Chapter 6

Jets at Colliders

Simone Marzani

Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Via Dodecaneso 33, 16146, Italy
simone.marzani@ge.infn.it

6.1 A Brief Introduction

Collisions at very high energies produce a plethora of particles that are collected by detectors surrounding the interaction point. In particular, because of the conspicuous magnitude of the coupling α_s, strongly interacting particles are abundantly produced in every such collision. This occurs for both lepton (e.g., e^+e^-) colliders and for experiments in which at least one hadron is brought to collision, such as, for instance, proton–proton (pp) collisions at the CERN Large Hadron Collider (LHC) or lepton–proton (ep) or, more generically, lepton–hadron (eh) collisions, such as the ones that will be investigated by the future BNL Electron Ion Collider (EIC).

Studies of hadronic final states in e^+e^- collisions have been instrumental to establish Quantum-Chromo Dynamics (QCD) as the theory of strong interactions. This is because the initial-state leptons carry no color charge and, consequently, QCD radiation can only be produced by the final state. More complex environments are found in ep and pp collisions because QCD radiation can also originate from the hadronic initial states. In this context, past ep experiments allowed us to reach a deep understanding of the structure of the proton in terms of parton distribution functions.

 https://doi.org/10.1142/9789819801107_0006

The successful physics program of the LHC, including the study of strong interactions at unprecedented energies, builds upon the knowledge acquired at previous particle colliders. Even more challenging is the study of collisions involving heavy ions, which allow us to probe new regions of the QCD phase diagrams, such as the color glass condensate and the quark-gluon plasma.

Studies of strong interactions in particle collisions come with enormous theoretical and experimental challenges. From the theory point of you, we can exploit a fundamental property of QCD, called asymptotic freedom, to perform perturbative calculations. In this framework, valid at high energies, i.e., far above the characteristic energy scale of hadron formation, typically denote by Λ or taken to be of the order of hadron masses, i.e., 1 GeV, the theory is weakly coupled and quarks and gluons, collectively referred to as partons, are good degrees of freedom. Thus, at high energy, QCD processes can be described in terms of scattering and production of these states.

Quarks and gluons cannot be directly detected in experimental apparatuses. We can imagine highly energetic quarks and gluons, which are produced in the collision, or from the decay of a high-mass intermediate particle, starting radiating further partons, thus reducing their energy. This process of successive splittings, usually referred to as *parton shower*, continues until one reaches the characteristic scale of hadron formation Λ. In this regime, QCD is no longer perturbative and, because of confinement, quarks and gluons form hadrons. Although some first-principle understanding of the hadronization process does exist, we often rely on phenomenological models implemented in Monte Carlo event generators to describe the transition from partons to hadrons.

One peculiar feature of parton showers is that, because of the structure of QCD matrix elements, QCD splittings preferentially happen at small angles, giving rise to a series of collimated quarks and gluons. This characteristic is not washed out by the hadronization process and hence hadrons resulting from high-energy interactions are not uniformly distributed in the detector but rather appear in a few collimated sprays that are named *jets*. This peculiar feature can be exploited to perform meaningful comparisons between theoretical calculations and experimental data. This is extremely useful because calculations in perturbative QCD feature a few final-state partons in fixed-order calculations, or a few tens of partons after the showering process, while a hadron-level event contains hundreds, if not thousands, of particles, the dynamics of which would be very difficult to individually determine. In some sense, jets constitute a portal between theory land and the real world.

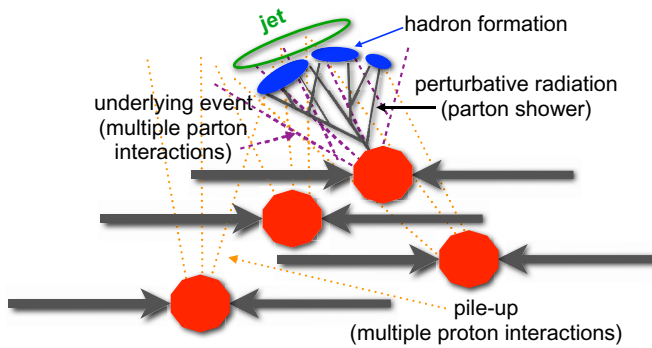

Figure 6.1. A cartoon representing jet formation in proton–proton collisions, such as the ones happening at the LHC. On top of highly energetic phenomena, which we can describe using perturbative field theory, jet formation is affected by soft, and hence non-perturbative, QCD effects, such as hadronization, the underlying event and pile-up.

Despite the remarkably successful application of perturbative calculations to describe collider phenomenology, we should bear in mind that actual collision events are much more complicated, as depicted in Fig. 6.1. Every time those two protons collide, multiple (semi-hard) partonic interactions can happen, giving rise to more hadronic activity, denoted by the term *underlying event*. Furthermore, in actual colliders, bunches of protons are brought to collisions and so multiple proton–proton interactions per bunch crossing can happen. This produces rather uniform soft radiation, usually referred to as *pile-up*. This is an unwanted consequence of the desire for higher and higher luminosity, which is necessary in order to probe rare events and pile-up mitigation is a very active area of research [1].

6.2 The Concept of Jets

The parton-shower picture described above, which may appear hand-wavy, finds its foundation on the factorization properties of QCD. However, it does simplify several aspects because it is essentially based on a semiclassical approximation of quantum field theory. If higher-order corrections are included, the concept of parton becomes ill-defined because both real emissions and virtual contributions must be taken into account. We discuss some of the issues we encounter when doing higher-order calculations in Section 6.2.1.1.

