) of string fragmentation in a single event that produces N hadrons are completely specified by a set of N absolute lightcone momentum fractions $\{z^+_{\mathrm{abs},i}\}$, where we are considering fragmentation right-to-left (but could just as well consider it left-to-right).

To conserve energy and momentum, these lightcone momentum fractions must add to unity, that is,

$$\sum_{i=1}^{N} z^+_{\mathrm{abs},i} = 1. \tag{3.1}$$

The area law and lightcone scaling properties of the Lund string model require that the relative lightcone momentum fractions $\{z^+_i\}$ of a single event (as defined in section 2.3) are all drawn from a given fragmentation function $f(z)$ — specifically, the Lund symmetric fragmentation function specified in equation (((insert equation here))).

Also established in section 2.3 is the fact that the quarks produced along the string must have masses m_q and transverse momenta $p_{\perp,q}$ drawn from a distribution $\Pr(m_q^2, p_{\perp,q}^2)$ with a Gaussian suppression, as in equation (((insert equation here))). The resulting hadrons formed from these quarks (and antiquarks) must have masses and transverse momenta distributed accordingly.

(((TODO: Is this a good description? Is this necessary? What about the distribution of N?)))

The actual implementation of string fragmentation in PYTHIA is given by the high-level pseudocode in (((link))). Here, we are considering the simplest fragmentation process where a quark q_0 and antiquark $\bar{q}_0$ move in opposite directions along the z-axis with centre-of-mass energy E_{CM}. An actual event in PYTHIA will contain many such processes between different partons produced in the parton shower, and will also require the consideration of gluon kinks along the strings. (((TODO, elaborate, make less vague))) However, as we will see, lightcone scaling is entirely violated in PYTHIA even in this minimal situation, and as such the rest of this thesis will be limited to simple $q\bar{q}$ hadronisation.

Algorithm 1 The default PYTHIA 8.3 algorithm for $q\bar{q}$ hadronisation

```
procedure FRAGMENT(E_CM, flav(q_0), flav(q̄_0))
    initialise event record event
    i ← 1
    p_x(q_0) ← 0.0
    p_y(q_0) ← 0.0
    p_x(q̄_0) ← 0.0
    p_y(q̄_0) ← 0.0
    loop
        fromPos ← true or false with equal probability
        flav(q_i) ← flavour according to Gaussian suppression ((equation)) and PYTHIA
weights
        flav(q̄_i) ← antiflavour of flav(q_i)
        p_x(q_i) ← transverse momentum according to Gaussian suppression ((equation))
        p_x(q̄_i) ← −p_x(q_i)
        p_y(q_i) ← transverse momentum according to Gaussian suppression ((equation))
        p_y(q̄_i) ← −p_y(q_i)                                  ▷ String break done.
        if fromPos then
            event[i].id ← hadron selected from combination of flav(q_{i−1} and flav(q̄_i)
            event[i].m ← mass selected according to Breit-Wigner distribution
            event[i].p_x ← p_x(q_{i−1}) + p_x(q̄_i)
            event[i].p_y ← p_y(q_{i−1}) + p_y(q̄_i)
        else
        end if
    end loop
end procedure
```

3.2 The Joining Step

1. Explain in close detail how the finalTwo joining step works in PYTHIA.
2. Explain the stopMass, stopNewFlav, and stopSmear parameters work.
3. Note the inherent violation of lightcone scaling. Explain how the manual and documentation claim this is resolved.

3.3 Performance of the Current finalTwo Procedure

1. Figure: dN/dy distributions in PYTHIA 8.3 vs 8.0 vs 6

2. Explain how bad the problem is with rapidity distributions. Also mention the tune loading bug.

3. Explain why this happens in terms of the stopMass parameter, rapidity spacing, and fragmentation functions, as well as the non-uniformity of the joining step rank.

4. Data: Example ratio differences and SSE of hadronic chemistry in the joining step, PYTHIA 6 vs 8.

5. Explain the bias introduced by the finalTwo failure rate and how this leads to an anomalous hadronic chemistry.

Chapter 4

Tuning Lightcone Scaling in PYTHIA

4.1 Restoring Lightcone Scaling by Tuning Parameters

1. Explain the necessity of tuning the joining step, in contradiction to what was asserted by the manual.

2. Explain the tradeoff between hadronic chemistry and kinematics of the joining step hadrons, and the difficulty in tuning for both with only one degree of freedom.

3. Data: Rapidity plateaus, SSEs of rapidity plateaus and hadronic chemistry across different joining step parameters and tunes.

4.2 The probRevertBreak parameter

1. Introduce the probRevertBreak parameter and the pseudocode. Explain how the algorithm works and what is changed. Include formulas for conditional spin switching.

2. Explain how the probRevertBreak parameter reduces bias and improves hadronic chemistry, as well as providing more freedom to tune the joining step.

3. Show results (SSEs and rapidity plateaus) of probRevertBreak and the possibility for improvement.

4.3 Limitations

1. Explain the limitations of this set of parameters, showing plots of rapidity differences at the joining step vs everywhere else.

2. Re-emphasise how issues like anomalous hadronic chemistry and the rapidity plateau are not fixed.

3. Note the issues with finalTwo failing. Cite a few papers that are affected by this.

Chapter 5

The Accordion Algorithm for String Fragmentation

5.1 The Accordion Algorithm

1. Motivate the concept behind the accordion algorithm. Re-emphasise the self-similarity and causal independence of breakup vertices.
2. Establish the goals of the algorithm - a flat rapidity plateau, correct hadronic chemistry, and a lower failure rate.
3. Explain how the algorithm works. Derive equations relating rapidity spacing and z fractions. Explain the accordion rescaling and why numerical solution is required.
4. Elaborate on limitations of the accordion rescaling, including the shaky Lorentz covariance and effect on fragmentation functions. Explain why fragmentation functions may not matter.

5.2 Pseudocode

1. Show and explain pseudocode of the accordion algorithm.

5.3 Results

1. Show off!

5.4 Limitations

1. Be honest. But in a nice way.
2. Not implemented for popcorn model.
3. Needs to be expanded to gluon kinks and string topolgies - not clear how this might work.
4. Necessity of tuning stopMass.

5. Need more investigation into correlations and fragmentation functions.

Chapter 6

Summary and Outlook

1. Summarise the achievements and results, re-establishing their context in the broader field.
2. Establish further avenues of investigation, including: expanding the scope of the algorithm, improving the rescaling step, improving the rapidity spacing sampling, investigating time complexity and performance, investigating the effects of the rapidity dip in other findings and papers

Bibliography

[1] Albert Einstein. On the electrodynamics of moving bodies. *Annalen der Physik*, 17:891–921, 1905.