

A Lorentz covariant model of string fragmentation

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 are extensively used to simulate high-energy particle-collision events. For analytically intractable aspects, they rely on phenomenological models. The so-called Lund model describes the non-perturbative hadronisation process as the fragmentation of a classical string with constant tension. Lorentz covariance then implies the self-similarity of this fragmentation process along the string. The current formulaion of the Lund model, used in the PYTHIA generator, violates this property in terms of both kinematic distributions and 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

1	Introduction	2
2	QCD, the Lund Model, and PYTHIA	4
2.1	Quantum Chromodynamics and Collider Physics	4
2.2	Monte Carlo Event Generators and PYTHIA	9
2.3	The Lund String Model	11
3	String Fragmentation in PYTHIA	15
3.1	The PYTHIA Fragmentation Algorithm	15
3.2	The Joining Step	18
3.3	Performance of the Current finalTwo Procedure	19
4	Tuning Lightcone Scaling in PYTHIA	23
4.1	Restoring Lightcone Scaling by Tuning Parameters	23
4.2	The probRevertFinal parameter	24
4.3	Limitations	25
5	The Slinky Model of String Fragmentation	26
5.1	The Slinky Model	26
5.2	Pseudocode	27
5.3	Results	27
5.4	Limitations	27
6	Summary and Outlook	28

Chapter 1

Introduction

The field of particle physics investigates 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 [1,2]. 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 [3–5].

The interactions investigated in particle-physics experiments occur at high energies, with low probabilities. Particle colliders, such as the LHC, are designed specifically to enable such experiments by accelerating charged particles, usually electrons or protons (and their antiparticles), to speeds up to 99.9999% of the speed of light. Most modern particle colliders are circular, and use high voltages and strong magnetic fields to accelerate particles and keep them within a thin beamline. The particle beams are then made to collide millions of times every second at interaction points, around which bespoke detector systems collect data on the final state [1,6,7].

For comparison against this data, it is very useful to be able to generate ensembles of large numbers of events based on physical models and assumptions. Because the physics involved is so complex and often lacks an analytic solution, Monte Carlo event generators such as PYTHIA provide a way to sample the phase space efficiently [1,8–10]. Powerful factorisation theorems allow for the various processes that occur at different energy scales to be modelled independently of each other, which greatly simplifies the problem [10,11]. The focus of this project is the process of hadronisation, in which partons (quarks and gluons) form hadronic bound states. This occurs at momentum transfer scales below the confinement energy of $\Lambda_{\text{QCD}} \sim 0.2 \text{ GeV}$ [10].

Due to asymptotic freedom, the coupling constant of the strong force becomes very large for momentum transfers in this range [12,13], and therefore the physics of hadronisation cannot be calculated perturbatively. Non-perturbative quantum chromodynamics has not yet been analytically solved, and numerical methods such as lattice QCD have limitations that render them inapplicable to hadronisation [14]. Instead, PYTHIA uses the Lund string model as a description of hadronisation, which is a phenomenological theory (and is not derivable from first principles) [8,15].

In the Lund string model, the strong field between a quark and antiquark is modelled as a string with constant tension $\kappa \sim 0.9 \text{ GeV}$ per femtometre, as measured in lattice

QCD simulations [10, 16]. At separation distances of $\gtrsim 1$ fm, there is enough potential energy stored in the string for it to ‘‘break’’, forming a new quark-antiquark pair. This leaves two quark-antiquark string subsystems, which proceed to break in the same manner as the original system, giving rise to a recursive self-similar process resulting in a set of outgoing bound states of quarks [8, 10, 15, 17].

This self-similarity of the fragmentation process along the string is one of the most fundamental properties of the Lund string model. The string breaks are all spacelike separated and hence cannot causally influence each other, and the tension along the string is constant and invariant under longitudinal boosts. It therefore follows that, away from the endpoints of the string, all fragmentation observables (like the number or species of hadrons produced from a region of the string) must be invariant under Lorentz boosts (or equivalently, rapidity translations) along the string axis. This means that distributions such as hadron density per unit rapidity should be flat except for endpoint effects (denoted the ‘‘rapidity plateau’’). It also implies that, if considered iteratively, the process by which each string subsystem is fragmented must be scale invariant [8, 15, 17].

((TODO: Rewrite this and next paragraph to be more clear about what `finalTwo` is, what it does, and its importance)). Despite the importance of the Lorentz covariance of string fragmentation in the Lund model, current simulations in PYTHIA reveal that this property is quite badly broken. Since the release of PYTHIA 8.0, the rapidity plateau has exhibited a significant dip in the central region. Initial investigations revealed this to be a consequence of the hadronisation algorithm used in PYTHIA, where the final two hadrons have kinematics forced by energy-momentum conservation and the mass-shell relation. In the Monash tune of PYTHIA 8.3, the rapidity spacing between these final two hadrons is larger than the typical rapidity spacing. This, combined with the non-uniformity of location of the joining step along the string, gives rise to this central rapidity dip. Furthermore, the `finalTwo` procedure that generates these hadrons fails about 50% of the time, introducing bias that skews the species makeup of these final two hadrons.

PYTHIA offers three tunable parameters that adjust the behaviour of this procedure — `stopMass`, `stopNewFlav`, and `stopSmear`. Prior to this project, it was believed that these parameters could be tuned to give a flat rapidity plateau, and would not need to be retuned when other fragmentation parameters change [8, 18]. This is not the case for the Monash tune, where it is impossible to achieve a flat plateau with these parameters without significantly worsening the failure rate of `finalTwo`.

In this thesis, chapter 2 provides a review of the theoretical background of quantum chromodynamics, PYTHIA, and the Lund string model. In chapter 3, the current behaviour of the PYTHIA hadronisation algorithm is analysed in more detail, and the various problems are described and explained. In chapter 4, the tuning of the `finalTwo` parameters is discussed. An additional tunable parameter `revertFinalBreak` is presented which provides an additional dimension of parameter space and allows for a flatter rapidity plateau to be obtained without as much of an impact on the failure rate. Finally, in chapter 5, an alternative algorithm for hadronisation in PYTHIA is developed, called the accordion join, which achieves an approximately flat rapidity plateau for any tune, and fails more than a hundred times less frequently.

Chapter 2

QCD, the Lund Model, and PYTHIA

2.1 Quantum Chromodynamics and Collider Physics

Before introducing PYTHIA and the Lund string model, we first provide an overview of the theoretical models and experimental techniques used in particle physics, with a focus on quantum chromodynamics.

The field of particle physics developed out of the study of atomic and nuclear physics in the early 20th century. Experiments like the discovery of the nucleus in 1911 [19] or the neutron in 1932 [20] paved the way for the development of more advanced technologies such as particle accelerators and colliders. The high centre-of-mass energies reached in these experiments allowed for the discovery of a slew of particles in the 1950s and 1960s, dubbed the “particle zoo” [2]. Alongside these experiments, the theoretical success of quantum mechanics inspired the quantisation of the electromagnetic field in the original formulation of quantum field theory (QFT) and quantum electrodynamics (QED) by Dirac, who notably predicted the existence of antimatter a number of years before its discovery [21,22]. QED was later unified with a description of the weak force (responsible for phenomena like beta decay) into electroweak theory, a Yang-Mills gauge theory consisting of a local $U(1) \times SU(2)$ gauge symmetry [1,23].

At the same time, quantum chromodynamics (QCD) emerged, describing the variety of particles discovered in the “particle zoo” as composed of more fundamental subatomic particles — quarks and gluons. Initial models like the Eightfold Way classified these various strongly interacting particles as hadrons emerging from a $SU(3)$ flavour symmetry between three quark flavours [24,25], as shown in Figure 2.1. These flavours were later denoted the up (u), down (d), and strange (s) quarks, and three more (charm, beauty, and top) were eventually found. Hadrons were classified into two groups — mesons, which are bosonic bound states consisting of a quark and an antiquark, and baryons, which are fermions consisting of three quarks or antiquarks. The experimental reality of these quarks outside of phenomenology was confirmed by deep inelastic scattering experiments, which showed pointlike constituents within the proton, called partons [6,23].

It was later found that these partons come in two types — the spin $\frac{1}{2}$ quark, and the spin 1 gluon [1]. The existence of the gluon was confirmed in the JADE experiment [26], which was notably one of the first uses of Monte Carlo event generators in the analysis of experimental data. In addition to their electromagnetic charge, quarks and gluons also carry colour charge, which has three components labelled red, green, and blue (with cyan,

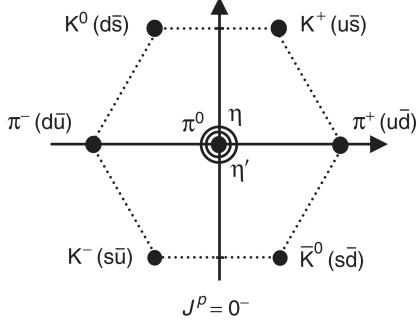


Figure 2.1: A visual depiction of the pseudoscalar mesons in the Eightfold Way, which form an octet and a singlet in SU(3) flavour space [24, 25]. The horizontal axis represents I_3 , an isospin component, and the vertical axis represents Y , the hypercharge, both of which are quantum numbers arising from the SU(3) flavour symmetry. Figure taken from [23].