From a more practical point of view, we immediately realize that the concept of jet is somewhat ambiguous. Assigning two particles (or two

partons in perturbative calculations) to the same jet, or to different ones, has some degree of arbitrariness because it depends on what we mean by two objects being collimated. In a more precise way, when talking about jets, we must introduce a resolution scale that allows us to separate objects in an event. This concept can be formalized by saying that we have to introduce a *jet definition*, i.e., a procedure that dictates how to reconstruct jets from the set of final-state hadrons (or partons) in a collision event. Jet definitions usually contain two parts:

- The *jet algorithm* is the set of rules that we must follow in order to map the set of final-state particles into jets. Most jet algorithms can be applied in an *inclusive* way, whereby the number of resulting jets is not fixed *a priori*, or in an *exclusive* mode, whereby an event is mapped into a specified number of jets. Jet algorithms feature free resolution parameters that are set by the user according to the physics case they are interested in. For example, a parameter that is present in most jet definitions for LHC studies is the jet radius, which sets the jet resolution scale in the azimuth-rapidity plane.
- The *recombination scheme* specifies how the kinematic properties of a jet, e.g., the jet four-momentum or its axis, are derived from the kinematics of the jet constituents. In most applications, the so-called *E-scheme* is employed. In this approach, the jet momentum is simply the vectorial sum of the four momenta of its constituents and the jet axis is aligned with the jet momentum. Although this choice does appear as the most natural one, specific applications may require different recipes. For instance, in the context of jet substructure studies, the so-called *Winner-Take-All* (WTA) [2] scheme is sometimes employed. In this scheme, the result of the recombination of two particles has the rapidity, azimuth, and mass of the particle with the larger transverse momentum, while the transverse momenta themselves are summed up. As a consequence, in the WTA scheme, the jet axis always lies along the direction of the hardest particle in the jet.

The design and the implementation of jet definitions are still an area of active research and a detailed discussion of the several algorithms that have been proposed in past few decades goes beyond the scope of this chapter.[1] Here, we limit ourselves to discuss and highlight, from both theoretical

[1] For an extensive review on jet definitions, we highly recommend the reading of Ref. [3].

and experimental viewpoints, the features of two main categories of jet definitions: the ones that feature *cone algorithms* and the ones based on *sequential recombination*. Before doing so, let us discuss the basic properties that jet definitions should respect.

6.2.1 *What experimenters want... what theorists want...*

Jets live at the boundary between theoretical and experimental high-energy physics. Thus, their definition should be meaningful both when applied to observable particles without considering detector effects (e.g., truth level) and also when applied to real data, which is to say detector signals. These signals include things like tracks left by charged particles or energy deposits in calorimeter cells. At the same time, the very same jet definitions should be used by theorists when performing perturbative calculations involving quarks, gluons, loops, and all that. In the 1990s, a group of theorists and Tevatron experimentalists formulated what is known as the Snowmass accord [5]. To date, this document represents the minimal set of fundamental criteria that any jet algorithm should satisfy:

(1) simple to implement in an experimental analysis;
(2) simple to implement in theoretical calculations;
(3) defined at any order of perturbation theory;
(4) yields finite cross-sections at any order of perturbation theory;
(5) yields cross-sections and distributions that are relatively insensitive to hadronization.

The first point of the list is the main demand that arises from experimental considerations. The information gathered from the various detector components, such as the trackers, the electromagnetic and hadronic calorimeters, and the muon spectrometer allows us to obtain a good picture of the types of particles that are produced in a given collision. However, jet reconstruction, often referred to as *clustering*, is typically performed at an early stage, when particle identification is still incomplete. In the first two runs of the LHC, ATLAS and CMS used different strategies to define jets. The former predominantly exploited topological clusters, so-called topoclusters, which are based on information obtained from the calorimeters, while the latter used so-called particle flow objects, which combine information from the tracker and the calorimeter to build a coherent single object. All major experimental collaborations have

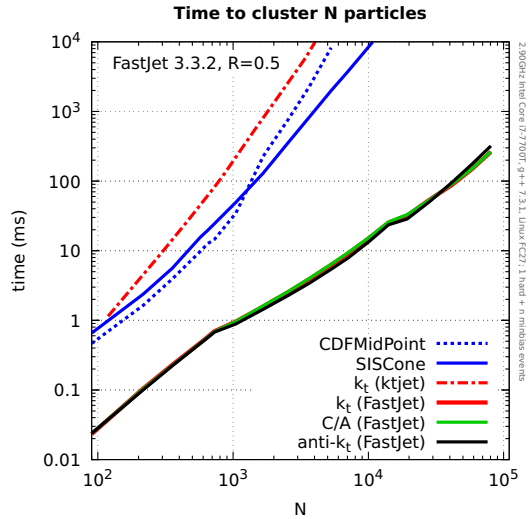

Figure 6.2. In this plot, the average clustering time for a set of representative algorithms is shown as a function of the event multiplicity N. Curves are obtained with either the algorithm original implementation or with the FastJet.

Source: Figure taken from Ref. [4].

dedicated groups actively working on the performance of jet definitions. For instance, the ATLAS collaboration has introduced for LHC Run 3 new Unified Flow Objects (UFOs) that aim to maximize performance across many orders of magnitude in the jet transverse momentum by combing the virtues of calorimetric and particle-flow approaches [6].

Once the inputs have been defined, jets must be reconstructed. Currently, the standard computer program for doing this step is `FastJet`[2] [7,8], used by both the experimental and theoretical communities. `FastJet` employs different strategies, including ideas from computational geometry, in order to speed up jet reconstruction. To illustrate this point, the plot in Fig. 6.2 shows the average time it takes to cluster an event with N particles into jets, for a few representative algorithms. There is a noticeable difference between the original `ktjet` implementation [9] of the k_t algorithm, which was deemed too slow, and the `FastJet` implementation which is faster by 2–3 orders of magnitude in the region relevant for phenomenology.

[2]See also http://fastjet.fr.

Conditions (2), (3), and (4) come from the theorists. We have already discussed the second one, namely, one should be able to use quarks and gluons as inputs to the jet algorithms. Conditions (3) and (4) have instead to do with *InfraRed and Collinear (IRC) safety*, a concept so important that deserves a separate discussion. We dive into this topic in Section 6.2.1.1, but before doing that, let us briefly comment on condition (5). Admittedly, this point is less precise and somewhat more subjective. Since jets are supposed to capture the "hard partons in an event," we should hope that observables built from jet quantities are as little sensitive as possible to non-perturbative effects like hadronization, the underlying event, and pile-up. Furthermore, jets should not be too sensitive to detector effects so that corrections deriving from moving from detector-level to particle-level quantities, the so-called unfolding procedure, remain under control.