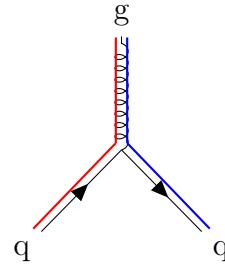


Figure 2.2: An illustration of a qqq vertex in QCD, showing the colour flow. The incoming and outgoing quarks have one colour, while the gluon has two (a colour and anticolour).

magenta, and yellow as their anticolour opposites) [1, 23].

In the 1970s, a formal theory of QCD based on a SU(3) gauge symmetry was developed, and by 1980, the Standard Model was formalised, which unified the electromagnetic, weak, and strong forces into a single unified gauge theory with three generations of fermions (6 quarks and 6 leptons). Collider experiments throughout the following years continued to confirm predictions of the Standard Model, and it has since become our most successful and fundamental model of physical reality [3, 23].

A full theoretical description of QFT and QCD is outside the scope of this review, and we will only provide a brief summary in order to establish the necessary theoretical background. The reader is directed to the textbook by Peskin and Schroeder [27] for a general overview of quantum field theory, or the textbook by Ellis, Stirling, and Webber [6] for closer detail on quantum chromodynamics or collider physics.

Quantum field theories model particles as excitations of underlying quantised fields, which are operator-valued at every point in space and time. Particles interact via the exchange of bosons as virtual particles, and in this way the concept of a “force” is reduced to an allowed interaction between particles. As in quantum mechanics, these particles can be more or less localised in coordinate or momentum space, with these two uncertainties linked by Heisenberg’s relation

$$\Delta x \Delta p_x \geq \frac{\hbar}{2}. \quad (2.1)$$

Typically, in particle physics, particles are described as entirely delocalised plane waves with precisely defined energies and momenta. This is justified because the length scales involved in particle collisions are much smaller than typical De Broglie wavelengths, and in this limit the physical extent of the particle’s wave packet is not relevant.

The physical laws governing these quantised fields are captured in the Lagrangian, and can be derived via the Euler-Lagrange equations. In a gauge theory, these Lagrangians are derived to obey underlying local gauge symmetries. QCD is based on a local SU(3) gauge symmetry, which implies that the wavefunction must carry three additional degrees of freedom representing the colour charge. The three fundamental colour states

corresponding to red, green, and blue are

$$r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (2.2)$$

A local SU(3) gauge transformation is a rotation of states in colour space, where the axis and amount of rotation can vary throughout time and space. The Noether invariant corresponding to this symmetry is colour charge, which is a conserved quantity in QCD.

In QCD, the strong interaction is mediated by the eight gluons, corresponding to the eight generators of the SU(3) group. Unlike quarks, gluons carry a combination of colour charge and anticharge. Since quarks are the only fermions that carry colour charge in the Standard Model, only quarks can couple to gluons. The Lagrangian density describing QCD is given by

$$\mathcal{L} = \bar{\psi}_q^i (i\gamma^\mu) (D_\mu)_{ij} \psi_q^j - m_q \bar{\psi}_q^i \psi_{qi} - \frac{1}{4} F_{\mu\nu}^\alpha F^{\alpha\mu\nu}. \quad (2.3)$$

Here, the first term describes the kinetic energy of quarks and their interaction with gluons, the second term endows quarks with their mass, and the third term allows for gluon-gluon interactions. Figure 2.2 shows how colour flows through a quark-gluon Feynman interaction vertex. Because the generators of SU(3) do not commute, QCD is a non-Abelian gauge theory and therefore gluon-gluon interactions are also possible. This is in contrast to quantum electrodynamics, where there is no photon-photon coupling term in the Lagrangian [1, 6, 23].

An important property of QCD is confinement. The quarks and gluons that make up hadrons have never been observed on their own as free particles. In nature, they seem to always be confined to hadronic bound states, and cannot be separated beyond the typical hadron size of ~ 1 fm. Confinement has not been mathematically proven—indeed, its proof is one of the Millennium Prize Problems (a consequence of the Yang-Mills existence and mass gap). However, it can still be understood as a consequence of the running coupling of QCD. Figure 2.3 shows how the coupling constant (describing the strength of the strong interaction) decreases with the energy scale of the interaction [1]. As such, at short distances (and high energies) the strong interaction is weak, but at longer distance the strong interaction becomes strong enough to forbid the separation of quarks and gluons. This phenomenon is known as asymptotic freedom [12, 13].

The strong potential between quarks and antiquarks is given by the Cornell potential,

$$V(r) = -\frac{4}{3} \frac{\alpha_s}{r} + \kappa r, \quad (2.4)$$

which combines a short-distance Coulomb potential with a stronger, longer-distance linear potential. Here, α_s is the strong coupling constant, and κ is a constant measured to be approximately 1 GeV fm^{-1} . Figure 2.4 shows lattice QCD calculations of the strong potential between a $q\bar{q}$ pair, exhibiting Coulombic behaviour at short distances and becoming linear for longer distances. Because QCD allows for gluon-gluon interactions, the linear potential can be understood as a result of the self-attraction of strong field lines, which compresses the field into a flux tube with constant tension. It is this linear potential that gives rise to the confinement distance of approximately 1 fm, at which it is energetically favourable for the strong field to break into a new quark-antiquark pair, forming a new hadron. This “string-breaking” model of hadron formation is the basis for the Lund string model, described further in section 2.3.

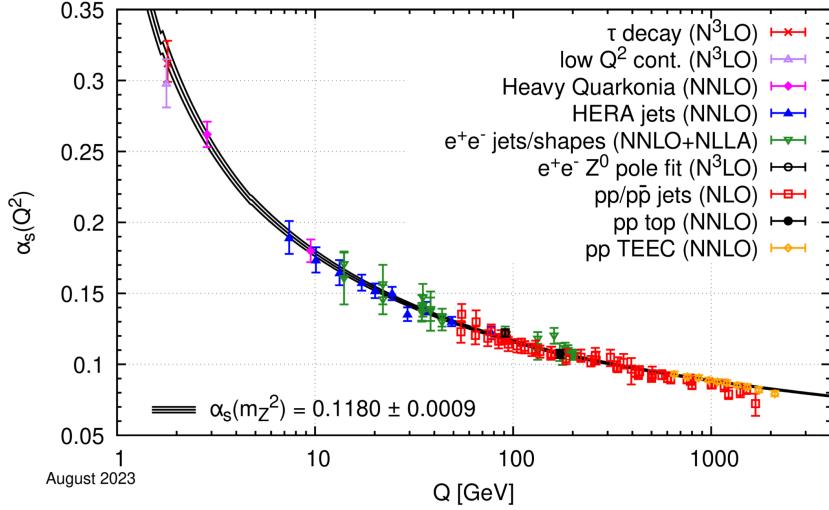


Figure 2.3: The running of the QCD coupling constant $\alpha_S(Q^2)$ with respect to the energy scale Q . The PDG average is shown as a solid line alongside data points from numerous experiments. The asymptotic freedom of QCD is evident from how the coupling constant is sufficiently small for perturbation theory at large values of Q , but becomes prohibitively large for smaller values of Q . Figure taken from [1].

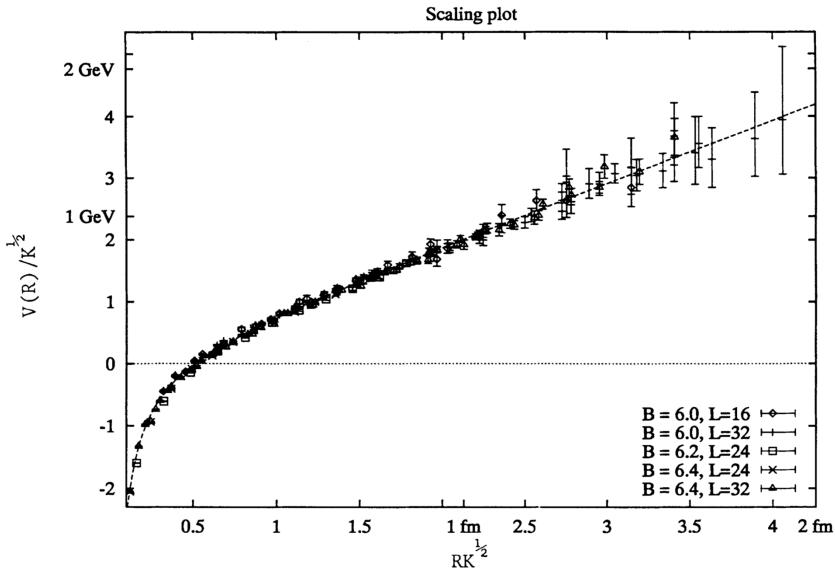


Figure 2.4: The potential between a quark and antiquark as a function of distance. Data points are obtained from lattice QCD simulations [16], and the dotted line shows the Cornell potential as defined in equation (2.4). Except for at short distances, the potential is approximately linear, motivating the treatment of the colour field as a classical string with constant tension. Figure taken from [16].