6.2.1.1 *A detour about IRC safety*

Following the Snowmass accord, we work with jet algorithms that are defined and yield finite cross-sections at any order of perturbation theory. In order to better understand the origin of this request, let us work through a simple example that initially does not involve jets. We consider the calculation of the total cross-section for the production of hadrons in e^+e^- collisions. In this discussion, we are going to mostly quote results of perturbative calculations and interpret them with physical arguments. We encourage the interested readers to actually perform such calculations, following one of the many beautiful textbooks about high-energy applications of perturbative quantum field theory.

As we have already mentioned, hadrons are bound states that cannot be described in perturbation theory. However, hadron formation happens at an energy scale that is much smaller than the scale of the hard interaction. For instance, at LEP1, leptons were brought to collision at an energy Q equal to the Z boson mass, which is two orders of magnitude bigger than the hadron formation scale Λ. We can separate, we say factorize, the production cross-section as follows:

$$d\sigma_{e^+e^-\to\text{hadrons}} = \sum_{\{i\}} d\sigma_{e^+e^-\to\{i\}} \times dF_{\{i\}\to\text{hadrons}} + \mathcal{O}\left(\frac{\Lambda^2}{Q^2}\right), \qquad (6.1)$$

where the $\{i\}$ sum runs over all partonic state that are possible at a given perturbative order. Thus, up to power corrections that are small at very high-energy colliders, we can separate a partonic cross-section,

which we can compute in perturbation theory, from a non-perturbative contribution that describes the fragmentation of partons into hadrons. Theorists usually focus on the former, computing higher and higher orders in the perturbative expansion. The calculation of the lowest order contribution is particularly straightforward. We only have to consider two Feynman diagrams, corresponding to the processes:

$$e^+e^- \to Z/\gamma^* \to q\bar{q}. \tag{6.2}$$

Note that the cross-section for this process at leading order (LO), or Born-level, only involves electroweak couplings. Its expression is a bit cumbersome because it involves the photon contribution, the Z one, and their interference. At energies much lower than the Z mass, but still larger enough than Λ, so that we can trust our factorized formula in Eq. (6.1), the photon contribution dominates and the inclusive, i.e., after integration over the phase space, Born cross-section has a particularly simple form:

$$\sigma_0^{\gamma^*} = \frac{4\pi\alpha^2}{3Q^2} N_C \sum_f Q_f^2, \tag{6.3}$$

where α is the fine-structure constant, the sum is over the quark flavors that are accessible at the energy Q considered here, Q_f is the fractional quark electric charge, and $N_C = 3$ is the number of colors in QCD.

We are now interested in the next-to-leading order (NLO) corrections, i.e., the $\mathcal{O}(\alpha_s)$ contributions, to the partonic cross-section. We have to consider two types of contributions. First, we can dress the LO diagram with loops involving quarks and gluons. At $\mathcal{O}(\alpha_s)$, we have only one such diagram, which is depicted in Fig. 6.3(2). Second, we should remember that we are ultimately interested in the inclusive cross-section for the production of hadrons, and according to Eq. (6.1), we must consider all possible partonic states $\{i\}$. At $\mathcal{O}(\alpha_s)$, this means that we should also consider the emission of a real gluon, as shown in Figs. 6.3(3) and 6.3(4):

$$\sigma_{\text{NLO}} = \int d\Phi_2(k_1, k_2) \mid \mathcal{M}_0 + \mathcal{M}_{\text{loop}} \mid^2 + \int d\Phi_3(k_1, k_2, k_3) \mid \mathcal{M}_{\text{real}} \mid^2, \tag{6.4}$$

where $d\Phi_n$ is the n-body Lorentz-invariant phase space. It goes beyond the scope of this presentation to describe the details of the calculation. Here, we simply state that both the integral over the loop momentum in the virtual amplitude and the one over the phase space of the real gluon

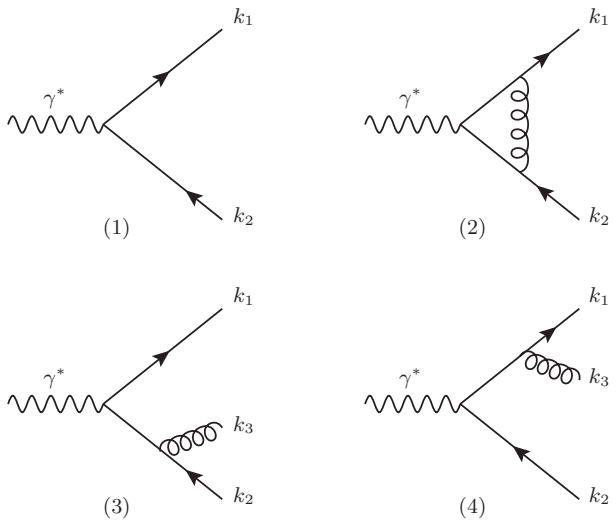

Figure 6.3. Feynman diagrams contributing to the cross-section of $e^+e^- \to q\bar{q}$ up to NLO. Diagram (1) gives the Born-level contribution, (2) the one-loop correction, and (3) and (4) describe the real-emission contribution.

are divergent. In order to understand the origin of these singularities, it is convenient to inspect the kinematics of the real emission. We find that the real emission contribution is singular when the gluon is either soft, i.e., with vanishing energy, or its three momentum becomes collinear to the directions of either the quark or the antiquark. This is a very general feature of massless gauge theories: infrared and collinear singularities arise when massless gauge bosons become soft or when two massless particles become collinear. It is interesting to note that in these singular limits, the kinematics of the three-body final state reduces to one of the two-body final states, i.e., one of the Born contribution and of the loop correction. This makes sense because we cannot resolve infinitely soft particles or two particles that are too close in angle. Thus, it is at least conceivable that the singular behavior of the real contribution may conspire with one of the loop diagrams, giving a finite result.[3] It is useful to rewrite the cross-section

[3]Loop amplitudes can also exhibit singularities in the ultra-violet, which can be dealt with the renormalization procedure.