As mentioned in the introduction, experiments in particle physics require very high centre-of-mass energies for the interactions under observation to be kinematically allowed. If two particle beams have energies E_1 and E_2 , then the resulting relativistic centre-of-mass (CM) energy is given by

$$E_{\text{CM}} \approx 2\sqrt{E_1 E_2}, \quad (2.5)$$

which reduces to $E_{\text{CM}} = 2E_b$ if the two beam energies are the same, which is common in modern colliders. The Large Hadron Collider has achieved CM energies up to 13.6 TeV.

Another important property of particle colliders is luminosity. While the beam and centre-of-mass energies describe the energies achieved by individual particles in the collider, the luminosity quantifies the flux of particles in the beam. Luminosities in modern colliders are quite large; the most recent LHC experiments have luminosities around $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

The “probability” of a given interaction occurring is quantified by the cross section σ . The cross section has units of area, but typical cross sections are so small that the standard unit is barns, where 1 barn is equal to $1 \times 10^{-28} \text{ m}^{-2}$. Despite not actually representing a physical area, the cross section generalises the notion of the area of a target representing its probability of being struck by an incoming particle. The variation of the interaction probability over the values of a variable, such as the solid angle Ω , is often expressed using the differential cross section $d\sigma/d\Omega$, which is related to the overall cross section by

$$\sigma = \int \frac{d\sigma}{d\Omega} d\Omega. \quad (2.6)$$

Here, Ω can be replaced with any number of variables, such as the Mandelstam variable s which quantifies the CM energy of the collision. If $\mathcal{L}(t)$ is the instantaneous luminosity of a particle beam, then the expected number of events with cross section σ is

$$N = \sigma \int \mathcal{L}(t) dt. \quad (2.7)$$

In this way, the cross section solely represents the underlying physics of an interaction, with the specific experimental collider setup factorised out.

The most important particle colliders for experimental QCD are typically proton-proton (pp) colliders, also called hadron colliders. The first hadron collider was the Intersecting Storage Rings (ISR) at CERN, which were operational from 1971, with a centre-of-mass energy of 62 GeV [1, 28, 29]. It was succeeded by the Tevatron which, as suggested by its name, reached centre-of-mass energies of 1 TeV [30]. Notable discoveries such as that of the top quark were made by the Tevatron [31, 32]. Currently, the largest hadron collider is the creatively named Large Hadron Collider (LHC) described above, which was operational from 2008 and is responsible for the monumental discovery of the Higgs boson in 2012 [4, 5, 7]. A photo of the Compact Muon Solenoid (CMS) detector, one of the particle detectors in the LHC, is shown in Figure 2.5.

Detectors are only able to measure the final state of an event, and ultimately output a list of the types, masses, and four-momenta of the outgoing particles. For high-luminosity colliders like the LHC, many billions of these event records are recorded. For many theoretical aspects of QCD, it is only possible to compare experimental data to theory by using Monte Carlo techniques to generate ensembles of events based on theoretical assumptions. As mentioned earlier, the discovery of the gluon in the JADE experiment utilised the JETSET generator. The next section outlines the functionality of these event generators, with a focus on PYTHIA, the modern successor to JETSET.



Figure 2.5: The ATLAS detector, the largest of the particle detectors at the LHC. It was involved in the discovery of the Higgs boson in 2012 [4]. The particle beams are contained within the thin tube visible at the top centre of the image, and the rest of the structure contains detectors that measure the energy and momentum of final state particles. Figure from [33].

2.2 Monte Carlo Event Generators and PYTHIA

Essential to any scientific undertaking is the ability to compare the predictions of a theoretical model with the results of an actual experiment. In many fields of physics, making a prediction of experimental results is a simple matter of a mathematical derivation or computation. Even within particle physics, discoveries such as that of the Higgs boson often have clear experimental signatures, such as resonances in distributions of particles over observables like invariant mass [4, 5], and these signatures can be identified with model fitting or machine learning to extract the signal from the background. On the other hand, some experimental predictions are very far removed from the underlying theory. For example, in the JADE experiment, the existence of the gluon was inferred not from any resonant peak, but rather from a comparison of the jet mass distribution with the results of detailed simulations of particle collisions with and without gluons. Similarly, the discovery of the top quark required use of simulations to determine the expected signal and background of various distributions [31, 32].

Simulating a particle collision (often referred to as an “event”) is not a simple task, for a variety of reasons. QCD has the property of asymptotic freedom, meaning the coupling constant at lower energies is too large for perturbation theory to apply [12, 13]. Non-perturbative QCD has not been analytically solved [1], and while techniques like lattice QCD exist to obtain numerical solutions, they are not suitable for event generation for a number of reasons. The energy scales involved in particle collisions span many orders of magnitudes beyond the energy ranges possible in lattice QCD simulations. Furthermore, the computational cost of lattice QCD is very high, and not appropriate for a situation where the number of required events is in the millions, or even billions. Lattice QCD also uses Euclidean spacetime, which is a very poor approximation when relativistic effects are prominent [14].

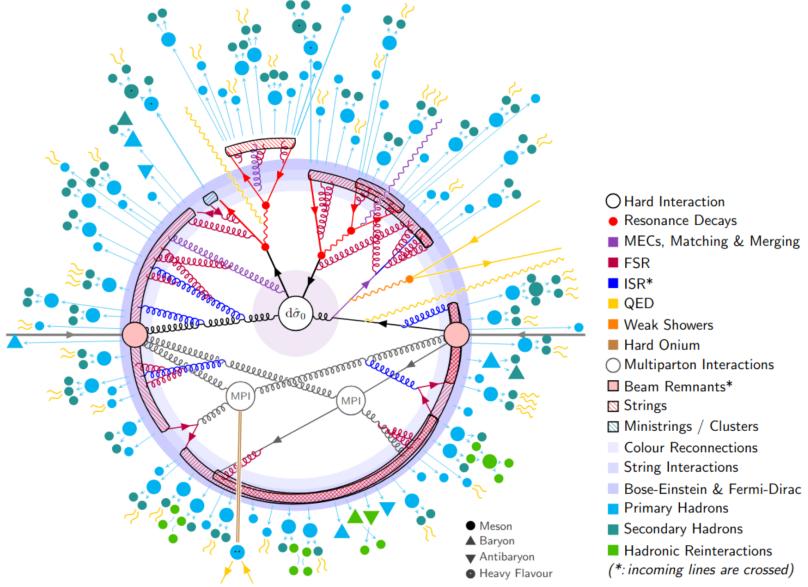


Figure 2.6: A visual representation of a $pp \rightarrow t\bar{t}$ event as generated in the PYTHIA MC event generator. The various processes involved are illustrated here with the hardest processes in the centre and the softest processes on the outer edge. Note in particular how the resultant partons from the parton shower are combined into strings, which go on to fragment into sets of outgoing hadrons (which in turn decay and emit further bremsstrahlung). Figure taken from [8].

Because of the implausibility of using an analytic or numerical solution to generate events, event generators usually use Monte Carlo (MC) techniques to sample the phase space of possible events. Such event generators exploit the fact that many aspects of particle collisions are described by random distributions. By randomly sampling from these distributions, MC event generators can generate ensembles of simulated events that closely approximate real-world events [1, 9, 10]. One of the first MC event generators to be developed was JETSET, which was based on the Field-Feynman model of hadronisation [34, 35] and played an essential role in the aforementioned discovery of the gluon [26]. In 1996, JETSET was merged into PYTHIA, which uses the Lund string model for hadronisation [8, 34]. Other commonly used event generators include HERWIG [36] and SHERPA [37], which both use the cluster model of hadronisation.

In this thesis, we focus on the PYTHIA event generator. Figure 2.6 shows a visual representation of the various processes that take place in a single $pp \rightarrow t\bar{t}$ event in PYTHIA. Here, the radial coordinate represents the hardness (momentum transfer) scale — so the hardest processes (like the initial scattering) take place in the centre, while soft processes like particle decays take place on the outer edge [8].

The hardest process in a particle collision is the hard scattering between the incoming particles (or constituent partons, in the case of pp collisions), which can be computed using perturbative techniques including Feynman diagrams and matrix elements. After this initial hard scattering, the outgoing particles will continue to emit gluon and photon radiation, which can in turn evolve into more partons (quarks and gluons) in a self-similar process known as the “parton shower”. This process continues until the energy scale is below the confinement scale of $\Lambda_{\text{QCD}} \sim 0.2 \text{ GeV}$, at which point colour confinement becomes relevant and the partons become bound within colour neutral states [8, 9].