separating out the divergent contributions:

$$\sigma_{\text{NLO}} = \int d\Phi_2(k_1,k_2)\,|\,\mathcal{M}_0 + \mathcal{M}_{\text{loop-finite}}\,|^2 + \int d\Phi_3(k_1,k_2,k_3)\,|\,\mathcal{M}_{\text{real-hard}}\,|^2 + \int d\Phi_2(k_1,k_2)\left[2\text{Re}\,\mathcal{M}_0^*\mathcal{M}_{\text{loop-div}} + \int d\Phi_1(k_3)\mid \mathcal{M}_{\text{real-IRC}}\mid^2\right] + \mathcal{O}\left(\alpha_s^2\right), \tag{6.5}$$

where we have exploited the factorization properties of phase-space integrals. The explicit computation of the problematic contributions reveals that

$$2\text{Re}\,\mathcal{M}_0^*\mathcal{M}_{\text{loop-div}} = -\int d\Phi_1(k_3)\mid \mathcal{M}_{\text{real-IRC}}\mid^2, \tag{6.6}$$

Thus, IRC singularities cancel and the cross-section that describes the process $e^+e^- \to$ hadrons can be safely computed in perturbation theory by considering the corresponding partonic process. The cross-section up to NLO reads

$$\sigma_{\text{NLO}} = \sigma_0\left(1+\frac{\alpha_s}{\pi}\right), \tag{6.7}$$

where σ_0 is the generalization of Eq. (6.3) that also includes the Z contribution and the Z/γ^* interference.

This important result is a manifestation of rather general theorems: the Bloch–Nordsieck [10] and Kinoshita–Lee–Nauenberg [11, 12] theorems state that observable transition probabilities are free of IRC singularities. However, as it stands, it leads to rather boring phenomenology because it holds for the inclusive cross-section. It is therefore interesting to investigate whether it can be generalized to more exclusive processes, such as the production of jets. In order to study this, we introduce a measurement function $J_r(\{k_i\})$ that takes as inputs the momenta of the final-state partons k_i and maps them into a set of jet momenta, with some resolution parameters r. More generally, we can consider measurement functions J_r that define physical observables, also characterized by one or more resolution scales r, with jets being a particular example. Let us go back to our e^+e^- example at $\mathcal{O}(\alpha_s)$ and consider the map J_r that produces two jets. Following the discussion about the inclusive cross-section, we write the

2-jet cross-section separating out the divergent contributions:

$$\begin{aligned}\sigma_{2\text{ jets}} = &\int d\Phi_2(k_1,k_2) \mid \mathcal{M}_0 + \mathcal{M}_{\text{loop-finite}} \mid^2 J_r(k_1,k_2) \\ &+ \int d\Phi_3(k_1,k_2,k_3) \mid \mathcal{M}_{\text{real-hard}} \mid^2 J_r(k_1,k_2,k_3) \\ &+ \int d\Phi_2(k_1,k_2) \Bigg[2\text{Re}\, \mathcal{M}_0^* \mathcal{M}_{\text{loop-div}}\, J_r(k_1,k_2) \\ &+ \int d\Phi_1(k_3) \mid \mathcal{M}_{\text{real-IRC}} \mid^2 J_r(k_1,k_2,k_3) \Bigg].\end{aligned} \tag{6.8}$$

Thus, thanks to Eq. (6.6), we obtain a finite 2-jet cross-section, provided that the 3-particle measurement function reduces to the 2-particle one, in the limit in which k_3 becomes soft and/or collinear to the fermions' directions. If the measurement function has this property, we say that the observable (or the jet algorithm) is Infra-Red and Collinear (IRC) safe and its cross-section can be computed in perturbation theory. Clearly, not all possible measurement functions J_r are IRC safe. For instance, a measurement function that simply counts the number of partons, irrespectively of their momenta, does not respect this criterion. Indeed, particle multiplicity, i.e., an observable that simply counts the number of particles in a region of phase space, is not IRC safe.

Different definitions of IRC safety exist in the literature. Here, we have adopted the one in Ref. [13] that ensures cancelation of IRC singularities to any order in perturbation theory:

$$J_r\left(k_1 \ldots, k_i, k_j, \ldots, k_n\right) \longrightarrow J_r\left(k_1 \ldots, k_i + k_j, \ldots, k_n\right) \qquad \text{if } k_i \parallel k_j, \tag{6.9}$$

$$J_r\left(k_1 \ldots, k_i, \ldots, k_n\right) \longrightarrow J_r\left(k_1 \ldots, k_{i-1}, \ldots, k_{i+1} \ldots, k_n\right) \quad \text{if } k_i \to 0. \tag{6.10}$$

IRC safe properties of jet cross-sections and related variables, such as event shapes and energy correlation functions, were first studied in Refs. [14–16]. We note that in the case of inclusive observables, for which $J_r = 1$, the cancelation between the soft and collinear contributions in Eq. (6.8) is complete and, consequently, the total cross-section remains unchanged by the emission of soft and collinear particles, as it should. In case of exclusive (but IRC safe) measurements, including jet definitions, although

the singularities cancel, the kinematic dependence of the observable can cause an unbalance between real and virtual contributions, which manifests itself with the appearance of potentially large logarithmic corrections to any orders in perturbation theory. There exist techniques to resum these large logarithmic corrections to all perturbative orders. In this context, the concept of recursive IRC safety is particularly useful [17]. Finally, we also mention that recent work [18–21] has introduced the concept of Sudakov safety, which enables to extend the reach of (resummed) perturbation theory beyond the IRC domain.

6.2.2 *Cone algorithms*

Cone algorithms were first introduced in a famous paper by Sterman and Weinberg [13]. They are based on the idea that jets represent dominant flows of energy in a collision event. According to this definition, a 2-jet event in e^+e^- collisions is such that all, but a fraction ε of the total energy is contained into two cones of opening angle δ. Considering the $\mathcal{O}(\alpha_s)$ calculation in Eq. (6.8), we have that the two-parton measurement function is equal to unity, $J_{\varepsilon,\delta}(k_1, k_2) = 1$, because if we only have two partons in the final states, they must be hard and well separated in angle. If instead we have three partons, the 2-jet condition becomes[4]

$$\begin{aligned} J_{\varepsilon,\delta}(k_1, k_2, k_3) = \Theta\left(\min(\theta_{12}, \theta_{13}, \theta_{23}) < \delta\right) \\ + \Theta\left(\min(\theta_{12}, \theta_{13}, \theta_{23}) > \delta\right) \Theta\left(\min(E_1, E_2, E_3) < \varepsilon\right), \end{aligned} \tag{6.11}$$

where we have introduced the angles θ_{ij} between the directions of motion of particle i and j and their energies E_i. The first Θ function says that if the angle between the three momenta of the closest pair of parton is below δ, then the two partons belong to the same jet and so the event has two jets. The second set of constraints tells us that a configuration in which the three partons are well separated in angle, but the energy of the softest particle is below threshold, leads to two jets. In the limit where two directions become collinear, the second line of Eq. (6.11) is never satisfied, while the first one becomes $\Theta(0 < \delta) = 1$. Similarly, in the soft limit, the energy constraints