The primary focus of this thesis is the energy scales below the confinement scale. At this scale, the parton shower is finished, leaving a large amount (often hundreds)

of outgoing quarks and gluons. These quarks and gluons will go on to combine into bound states — mesons and baryons — in a process called hadronisation. Because the physics involved is non-perturbative, hadronisation is poorly understood, and the lack of an analytic solution necessitates the use of phenomenological models to describe the underlying physics [8, 10]. A number of such models have been developed, the most commonly used of which are the Lund [15, 17] and cluster models [1, 9].

Many of the hadrons produced in hadronisation are unstable and have short lifetimes, and will therefore go on to decay into stable states. This is the final, softest step of event generation, and results in a list of outgoing particles and their four-momenta as they are observed in detectors [8, 9]. Integral to the functionality of event generators are the various factorisation theorems, which together imply that processes at different energy scales can be considered independently of one another [10, 11]. More specifically, the differential cross section for pp collisions with respect to an observable \mathcal{O} can be written as

$$\frac{d\sigma}{d\mathcal{O}} = \sum_{i,j} \int_0^1 dx_i dx_j \sum_f \int d\Phi_f f_{i/h_1}(x_i, \mu_F^2) f_{j/h_2}(x_j, \mu_F^2) \frac{d\hat{\sigma}_{ij \rightarrow f}}{d\hat{\mathcal{O}}} D_f(\hat{\mathcal{O}} \rightarrow \mathcal{O}, \mu_F^2), \quad (2.8)$$

where $\hat{\mathcal{O}}$ denotes the observable as evaluated on the final partonic state (after the parton shower, and before hadronisation). Here, f_{i/h_1} and f_{j/h_2} denote the parton densities (which describe the probabilities of interactions involving the various partons inside the colliding protons), $d\hat{\sigma}_{ij \rightarrow f}$ is the partonic cross section (describing the parton shower), and D_f describes the hadronisation of partons into final-state hadrons [10]. Each of these different processes are factorised out and different models and approximations can be used to describe each of them.

Because of the various approximations involved, MC generators involve dozens of free parameters which must be adjusted to fit experimental data. A set of such parameters is known as a “tune” [8, 9]. The JETSET generator, as well as earlier versions of PYTHIA, used the JETSET tune — the most up-to-date version of PYTHIA, PYTHIA 8.316, uses the Monash tune, which was developed at Monash University in 2014 [8, 38].

((TODO: Talk about recent PYTHIA results that raise concerns about the hadronisation??))

Having established the fundamentals of event generators and quantum chromodynamics, we can now proceed to a treatment of the Lund string model, the phenomenological model developed at Lund University in the 1980s that forms the basis of the PYTHIA generator.

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.

The Lund string model is a phenomenological model of hadronisation which models the strong colour field between colour charged particles as a classical string with constant tension, and describes hadronisation in terms of the fragmentation of these strings. In this section, we provide an overview of the Lund model, with a focus on the underlying symmetry of Lorentz covariance. This review broadly follows the textbook by Andersson [17], as well as the primary review [15], and the reader is encouraged to consult these sources for a more detailed description.

The fundamental assumption of the Lund model is that when a colour charge and anticharge are separated by a distance r , the potential takes the asymptotic form of the Cornell potential (2.4),

$$V(r) = \kappa r, \quad (2.9)$$

where $\kappa \sim 1 \text{ GeV fm}^{-1}$ is the string tension [16]. Because of the gluon-gluon interaction in QCD, this strong field is compressed into a thin flux tube, justifying the approximation of a classical string with constant tension.

At a high level, the process of hadronisation in the Lund model is illustrated in Figure 2.7. First, a $q\bar{q}$ pair is created in a particle collision, and a string forms between them. At a sufficient separation distance, the string has enough energy to break, forming a new $q\bar{q}$ pair. There are now two $q\bar{q}$ pairs, each separated by a string, each of which will continue to fragment in a self-similar manner until there is no longer enough energy for further string breaks to occur. At this point, the final state consists of a number of outgoing string pieces in the two jet directions, which are the outgoing hadrons.

In the Lund model, mesons are described as “yo-yo modes”, which consist of a quark and an antiquark connected by a string and oscillating back and forth, as depicted in Figure XX. Baryons, which consist of three quarks or three antiquarks, are also modelled as yo-yo modes where one of the string endpoints is a diquark (a grouping of two quarks or antiquarks).

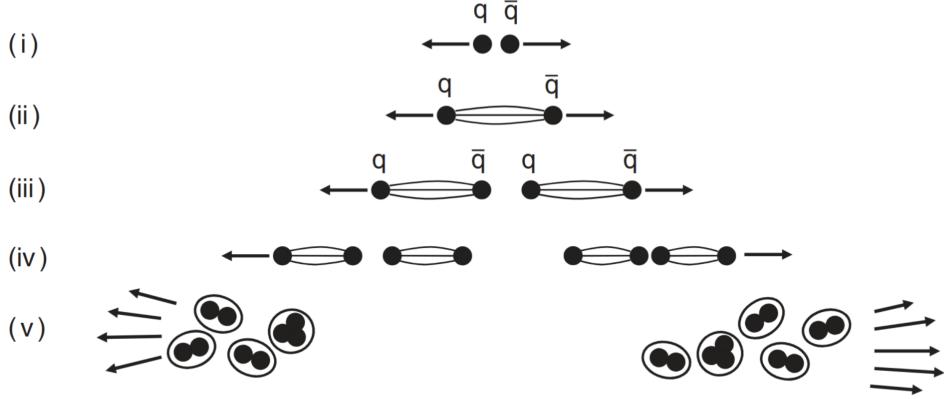


Figure 2.7: A schematic representation of hadronisation in the Lund string model. (i) A $q\bar{q}$ pair is created. (ii) As the q and \bar{q} move away from each other, the strong field between them forms a string. (iii) The string breaks, forming a new $q\bar{q}$ pair and leaving two strings. (iv) These smaller strings proceed to break in the same way, creating further $q\bar{q}$ pairs. (v) The final state consists of jets of hadrons (bound quark-antiquark states) in either direction. Figure taken from [23].

String breaking in the Lund model is modelled analogously to the Schwinger mechanism of quantum electrodynamics, where the tunnelling probability of forming a quark-antiquark pair of mass m_q and transverse momentum $p_{\perp,q}$ is given by

$$\begin{aligned} \text{Pr}(m_q, p_{\perp,q}) &\propto \exp\left(\frac{-\pi m_q^2}{\kappa}\right) \exp\left(\frac{-\pi p_{\perp,q}^2}{\kappa}\right) \\ &= \exp\left(\frac{-\pi m_{\perp,q}^2}{\kappa}\right). \end{aligned} \quad (2.10)$$

Here, $m_{\perp,q}$ is defined as the transverse momentum of the quark, where $m_{\perp,q}^2 = m_q^2 + p_{\perp,q}^2$. There is therefore a Gaussian suppression of string breaks in both transverse momentum and quark mass. Practically, PYTHIA implements the mass suppression factor using quark-specific probabilities that are tunable parameters, rather than using the Gaussian selection here, but the probabilities are approximately similar.

Equation (2.10) governs the flavour and transverse momentum selection of the quarks created in string breaks in the Lund model, which in turn determines the species and transverse momentum distributions of hadrons produced along the string in hadronisation. Before describing the longitudinal momentum selection process, we must first introduce the longitudinal lightcone coordinates and rapidity variable. In position space, the lightcone coordinates are defined as

$$z^\pm = z \pm t, \quad (2.11)$$

where x^+ describes the position of a particle along the positive worldline and x^- describes the position of a particle along the negative worldline. In momentum space, lightcone coordinates are defined as

$$p_z^\pm = E \pm p_z, \quad (2.12)$$

where p_z is the momentum in the longitudinal direction. Finally, rapidity is defined (in

the longitudinal direction) as

$$\begin{aligned}
y &= \frac{1}{2} \ln \left(\frac{1+v}{1-v} \right) \\
&= \frac{1}{2} \ln \left(\frac{E+p_z}{E-p_z} \right) \\
&= \frac{1}{2} \ln \left(\frac{p^+}{p^-} \right),
\end{aligned} \tag{2.13}$$

where v is velocity and p^\pm are the lightcone momenta as defined in Equation (2.12).

Chapter 3

String Fragmentation in PYTHIA

3.1 The PYTHIA Fragmentation Algorithm

Having established how the Lund string model describes hadronisation, we can now proceed to a description of how this model is implemented in the PYTHIA event generator. Actual events involve many hundreds of outgoing partons from the parton shower, and these will combine into various string topologies. For the purposes of this project, we will focus solely on the fragmentation of a $q\bar{q}$ string into a number of outgoing primary hadrons, prior to decays, and without any gluon kinks, loops, or junctions. As we will see, Lorentz invariance is entirely violated in PYTHIA even in this minimal situation.

In the Lund model, the Schwinger mechanism describes the relative probabilities of quark flavours in a string breakup. In turn, the probabilities of forming the various possible hadrons from a combination of quarks and antiquarks are theoretically dependent only on the various mixings of quark states that comprise the hadrons. In PYTHIA, these probabilities are implemented as tunable parameters, but the overall process is the same. Also described by the Schwinger mechanism is the transverse momenta of the quarks formed in string breakups, and therefore the transverse momenta of all final-state hadrons.