[4] We introduce the following notation for the Heaviside step function: $\Theta(a > b) = 1$, if $a > b$, and $\Theta(a > b) = 0$, if $a < b$.

are always satisfied and we obtain

$$\Theta\left(\min(\theta_{12}, \theta_{13}, \theta_{23}) < \delta\right) + \Theta\left(\min(\theta_{12}, \theta_{13}, \theta_{23}) > \delta\right) = 1.$$

Thus, Sterman–Weinberg cones are IRC safe, at least to $\mathcal{O}(\alpha_s)$.

In realistic hadron-collider environments, cone algorithms rely on the concept of a *stable cone*, i.e., the sum of all particles' momenta in the cone should point in the direction of the center of the cone. In order to find stable cones, the JetClu [22] and (various) midpoint-type [23, 24] cone algorithms use a procedure that starts with a given set of seed particles. Taking each of them as a candidate cone center, one calculates the cone contents, finds a new center based on the four-vector sum of the cone contents, and iterates until a stable cone is found. However, stable cones in a given event can overlap, meaning particles can belong to more than one cone. The most common approach is to run a split–merge procedure once the stable cones have been found. This iteratively takes the most overlapping stable cones and either merges them or splits them depending on their overlapping fraction. The procedure is repeated until one is left with non-overlapping objects that can be identified as jets.

Cone algorithms were widely used by the Tevatron experiments. For instance, the JetClu algorithm, used during Run I at the Tevatron, takes the set of particles as seeds, optionally above a given threshold in transverse momentum. This can be shown to be IRC unsafe for configuration for which two hard particles are within a distance smaller than twice the cone radius, rendering JetClu unsatisfactory for theoretical calculations. Midpoint-type algorithms, used for Run II of the Tevatron, added to the list of seeds the intermediate points between any pair of stable cones found by JetClu. This is still infrared unsafe, this time when 3 hard particles are in the same vicinity, i.e., one order later in the perturbative expansion than the JetClu algorithm. This IRC issue was solved by the introduction of the SISCone [25] algorithm, which provably finds all possible stable cones in an event, making the stable cone search IRC safe.

6.2.3 *Sequential recombination algorithms*

Due to the aforementioned problems related to IRC safety, the use of cone algorithms in modern high-energy physics experiments has dwindled in favor of approaches that form jets by successive pairwise combinations of more elementary objects. These sequential recombination algorithms are

based on the idea that, from a perturbative QCD viewpoint, jets are the product of successive parton branchings, as we discussed at the beginning of this chapter. Thus, if jets are supposed to capture the properties of the very energetic partons produced in the hard collision, jet algorithms attempt to invert the parton shower process by successively recombining pairs of particles, which are close to each other, according to some user-defined (and physics-inspired) metric, into objects that can be taken as proxies to the hard partons. The metric used in this process determines the type of algorithm.

6.2.3.1 *JADE algorithm*

A natural choice for the distance metric is the invariant mass of the pair under examination $m_{ij}^2 = (p_i + p_j)^2$. This is clearly a Lorentz-invariant measure that reflects important features of QCD, namely, collinear splittings and soft emissions, which both produce small invariant masses, are favored. The sequential recombination algorithm that exploits this distance measure was first introduced by the JADE collaboration at the PETRA e^+e^- collider and it is therefore called the JADE algorithm [26, 27]. It is formulated as follows:

(1) Take the particles in the event as the initial list of objects.
(2) For each pair of particles i, j work out the distance

$$y_{ij} = \frac{2E_iE_j(1-\cos\theta_{ij})}{Q^2}, \tag{6.12}$$

where Q is the total energy. If particles i and j are massless, then y_{ij} is the just their squared invariant mass, normalized to the square of the total energy.
(3) Find the minimum $y_{\min}$ of all the y_{ij}.
(4) If $y_{\min}$ is below some *jet resolution threshold* y_{cut}, then recombine i and j into a single new particle (or "pseudojet") and repeat from step 2.
(5) Otherwise, declare all remaining particles to be jets and terminate the iteration.

The parameter y_{cut} plays the role of the resolution variable of the algorithm. In particular, as y_{cut} grows smaller, softer and/or more collinear radiation is resolved into separate jets. Thus, the number of jets found by the JADE algorithm is controlled by a single parameter rather than the two parameters (ε and δ) of Sterman–Weinberg cones.

The JADE algorithm is IRC safe because soft particles are recombined at the beginning of the clustering, as they produce small invariant masses with any other particle, as do pairs of collinear particles. However, the presence of the product $E_i E_j$ in the distance measure means that two very soft particles moving in opposite directions may be recombined into a single particle in the early stages of the clustering, which is at odds with the intuitive picture of a jet as a stream of collimated particles. This peculiar behavior is reflected in a rather intricate structure of higher-order corrections for the distributions of the JADE resolution scale [28–30]. In a modern language, it is possible to show that despite being IRC safe, the JADE algorithm lacks recursive IRC safety [17].

6.2.3.2 *Generalized k_t algorithm*

Due to the unwanted features of the JADE algorithm, sequential recombination algorithms with alternative metrics have been suggested since the early 1990s. Here, instead of a historical discussion, we group these algorithms into a one-parameter family, the *generalized k_t algorithm* [8], discussing the most common examples. We present the algorithm in its incarnation for hadron–hadron collisions, although it can also be applied to e^+e^-, with small modifications.[5] The algorithm proceeds as follows:

(1) Take the particles in the event as the initial list of objects.
(2) From the list of objects, build two sets of distances: a pairwise distance

$$d_{ij} = \min(p_{t,i}^{2p}, p_{t,j}^{2p}) \Delta R_{ij}^2, \tag{6.13}$$

where p is a free parameter and $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ is the geometric distance in the rapidity-azimuthal angle plane, and a "beam distance":

$$d_{iB} = p_{t,i}^{2p} R^2, \tag{6.14}$$

with R the algorithm resolution parameter, often called the jet radius.