The Lund model describes the selection of hadron longitudinal momenta in terms of an iterative left-right (or right-left) ordering of string breakups, and the fragmentation function $f(z)$. This model is derived from the causal independence of breakup vertices and the resulting Lorentz covariance and self-similarity of the string fragmentation. It is important to note, however, that the Lund model provides no description of how energy and momentum are globally conserved in string fragmentation. If string fragmentation is considered iteratively, there will eventually come a point where there is insufficient energy left in the string to create a hadron with the required mass. PYTHIA enforces energy conservation by halting the string fragmentation when the remaining string energy is below a certain value (with some smearing), and then creating two final hadrons. Energy-momentum conservation and the on-shell condition fix the kinematics of these final two hadrons entirely. In the event where there is insufficient energy to create these two hadrons, the event is thrown out and string fragmentation is restarted.

This implementation of energy conservation manifestly breaks Lorentz covariance. The manner in which the kinematics of these final two hadrons are determined is different from the rest of the hadrons along the string. Furthermore, refragmenting the entire string when it is not possible to create the final two hadrons introduces bias in the species makeup of these final two. This is in contradiction with the expectation that the string fragmentation

process should be self-similar along the string. The PYTHIA documentation (and online manual, prior to updates that were part of this project) describe how this violation of Lorentz covariance is compensated for in a few ways. Firstly, instead of fragmenting left-to-right or right-to-left, hadrons are fragmented from either end with equal probability, in order to smear the position of the joining step uniformly along the string. Secondly, the parameters that control when string fragmentation is stopped are carefully chosen in order to give a flat rapidity plateau [8, 18]. It was thought that the dependence of these final two stopping parameters on the rest of the fragmentation parameters (such as the Lund a and b parameters, or the quark and hadron probabilities) was minimal, and that as such these parameters would not need to be adjusted between different PYTHIA tunes [39]. As we will see, none of these statements are true.

Algorithm 1 shows the algorithm used in PYTHIA 8.316¹ for $q\bar{q}$ string fragmentation. It takes in the two string endpoint flavours as well as the centre-of-mass energy, and populates the event record event with a set of hadrons, each with a specified species, mass, and four-momentum. The transverse momenta of the string break quarks (or antiquarks) q_{break}^{\pm} are selected according to (2.10), and the selection of their flavours is encapsulated by the procedure FLAVSEL. The details of this procedure depend on the various PYTHIA parameters that govern flavour selection, which have the prefix `StringFlav`. The dependence on the current string end quark flavour is only to determine whether the new parton should be a quark or antiquark (or a diquark or antidiquark), such that the resulting hadron is either a meson or a baryon.

The procedure COMBINE takes two quark flavours/antiflavours and combines them into a resultant hadron. Again, this procedure depends on a number of PYTHIA parameters with the prefix `StringFlav`. Neither the details of FLAVSEL or COMBINE are within the scope of this project, beyond the fact that they encapsulate the species makeup of the hadrons produced along the string. Similarly, the procedure zFRAG combines the Lund fragmentation function ((EQREF)) with other fragmentation functions used depending on the tune and quark flavours, and generally selects a z fraction for the new hadron depending on the quark flavours and hadron transverse mass m_{\perp} . The mass of each hadron is selected based on its species, usually according to a Breit-Wigner distribution, in behaviour contained within the procedure MSEL.

This z fraction specifies the fraction of the positive (or negative, depending on which end fragmentation is currently from) lightcone momentum remaining in the string. The product p^+p^- of the lightcone momenta of the string gives the remaining string mass

$$W_{\text{rem}} = \sqrt{p^+p^-}. \quad (3.1)$$

If $W_{\text{rem}} \leq W_{\text{min}}$, where W_{min} is selected according to the procedure WMIN, then string fragmentation is stopped and the final two hadrons are created. Note that if creating the most recent hadron used more energy than was available in the string, then string fragmentation will also be stopped at this step, and FINALTWO will fail automatically. If FINALTWO does fail, which can occur for this reason or if there is insufficient energy in the string to create the final two hadrons, then the entire string fragmentation process is restarted.

¹Version 8.316 of PYTHIA was released on October 3, 2025, during the writing of this thesis. It contains a number of changes, including a fix to a bug where tunes were incorrectly loaded when reading subrun settings from a file, which was uncovered as part of this project. The default value of `StringFragmentation:stopMass` was also changed, for reasons explained later in this thesis. There are no actual changes to the fundamental algorithm used in fragmentation or the final two between PYTHIA 8.316 and PYTHIA 8.315.

Algorithm 1 The default PYTHIA 8.315 algorithm for $q\bar{q}$ hadronisation.

```

procedure FRAGMENT( $E_{\text{CM}}$ ,  $\text{flav}_{+end}$ ,  $\text{flav}_{-end}$ )
  repeat
    initialise event record event
     $\vec{p}_{\perp,+end} \leftarrow \vec{0}$ 
     $\vec{p}_{\perp,-end} \leftarrow \vec{0}$ 
     $p^+ \leftarrow E_{\text{CM}}$ 
     $p^- \leftarrow E_{\text{CM}}$ 
  loop
    fromPos  $\leftarrow$  true or false with equal probability  $\triangleright \pm$  and  $\mp$  reflect this selection
     $\vec{p}_{\perp,\pm\text{break}} \leftarrow$  transverse momentum selected according to (2.10)  $\triangleright$  Select transverse momentum
     $\vec{p}_{\perp,\mp\text{break}} \leftarrow -\vec{p}_{\perp,\pm\text{break}}$ 
     $\vec{p}_{\perp} \leftarrow \vec{p}_{\perp,\pm end} + \vec{p}_{\perp,\pm\text{break}}$ 
     $\vec{p}_{\perp,\pm end} \leftarrow \vec{p}_{\perp,\mp\text{break}}$ 
    repeat  $\triangleright$  Select string break flavour and new hadron species, mass, and  $z$  fraction
       $\text{flav}_{\pm\text{break}} \leftarrow \text{FLAVSEL}(\text{flav}_{\pm end})$ 
       $\text{flav}_{\mp\text{break}} \leftarrow \text{anti}(\text{flav}_{\pm end})$ 
       $\text{species} \leftarrow \text{COMBINE}(\text{flav}_{\pm end}, \text{flav}_{\pm\text{break}})$ 
       $m \leftarrow \text{MSEL}(\text{species})$ 
       $m_{\perp} \leftarrow \sqrt{m^2 + |\vec{p}_{\perp}|^2}$ 
       $z \leftarrow \text{ZFRAG}(\text{flav}_{\pm end}, \text{flav}_{\pm\text{break}}, m_{\perp})$ 
       $\text{flav}(q_{\pm end}) \leftarrow \text{flav}(q_{\mp\text{break}})$ 
    until COMBINE succeeds
    if  $\sqrt{p^+ p^-} \leq \text{WMIN}(\text{flav}_{+end}, \text{flav}_{-end}, \text{flav}_{+\text{break}}, \text{flav}_{-\text{break}})$  then
      break
       $p_{\text{hadron}}^{\pm} \leftarrow z p^{\pm}$   $\triangleright$  Calculate longitudinal momentum
       $p^{\pm} \leftarrow (1 - z) p^{\pm}$ 
       $p_{\text{hadron}}^{\mp} \leftarrow p^{\pm} / m_{\perp}^2$ 
       $p^{\mp} \leftarrow p^{\mp} - p_{\text{hadron}}^{\mp}$ 
       $E \leftarrow p_{\text{hadron}}^+ + p_{\text{hadron}}^-$ 
       $p_z \leftarrow p_{\text{hadron}}^+ - p_{\text{hadron}}^-$ 
      append new hadron with ( $\text{species}, m, \vec{p}_{\perp}, p_z, E$ ) to event
      FINALTWO(event,  $p^+, p^-, q_{+end}, q_{-end}$ ,  $\text{species}, m, \vec{p}_{\perp}$ )
    until FINALTWO succeeds
  return event

```

Worth mentioning is a paper by Edén investigating the deviations from the Lund model area law inherent in a procedure where string ends are joined once the remaining energy is below some threshold, even when the final two hadrons are created in a Lorentz invariant manner [40]. Edén presents theoretical solutions to these deviations, but they are a separate but related problem to the issue we will see of the kinematics and hadronic chemistry of the final two being different to the hadrons along the rest of the string. As such, the problems presented by Edén will not be considered further within the scope of this thesis.

3.2 The Joining Step

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

Having established the overall structure of the PYTHIA fragmentation algorithm, we can now look more closely at how and when the final two hadrons are created. The remaining string mass at which the FINALTWO procedure fails is stipulated by the procedure `WMIN`, shown in Algorithm 2. The behaviour of `WMIN` depends on three parameters in PYTHIA. The first is `StringFragmentation:stopMass`, which allows the mean W_{\min} to be adjusted. The next is `StringFragmentation:stopNewFlav`, which also adjusts the mean of W_{\min} but with a dependence on the new string break quark mass. This allows for fragmentation to be stopped earlier in order to have sufficient energy to create heavier final hadrons (although there is still not always enough energy). Finally, `StringFragmentation:stopSmear` controls the spread of the W_{\min} values that are used. The default values of these parameters in PYTHIA are shown in Table (((REF))). Note how `stopMass` was changed from 0.8 to 1.0 between PYTHIA 6 and 8.315. This change has since been reverted in PYTHIA 8.316.