[5] At hadron colliders, we typically express the kinematics in terms of transverse momentum, rapidity, and azimuth, while, as we have already seen, in lepton–lepton colliders, energy and (polar) angle are preferred.

(3) Find the minimum of all d_{ij} and d_{iB}.
(4) If the smallest distance is a d_{ij}, then objects i and j are removed from the list and recombined into a pseudo-jet which is itself added to the list.
(5) If the smallest is a d_{iB}, object i is called a jet and removed from the list.
(6) Go back to step 2 until all the list of objects is empty.

In all cases, we see that if two objects are close in the rapidity-azimuth plane, as would be the case after a collinear splitting, the distance d_{ij} becomes small and the two objects are more likely to recombine. Similarly, when $\Delta R_{ij} > R$, the beam distance becomes smaller than the inter-particle distance and objects are no longer recombined, making R a typical measure of the size of the jet. Indeed, if we only have two particles, any member of the generalized k_t family will cluster them together if their distance in the rapidity-azimuth plane is less than R, irrespective of the value of the parameter p:

$$\min(p_{t,i}^{2p}, p_{t,j}^{2p})\Delta R_{ij}^2 < \min\left(p_{t,i}^{2p}R^2, p_{t,j}^{2p}R^2\right) \quad \Rightarrow \quad \Delta R_{ij} < R. \tag{6.15}$$

The situation changes if we consider three or more particles and indeed the shape of realistic jets strongly depends on the value of the parameter p, as we are about to discuss.

***k_t* algorithm:** The first solution to alleviate the issues related to the JADE algorithm, while preserving the idea of clustering soft particles first, was the so-called k_t algorithm [31, 32], which corresponds to taking $p = 1$ above. According to this metric, emissions with small transverse momentum are close and therefore are recombined early in the clustering, in accordance with the parton-shower picture. However, the presence of the "minimum" in the distance measure, instead of the product, guarantees that two soft objects geometrically far apart are not recombined, thus avoiding the issues encountered with JADE. It should be noted that, while physically motivated, the k_t distance enhances sensitivity to all sorts of low-energy, non-perturbative, effects, such as the underlying event and pile-up, and for this reason, k_t jets are seldom used in hadron–hadron collisions.

Cambridge/Aachen algorithm: Another specific incarnation is the Cambridge/Aachen algorithm [33, 34], which is obtained by setting $p = 0$ above. With this choice, the metric measures a purely geometrical distance

in the rapidity-azimuth plane and particles close in angles are recombined first. This choice is physically motivated because of the collinear enhancement of QCD splittings and it suffers less from the contamination due to soft backgrounds than the k_t algorithm does.

Anti-k_t algorithm: In the context of LHC physics, jets are almost always reconstructed with the anti-k_t algorithm [35], which corresponds to the generalized k_t algorithm with $p = -1$. This choice seems at first rather unnatural because it is at odds with the picture emerging from the QCD parton shower. However, its primary advantage consists in the fact that the anti-k_t metric favors clusterings between hard particles. Thus, anti-k_t jets grow by successively aggregating soft particles around a hard core, until the jet has reached a (geometrical) distance R away from its axis. Since two soft particles are always far away with the anti-k_t metric, anti-k_t jets have very little sensitivity to soft radiation and they appear to have circular shapes in the azimuth-rapidity plane. Indeed, anti-k_t behaves as a rigid cone in the soft limit, which simplifies all-order calculations of jet properties. From an experimental point of view, the resilience against soft radiation implies that anti-k_t jets are easier to calibrate. This is the main reason why it was adopted as the default jet clustering algorithm by all the LHC experiments.

6.2.4 *Sensitivity to soft physics*

The effect of soft radiation on jets clustered with different algorithms is shown in Fig. 6.4. The three-dimensional plots show calorimeter cells in the azimuth-rapidity plane, with the vertical axis measuring the transverse momentum carried by the particles in each cell. The shaded regions correspond to the active catchment area of each jet [36], which is obtained by adding infinitely soft particles (usually called ghosts) that are clustered with the hard jets, thus determining their boundaries. Anti-k_t jets have sharp and round boundaries, demonstrating resilience against soft physics. In actual experimental situations, this translates into reduced sensitivity to the underlying event and pile-up.

Another measure of a jet resilience to soft backgrounds is the back-reaction. Let us suppose to have a hard scattering event that produces a set of jets, with given properties. If we then add soft radiation to this event and we rerun the same jet algorithm, we will obtain a different set of jets. In particular, not only jets can acquire additional soft constituents, but we are also not guaranteed that a given jet will contain the same hard particles

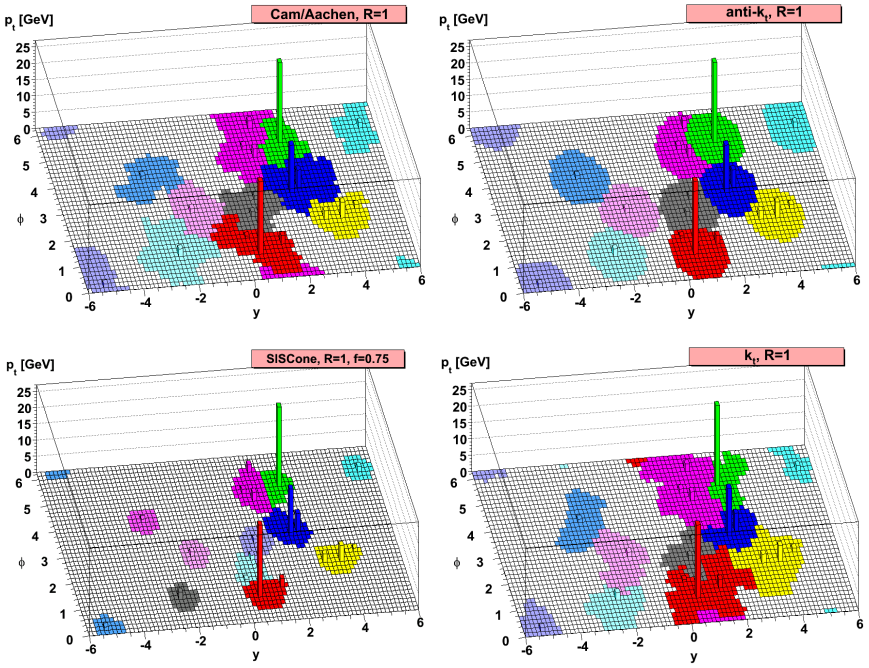

Figure 6.4. Representation of jets in the azimuth (ϕ) and rapidity (y) plane obtained with SISCone and with the three members of the generalized k_t family discussed here. All algorithms have $R = 1$, while $f = 0.75$ is the overlap parameter for the SIScone algorithm. While the jets obtained with the Cambridge/Aachen and k_t algorithm have irregular boundaries, the hard jets obtained with anti-k_t clustering are almost perfectly circular. SIScone produces smaller jets, which become more irregular as the number of constituents increases.

Source: Figure taken from Ref. [35].

of the original hard event. The back-reaction is precisely the deformation of the original jets because of the presence of the soft background. This is illustrated by the cartoon on the left-hand side of Fig. 6.5. The black dots represent the particles from the hard scattering, while the gray ones the (almost uniform) soft radiation, e.g., pile-up. The original jet, which is represented by the light gray area, is modified because of its interaction with the soft background (dark gray area).

The impact of the back-reaction on the transverse momentum of a jet is illustrated in Fig. 6.5, on the right, for different jet definitions. Positive values of $\Delta p_t^{(B)}$ correspond to transverse momentum gain, while negative ones to loss of p_t. We clearly see that back-reaction effects are strongly

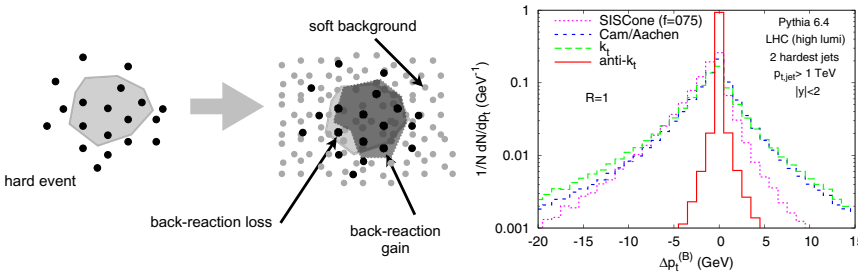

Figure 6.5. On the left, we show a cartoon describing the back-reaction effect, i.e., the modification of a hard jet due to its interactions with a soft background. On the right, we show the distribution of the transverse momentum change due to back-reaction for the anti-k_t algorithm as compared to k_t, Cambridge/Aachen, and SISCone.

Source: Figure taken from Ref. [35].

suppressed for the anti-k_t algorithm relative to the others, a feature that can help reduce the smearing of jets' momenta due to the underlying event and pile-up.

6.3 Jets as Tools

Jets are ubiquitous objects in collider phenomenology. They are employed in dedicated measurements that aim to stress-test our understanding of the Standard Model to the highest accuracy. In this context, we mention, for instance, measurements of electroweak bosons in association with many jets. Jets also appear in numerous searches for new physics, e.g., cascades of supersymmetric particles, events with one jet produced in association with missing energy in searches for dark matter, and, generically, searches for heavy states decaying into hadrons. Let us consider, for instance, a search for a new resonance X, which decays into quarks. If the mass of this new resonance is very large, it is most likely produced with a small velocity in the laboratory frame or, equivalently, with small transverse momentum. Then, its decay products move in opposite direction, fragmenting into well-separated jets, as depicted in the left-hand cartoon of Fig. 6.6. The most basic search strategy in this scenario is then to look for resonance peaks (the so-called "bump hunt") in the invariant mass distributions of the two jet with the highest transverse momenta.

We might also be interested in studying the hadronic decays of particles with mass around the electroweak scale. These can be Standard Model particles like electroweak and Higgs bosons or top quarks but also any

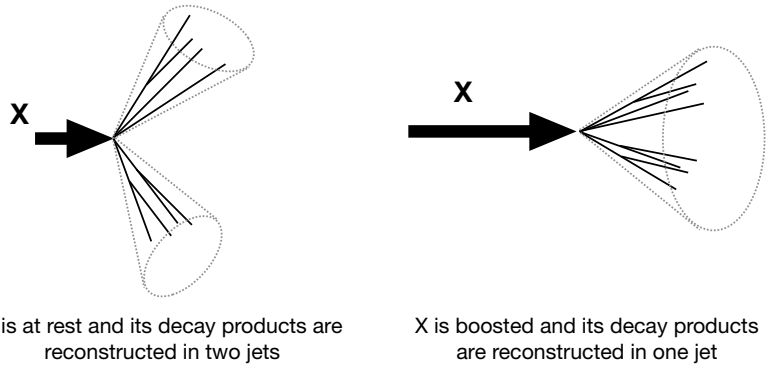

Figure 6.6. If a heavy state X is produced at rest, in the laboratory frame, its hadronic decay products are reconstructed as two (or more) well-separated jets, as depicted on the left. However, if its transverse momentum is large, $p_t \gtrsim 2m/R$, its decay products are collected in a single jets of radius R.

new particle with a mass of the order of the electroweak scale. Due to its unprecedentedly high colliding energy, the LHC is reaching energies far above the electroweak scale. Therefore, analyzes and searching strategies developed for earlier colliders, in which electroweak scale particles were produced with small velocities, had to be fundamentally reconsidered. In particular, as the transverse momentum of the decaying particle grows larger, its decay products become more collimated. If $p_t \gtrsim \frac{2m}{R}$, the decay products are reconstructed into a jet of radius R, as depicted in the right-hand cartoon of Fig. 6.6.

At the LHC, this scenario is particularly relevant for Higgs physics and, in particular, in the context of measurements of the couplings of the Higgs boson to the fermions. This is a crucial test for the Higgs mechanism of electroweak symmetry breaking, which predicts that the couplings to the fermions should be proportional to their masses. Despite the fact that the branching ratios into heavy (beauty b and charm c) flavors are not small, these measurements are challenging because of the large QCD background. However, when the Higgs boson is produced with a large transverse momentum, its decay products are likely to be reconstructed in a single jet. The presence of the Higgs boson can be then inferred by studying the substructure of this jet [37–39]. Consequently, jet substructure has emerged as an important tool for searches at the LHC, and a vibrant

field of theoretical and experimental research has developed in the past decade, producing a variety of studies and techniques [4, 40–46].