Figures (((REF, REF, and REF))) show the effects of adjusting these three parameters on the probability distribution of actual W_{rem} values encountered at the final two step, as simulated in PYTHIA 8.315 for $d\bar{d}$ hadronisation at 200 GeV (with default settings, including the Monash tune). These plots show how `stopMass` and `stopNewFlav` affect the mean of W_{\min} , while `stopSmear` affects the spread.

Algorithm 2 The procedure to determine the W_{\min} string mass threshold to stop fragmentation.

```

procedure WMIN(flav+end, flav-end, flav+break, flav-break)
     $W_{\min} \leftarrow \text{stopMass} + m_{+end} + m_{-end}$ 
     $W_{\min} \leftarrow W_{\min} + \text{stopNewFlav} * m_{+break}$ 
     $W_{\min} \leftarrow (1 + \text{stopSmear} * \text{RANDOM}(-1, 1)) * W_{\min}$ 
    return  $W_{\min}$ 
```

When the remaining string energy is below W_{\min} , the FINALTWO procedure creates the final two hadrons, as shown in Algorithm 3. In this procedure, the already-selected

string break flavour and transverse momenta are used, and hence the species, mass, and transverse momenta (species, m, \vec{p}_\perp) of one of the final two hadrons is already determined. The leftover quark flavours and transverse momenta are used to determine the properties (species_{other}, $m_{\text{other}}, \vec{p}_{\perp,\text{other}}$) of the other final hadron. Finally, the remaining string energy is used to give these final hadrons their longitudinal momenta, which are fixed by energy conservation and the on-shell requirement.

Algorithm 3 The procedure to create the final two hadrons in PYTHIA 8.316.

```

procedure FINALTWO(event,  $p^+, p^-, q_{+\text{end}}, q_{-\text{end}}$ , species,  $m, \vec{p}_\perp$ )
  if  $p^+ p^- \leq 0$  then
    return failure                                > Fails if last hadron used more energy than available
  if flav+end and flav-end are both diquarks then
    return failure                                > Cannot join two diquarks
   $\vec{p}_{\perp,\text{other}} \leftarrow \vec{p}_{\perp,+}\text{end} + \vec{p}_{\perp,-}\text{end}$ 
  repeat
    speciesother  $\leftarrow \text{COMBINE}(\text{flav}_{+\text{end}}, \text{flav}_{-\text{end}})$ 
     $m_{\text{other}} \leftarrow \text{MSEL}(\text{species}_{\text{other}})$ 
  until COMBINE succeeds
   $m_{\perp,\text{other}} \leftarrow \sqrt{m_{\text{other}}^2 + |\vec{p}_{\perp,\text{other}}|^2}$ 
  if  $\sqrt{p^+ p^-} < m_\perp + m_{\perp,\text{other}}$  then
    return failure                                > Not enough energy to create final two hadrons
   $p_z, p_{z,\text{other}}, E, E_{\text{other}} \leftarrow$  values calculated to use remaining string energy and have
  final two hadrons on-shell
  append hadron with (species,  $m, \vec{p}_\perp, p_z, E$ ) to event
  append hadron with (speciesother,  $m_{\text{other}}, \vec{p}_{\perp,\text{other}}, p_{z,\text{other}}, E_{\text{other}}$ ) to event
  return event

```

There are three possible ways for this procedure to fail, as shown in the pseudocode. Firstly, if the remaining string energy is already negative when FINALTWO is called, it will fail. This occurs if the previous hadron created in the fragmentation loop took more energy than was remaining in the string. Secondly, if the two leftover flavours for the other final hadron are both diquarks, then FINALTWO also fails, since there are no tetraquark states in PYTHIA. Finally, if there is insufficient energy in the string to create the final two hadrons with their required transverse momenta and masses, FINALTWO fails. In all of these cases, the event is discarded, and fragmentation begins again from the beginning.

3.3 Performance of the Current finalTwo Procedure

The fragmentation algorithm used in PYTHIA 8.316 manifestly violates Lorentz covariance in how it handles the final two hadrons. The longitudinal kinematics of these hadrons are determined to enforce energy conservation, without the degrees of freedom available to the rest of the hadrons produced along the string. Unless the three parameters controlling W_{\min} are tuned perfectly (if such a thing is possible — we will find it is not), the kinematics of the final two hadrons will therefore be differently distributed than the hadrons produced along the string. We will denote these hadrons produced outside of the final two and not adjacent to the endpoints as “regular” hadrons. On top of a deviation in kinematics, the fact that FINALTWO can fail introduces biases in the species composition of the final two, which we will denote the “hadrochemistry”. As such, the hadrochemistry

of the final two hadrons will be different from the hadrochemistry of regular hadrons. These deviations in the kinematics and hadrochemistry of the final two hadrons comprise the violation of Lorentz covariance introduced by this handling of energy conservation in PYTHIA’s string fragmentation algorithm.

A simple picture of the Lorentz covariance of the kinematics of hadrons along the string can be obtained from the shape of the rapidity plateau. Figure (((REF))) shows the rapidity plateaus (dN/dy distributions) for $d\bar{d}$ hadronisation across PYTHIA 6.4XX (which uses the JETSET tune), PYTHIA 8.1XX (which uses the Hoeth tune), and PYTHIA 8.315 (which uses the JETSET tune)². These results are for a rather long 200 GeV string, for which a very flat central region would be expected in the rapidity plateau. Indeed, the plateau in PYTHIA 6.4XX is entirely flat in the middle, with the “ears” on either side explainable by endpoint effects, since Lorentz covariance only applies away from the endpoints.

On the other hand, the plateaus for both PYTHIA 8 versions show a significant dip in the middle (((DO I NEED TO QUANTIFY))). To factor out effects of different PYTHIA versions being coded differently, Figure (((REF))) shows the same rapidity plateaus for these three tunes, all within PYTHIA 8.315. Notably, the plateau for the JETSET tune is not as flat. This is a result of a bug in how hadron kinematics are calculated in PYTHIA 8³ [41]. It is, however, mostly flat, and there is still a much larger dip for the Hoeth and Monash tunes. Figure (((REF))) shows how the rapidity plateau for the Monash tune varies across 20, 50, 100, and 200 GeV strings.

The presence of this dip in the rapidity plateaus of the Monash and Hoeth tunes exposes the violation of Lorentz covariance in the PYTHIA string fragmentation algorithm, and disproves the assertion that the values of the final two parameters (`stopMass`, `stopNewFlav`, and `stopSmear`) do not need to be adjusted between tunes. As mentioned earlier, `stopMass` was changed from 0.8 to 1.0 in the move from FORTRAN to C++. The rapidity plateau for a `stopMass` of 0.8 in the Monash tune is also shown in Figure (((REF))), and this change is evidently not enough to explain the violation of Lorentz covariance.

The reason for this dip can be understood in the following way. A useful measure for the longitudinal kinematics of the final two hadrons (especially in terms of how they affect the rapidity plateau) is the rapidity difference Δy_{final} between them. When the final two hadrons are created, the remaining string energy W_{rem} is used up to create the masses of the final hadrons, and the kinetic energy of the final hadrons. If W_{rem} is on average larger, then Δy_{final} would also be larger, since there would be more kinetic energy available to create a larger rapidity spacing. On the other hand, if W_{rem} is on average smaller, then Δy_{final} would also be smaller.

This behaviour is visible in Figure (((REF))), which shows how, for the Monash and JETSET tunes, the mean final rapidity difference $\overline{\Delta y}_{\text{final}}$ varies for differing values of `stopMass` (the parameter with the most direct impact on the mean of W_{rem}). As expected, $\overline{\Delta y}_{\text{final}}$ increases as `stopMass` increases and more kinetic energy is available for the final two. The horizontal dotted lines show the mean rapidity differences between regular hadrons along the string, $\overline{\Delta y}_{\text{reg}}$. The mean rapidity differences for both regular and final

²PYTHIA 8.315 is used here rather than PYTHIA 8.316 because the change to `stopMass` in PYTHIA 8.316 provides a small improvement in the flatness of the rapidity plateau.

³This bug was introduced when PYTHIA was ported to C++ from FORTRAN, between versions 6 and 8. It was discovered by Stephen Mrenna a few weeks before this thesis was submitted [41], and there was not enough time in this project to properly investigate the implications.

hadrons vary with the tune. This is because the kinematics of both regular and final hadrons are dependent on the various PYTHIA parameters that control fragmentation functions and relative quark/hadron flavour/species probabilities. The dependence is complicated, and an analytic expression is not necessary here.

Notably, the value of `stopMass` required to match the mean rapidity spacing between the final two and the rest of the string is much higher for the JETSET tune than it is for the Monash tune. Without a decrease in `stopMass` moving from the JETSET to the Monash tune (indeed, `stopMass` was actually increased), the mean rapidity spacing of the final two hadrons is much larger than the mean rapidity spacing of regular hadrons. This has the effect of “pushing out” the final two in rapidity space.

If the rank of the final two hadrons along the string was truly uniformly smeared along the string, as is intended by the random fragmentation from the left or right, then the effects of this “pushing out” would also be uniformly smeared. However, this random fragmentation implies the number of string breaks that are done from the left (or right) is binomially distributed, and therefore the rank of the final two hadrons is on average in the centre of the string. Figure (((REF))) shows a histogram of the ranks of the final two hadrons in $200\text{ GeV }d\bar{d}$ hadronisation in PYTHIA 8.316, limited specifically to events with 14 hadrons. An overall symmetrical Gaussian distribution is visible.

The dip in the rapidity plateau occurs because of these two effects. The final two hadrons are on average produced in the middle of the string, and they are on average spaced further apart in rapidity space than the rest of hadrons along the string, causing a relative underproduction of hadrons in rapidity space at the centre of the string. It is also worth noting that even in PYTHIA 6, where the rapidity plateau is flat, the shape of the distribution of Δy is different for regular and final hadrons. Figures (((REF))) and (((REF))) show how Δy is distributed for final and regular hadrons in the JETSET and Monash tunes respectively. In both tunes, there is a dip at $\Delta y = 0$ and the overall shape is obviously not the same. This indicates a violation of Lorentz covariance in hadron kinematics beyond what is visible in the rapidity plateau.

We also note here the large deviations in hadrochemistry of the final two hadrons. Figure (((REF))) shows the percentage differences in the proportions of hadron species produced in the final two compared to regular hadrons. Hadrons are ordered by mass along the horizontal axis, and antiparticles are counted as the same species as particles (since they have identical masses, and string fragmentation is symmetric under charge conjugation). The dependence of the deviations (or their direction) on mass is complex. It would be expected that the bias introduced by refragmenting the string when there is insufficient energy to create the final hadrons would lead to a bias towards lighter hadrons. This is somewhat visible — π^0 is overproduced by (((VALUE))) and Δ^0 is underproduced by (((VALUE))). However, other biases are introduced by the diquark joining failures of FINALTwo as well as effects of `stopNewFlav`. There is also an additional bias introduced since quark flavours are reselected in Algorithm 1 when COMBINE fails, but quark flavours are never reselected in Algorithm 3. As a whole, these deviations in hadrochemistry can be expressed by the sum of squared errors, given by

$$\text{SSE} = \sum_{\{i\}} (p_{i,\text{final}} - p_{i,\text{reg}})^2, \quad (3.2)$$

where $\{i\}$ runs over all hadron species, $p_{i,\text{final}}$ is the proportion of hadron species i in the final two hadrons, and $p_{i,\text{reg}}$ is the proportion of hadron species i in regular hadrons.

The values of the sum of squared errors for the JETSET and Monash tunes are shown

in Figure (((REF))), alongside the average number of failures per generated event. Evidently, for both tunes, the failure rate of `finalTwo` is high, and there are notable deviations in hadronic chemistry.

In conclusion, the algorithmic implementation of the Lund model in PYTHIA implements energy conservation by manifestly breaking Lorentz covariance. This violation of Lorentz covariance did not lead to a non-flat rapidity plateau in earlier versions of PYTHIA because in the JETSET tune the rapidity spacing was similar between final and regular hadrons. However, in the Hoeth tune (and later the Monash tune), the value of `stopMass` required for a flat plateau became much smaller, and since this value was not actually adjusted, a dip was introduced to the rapidity plateau, where it remained for over a decade. Aside from the deviation in kinematics, the hadrochemistry of the final hadrons is also distributed differently than regular hadrons across all tunes because of the high failure rate of `finalTwo`. This failure rate also causes difficulties in statistical analysis and machine learning using PYTHIA (((CITATIONS))).

As a whole, the violation of Lorentz covariance in PYTHIA hadronisation is quite severe, and represents quite a large deviation from the theoretical predictions of the Lund model. As an immediate solution, `stopMass` was reverted to 0.8 in PYTHIA 8.316, which provides a slightly flatter rapidity plateau. The PYTHIA online manual was also updated to reflect this new understanding. The rest of this thesis explores techniques for restoring or preserving Lorentz covariance in PYTHIA, through tuning the W_{\min} parameters, introducing a new parameter, and redesigning the fragmentation model.

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. Figure: Plots of SSEs across parameters and tunes.

Having established the extent and causes of the breaking of Lorentz covariance in PYTHIA hadronisation, we can now move on to exploring solutions to this problem. It would be easiest if we could at the very least tune for a flat rapidity plateau (and ideally restore Lorentz covariance) using the three parameters that control when the final two hadrons are created — `stopMass`, `stopNewFlav`, and `stopSmear`. Figure (((REF))) shows that, in the Monash tune, a flat rapidity plateau can be obtained by decreasing `stopMass` to approximately (((VALUE))), and that the deviation in Δy increases linearly with `stopMass`. Figure (((REF))) shows the rapidity plateau for 200 GeV $d\bar{d}$ hadronisation for varying values of `stopMass`, and the effect on the flatness is clear, with a sweet spot shown at the value of (((VALUE))) where the Δy means are equal across final and regular hadrons.

The issue with simply turning down `stopMass` (or reducing W_{\min} in general) is that the bias towards lighter hadrons in the final two hadrochemistry is significantly increased for smaller W_{\min} , and in the extreme case of `stopMass` = 0.05 failures are common enough to cause noticeable slowdowns and pions make up (((VALUE)))% of all final hadrons. Figure (((REF))) shows how the hadrochemistry SSE varies with `stopMass` — worse, there is no value that brings it close to zero.

Similar to `stopMass`, `stopNewFlav` affects the mean of W_{\min} , but with a more favourable effect on hadronic chemistry since this parameter accounts somewhat for the mass of the

final hadrons that will be produced in its effect on W_{\min} . Figure (((REF))) shows how the hadrochemistry SSE varies with `stopNewFlav`. (((COMMENT))).

Since `stopSmear` primarily affects the spread of W_{\min} , its main utility is in aligning the shape of the Δy distribution between the final and regular hadrons. Figure (((REF))) shows, however, how these distributions are inherently shaped differently beyond just a dilation. Through experimentation, it was found that the values (((VALUES))) give a good improvement on the flatness of the rapidity plateau without a significant worsening of hadrochemistry SSE, as well in Figure (((REF))) which compares the resulting rapidity plateau and SSE with the PYTHIA 8.316 defaults.

In general, for the Monash tune, and a wide range of potential tunes, there is no way to restore Lorentz covariance using the existing `finalTwo` parameters in PYTHIA. Even a flat rapidity plateau cannot be obtained without extreme violations in Lorentz covariance of hadrochemistry along the string. Fundamentally, the issue is that there is only one degree of freedom — the mean of W_{\min} and hence Δy — with which to adjust for a flat rapidity plateau. Increasing the mean improves the hadronic chemistry but worsens the dip in the rapidity plateau, and vice versa.

The next section describes a new parameter, `probRevertFinal`, which provides some additional freedom in parameter space to adjust for Lorentz covariance.

4.2 The `probRevertFinal` parameter

1. Introduce the `probRevertFinal` 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.

The next option for restoring Lorentz covariance would ideally provide a way to improve hadrochemistry without a significant affect on the rapidity plateau. One way to do this is to reduce the bias in the final two hadrochemistry when there is insufficient energy to create the final two hadrons. By default, PYTHIA reframents the entire string in this case. Instead of reframenting from the start, we could instead undo the previous string break from the end we are fragmenting from, and then try to create the final two hadrons again. The parameter `probRevertFinal` was implemented in PYTHIA 8.316, which gives a probability that the final string break will be reverted when `finalTwo` fails because of insufficient string mass. Algorithm (((REF))) shows how the `FRAGMENT` procedure is adjusted to implement this parameter — changes are highlighted in red, and some lines are collapsed for brevity.

When one of the final hadrons is reselected from its new constituent flavours after the string break is reverted, the selection between a light and heavy (pseudoscalar vs. scalar, or vector vs. axial vector) hadron spin is flipped with a probability $Pr(\text{flip})$, given by

$$Pr(\text{flip}) = \frac{Pr(\text{current heavy spin} | \text{current flavour})}{Pr(\text{previous heavy spin} | \text{previous flavour})}. \quad (4.1)$$

Figure (((REF))) illustrates how this equation has the effect of (as much as possible) ensuring the “same” random number is used to select both spins. Algorithm (((REF))) has high level pseudocode of the COMBINECONDITIONAL procedure which takes in the old flavour and spin and selects a hadron species accordingly.

Figure (((REF))) shows how the rapidity plateau and hadrochemistry SSE vary as `probRevertFinal` varies from 0 to 1. As expected, increasing `probRevertFinal` decreases the SSE, and while there is some worsening of the rapidity plateau (since reverting the final break on average increases the energy available for the final two), the effect is quite mild. Through experimentation with all four parameters, the combination (((VALUES))) was found to give a good improvement in the rapidity plateau with a lessened effect on SSE compared to the same values with `probRevertFinal = 0`, as shown in Figure (((REF))).

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.

The `probRevertFinal` parameter successfully provides more parameter space and allows for Lorentz covariance to be restored further. However, the rapidity plateau in Figure (((REF))) is not as flat as that of the JETSET tune in Figure (((REF))), and the hadrochemistry SSE is enough for noticeable distortions in particle spectra across the string. The shape of the Δy distribution is also different between regular and final hadrons. Finally, the manifest breaking of Lorentz covariance in the structure of the algorithm means that restoring Lorentz covariance by obtaining a flat rapidity plateau and a negligible hadrochemistry SSE is impossible without using a different model to conserve energy in string fragmentation.

One might assume the thesis would end here by describing the impact of this research and the possibilities for further investigations. The attuned reader, though, would realise that this feels like a narratively unsatisfying ending to the paper. The reader can rest assured that the humble author is goated. In the next section, we describe a new model of string fragmentation that enforces energy conservation in a Lorentz covariant way, restoring Lorentz covariance to the PYTHIA hadronisation algorithm.

Chapter 5

The Slinky Model of String Fragmentation

5.1 The Slinky Model

To motivate this new model, it is useful to take a step back to the basics of the Lund model — what properties we expect string fragmentation to have, what these say about how the algorithm can be structured, and the nature of the actual physical process of hadronisation. The Lund model describes $q\bar{q}$ string fragmentation in terms of causally independent string breaks via the Schwinger mechanism on a string worldsheet. This property combined with the possibility of considering string breaking iteratively and left-right symmetry imply the form of the Lund symmetric fragmentation function in Equation (2.10). In the limit where the remaining string mass is infinite, the longitudinal momenta (or the Lorentz invariant rapidity spacings) of hadrons can be selected according to this fragmentation function, and this provides a Lorentz covariant fragmentation.

More fundamentally, hadronisation is the evolution of an incoming partonic $q\bar{q}$ state into an outgoing state consisting of a set of hadrons, subject to the constraint of energy conservation and the requirement that the evolution into the final state is Lorentz covariant. Although the breakup vertices are causally independent, they are also quantum mechanically correlated in the enforcement of energy conservation. If we observe that the first rank hadron from the positive end takes the majority of the string mass, then the rest of the hadrons must have small masses or kinetic energies. As such, the effect of the energy conservation constraint on the fragmentation process should be global. A Lorentz covariant model for string fragmentation should therefore enforce energy conservation in



Figure 5.1: Slinky.

a way that is both global and Lorentz covariant. Physically, this would correspond to the quantum entanglement of the breakup vertices, alongside their causal disconnection. As a phenomenological model, we cannot expect to accurately model quantum entanglement effects, but this at least provides a justification for the model exhibiting global behaviour. At least metaphorically, the partonic state evolves “simultaneously” to the final state, in the sense that while nothing that occurs at any point along the string has any causal or time-ordered influence on any other point (so any ordering is theoretically appropriate), there are also global considerations that forbid a strictly iterative model.

The accordion model of string fragmentation preserves Lorentz covariance by allowing energy conservation to be violated and sampling rapidity spacings for the final two hadrons based on selections of trial z fractions from fragmentation functions, so that the rapidity spacings and particle spectra are identically distributed along the string. Energy conservation is then restored by a global Lorentz covariant rapidity dilation $y \rightarrow ky$, where a solution to k is obtained numerically and negligible corrections are done to ensure energy conservation is exact. Such a rapidity dilation preserves a flat rapidity plateau by stretching or shrinking all rapidity differences like a Slinky, as shown in Figure 5.1. Lorentz covariance is therefore maintained in fragmentation.

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 topologies - 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] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001.
- [2] F. Gross et al., *50 Years of Quantum Chromodynamics*, *Eur. Phys. J. C* **83** (2023) 1125 [[2212.11107](#)].
- [3] M.D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press (3, 2014).
- [4] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [[1207.7214](#)].
- [5] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [[1207.7235](#)].
- [6] R.K. Ellis, W.J. Stirling and B.R. Webber, *QCD and collider physics*, vol. 8, Cambridge University Press (2, 2011), [10.1017/CBO9780511628788](#).
- [7] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001.
- [8] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8 [[2203.11601](#)].
- [9] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145 [[1101.2599](#)].
- [10] P. Skands, *Introduction to QCD*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pp. 341–420, 2013, DOI [[1207.2389](#)].
- [11] J.C. Collins and D.E. Soper, *Parton Distribution and Decay Functions*, *Nucl. Phys. B* **194** (1982) 445.
- [12] D.J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories - I*, *Phys. Rev. D* **8** (1973) 3633.
- [13] H.D. Politzer, *Asymptotic Freedom: An Approach to Strong Interactions*, *Phys. Rept.* **14** (1974) 129.
- [14] G.P. Salam, *Elements of QCD for hadron colliders*, in *2009 European School of High-Energy Physics*, 11, 2010 [[1011.5131](#)].

- [15] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31.
- [16] G.S. Bali and K. Schilling, *Static quark - anti-quark potential: Scaling behavior and finite size effects in SU(3) lattice gauge theory*, *Phys. Rev. D* **46** (1992) 2636.
- [17] B. Andersson, *The Lund Model*, vol. 7, Cambridge University Press (1998), [10.1017/9781009401296](https://doi.org/10.1017/9781009401296).
- [18] T. Sjostrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [[hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175)].
- [19] E. Rutherford, *The scattering of alpha and beta particles by matter and the structure of the atom*, *Phil. Mag. Ser. 6* **21** (1911) 669.
- [20] J. Chadwick, *The Existence of a Neutron*, *Proc. Roy. Soc. Lond. A* **136** (1932) 692.
- [21] P.A.M. Dirac, *The quantum theory of the electron*, *Proc. Roy. Soc. Lond. A* **117** (1928) 610.
- [22] C.D. Anderson, *The Apparent Existence of Easily Deflectable Positives*, *Science* **76** (1932) 238.
- [23] M. Thomson, *Modern particle physics*, Cambridge University Press, New York (10, 2013), [10.1017/CBO9781139525367](https://doi.org/10.1017/CBO9781139525367).
- [24] Y. Ne'eman, *Derivation of strong interactions from a gauge invariance*, *Nucl. Phys.* **26** (1961) 222.
- [25] M. Gell-Mann, *Symmetries of baryons and mesons*, *Phys. Rev.* **125** (1962) 1067.
- [26] JADE collaboration, *Observation of Planar Three Jet Events in e+ e- Annihilation and Evidence for Gluon Bremsstrahlung*, *Phys. Lett. B* **91** (1980) 142.
- [27] M.E. Peskin and D.V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA (1995), [10.1201/9780429503559](https://doi.org/10.1201/9780429503559).
- [28] K. Hubner, *Design and construction of the ISR*, in *40th Anniversary of the First Proton-Proton Collisions in the CERN Intersecting Storage Rings (ISR)*, 6, 2012 [[1206.3948](https://arxiv.org/abs/1206.3948)].
- [29] S. Erhan, W.S. Lockman, T. Meyer, J. Rander, P. Schlein, R. Webb et al., *Hyperon production in pp interactions at $\sqrt{s} = 53$ and 62 GeV*, *Phys. Lett. B* **85** (1979) 447.
- [30] “Design Report Tevatron 1 project.”
<https://lss.fnal.gov/archive/design/fermilab-design-1984-01.pdf>, 1984.
- [31] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) 2632 [[hep-ex/9503003](https://arxiv.org/abs/hep-ex/9503003)].
- [32] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626 [[hep-ex/9503002](https://arxiv.org/abs/hep-ex/9503002)].

- [33] CERN, “Atlas images gallery.”
<https://home.cern/resources/image/experiments/atlas-images-gallery>, 2011.
- [34] T. Sjöstrand, *The PYTHIA Event Generator: Past, Present and Future*, *Comput. Phys. Commun.* **246** (2020) 106910 [[1907.09874](#)].
- [35] T. Sjostrand, *The Lund Monte Carlo for Jet Fragmentation and e+ e- Physics: Jetset Version 6.2*, *Comput. Phys. Commun.* **39** (1986) 347.
- [36] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010 [[hep-ph/0011363](#)].
- [37] SHERPA collaboration, *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034 [[1905.09127](#)].
- [38] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J. C* **74** (2014) 3024 [[1404.5630](#)].
- [39] P. collaboration, “Pythia 8 online manual.”
<https://pythia.org/manuals/pythia8315>Welcome.html>, 2025.
- [40] P. Eden, *On energy conservation in Lund string fragmentation*, *JHEP* **05** (2000) 029 [[hep-ph/0004132](#)].
- [41] S. Mrenna. private communication.