We have already said that, in the context of resolved analyzes, the key observable to look at is the invariant mass distribution of the two jets. We can try and play the same strategy in the case of analyzes in the boosted regime and look at the jet invariant mass:

$$m^2_{\text{jet}} = \left(\sum_{i \in \text{jet}} p_i \right)^2 , \tag{6.16}$$

where p_i are the four momenta of the jet's constituents. If the jet comprises all the debris of the decay, then its invariant mass distribution should peak around the decaying particle mass. On the other hand, background, i.e., QCD, jets have no intrinsic mass scale[6] and therefore their invariant mass must be proportional to the jet transverse momentum. Thus, one may hope that a cut on the jet invariant mass distribution will do the trick. It turns out that, despite being an important discriminant, the jet mass distribution is not enough. For instance, the jet mass turns out to be very sensitive to soft contamination, such as the underlying event and pile-up, resulting in degradation of its performance. We can see a striking example of this in Fig. 6.7, on the left. The invariant mass distribution of the leading QCD jet is shown, as measured by the ATLAS collaboration during the first run of the LHC. The different curves correspond to different pile-up situations, as measured by the number of reconstructed interaction vertices. Despite the transverse momentum of the jet being rather high, $p_t \in [600, 800]$ GeV, we can see that pile-up has a huge effect on the distribution, causing a shift of several tens of GeV. Thus, if we want to develop tools that can successfully discriminate signal and background jets in the boosted regime, we must move beyond the standard jet invariant mass and find new strategies to scrutinize the substructure of jets.

6.3.1 *Grooming and tagging*

The two key concepts in jet substructure go under the names of *grooming* and *tagging*. Broadly speaking, a grooming procedure takes a jet as an input and tries to clean it up by removing constituents which, being at wide

[6]The hadron-formation scale Λ is always present, but it is much lower than the energy scales considered here.

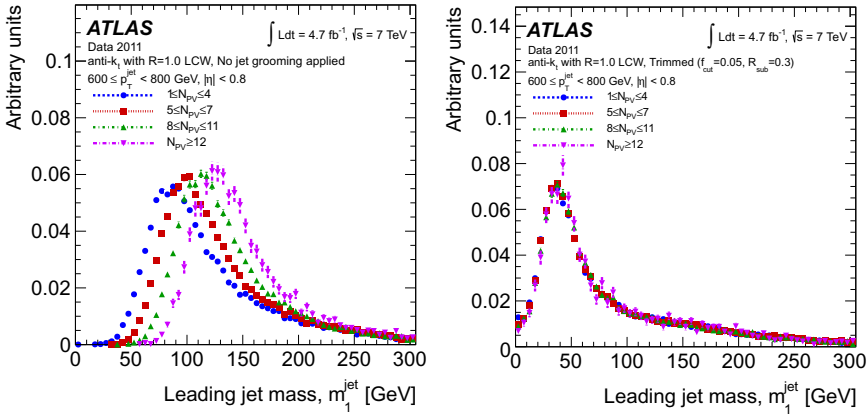

Figure 6.7. The leading jet mass distribution as measured by the ATLAS collaboration during LHC Run 1. The curves correspond to different numbers of primary vertices, a measure of the pile-up environment. The plot on the left is for standard jets, and the plot on the right for jets groomed with trimming [47].

Source: Figure taken from Ref. [48].

angle and relatively soft, are likely to come from contamination, such as the underlying event or pile-up. After this contamination has been removed, we are left with groomed jets that should be closer to our partonic picture. At this stage, we can perform a tagging step, namely, a cut on some kinematical variable that is able to distinguish signal from background. For instance, in electroweak boson decays, the energy sharing between the two daughters is symmetric. This is in contrast to QCD splittings $q \to qg$, for which the gluon tends to be soft. Thus, the energy sharing between subjets in the jets can be used as a tagging variable. We can build on this idea by noticing that high-p_t QCD jets are likely to appear as containing one prong, i.e., a hard core surrounded by a cloud of soft radiation. Electroweak (and Higgs) jets are instead two-pronged because they are initiated by a two-body decay into quarks. Jets that contain boosted top quarks feature three prongs because the top is so massive that goes through an electroweak decay before hadronizing, $t \to Wb$. If the W decays hadronically, then the top jet will contain three main subjets: one originated by the b quark and two from $W \to q\bar{q}'$. Thus, we can build tagging algorithms that distinguish jets according to the number of prongs they feature. The most famous example of such a tagger is called N-subjettiness [49, 50].

Many grooming algorithms have been developed, successfully tested, and are currently used in experimental analyzes, e.g., the mass-drop tagger [39], trimming [47], and pruning [51, 52]. A successful application of jet trimming by the ATLAS collaboration is shown in the right-hand plot of Fig. 6.7. The invariant mass distribution of the leading QCD jet is shown, but this time, jets are trimmed. We see that, in contrast to standard jets (on the left), no sensitivity to pile-up is found.[7]

By staring at the two plots in Fig. 6.7, we note a second interesting feature. The trimmed jet mass distribution is insensitive to pile-up, but it is not the same as the standard jet mass distribution, in the absence of pile-up. Thus, trimming is modifying standard jets, possibly carving away perturbative radiation too. This is something we should investigate because we do not want to undermine our perturbative understanding of jets. Regardless of their nature, substructure algorithms try to resolve jets on smaller angular and energy scales, thereby introducing new parameters. This challenges our ability of computing predictions and indeed most of the early theoretical studies of substructure tools were performed using Monte Carlo event generators. While these are powerful general-purpose tools, their essentially numerical nature offers little insight into the results produced or their detailed and precise dependence on the algorithms' parameters. A deeper, first-principle, understanding of the most used grooming and tagging techniques, both in the presence of background [53, 54] and signal jets [55, 56], was achieved when perturbative (all-order) techniques were employed to describe jet substructure. When this understanding was put at work, a second generation of substructure algorithms, which combined efficient signal-from-background discrimination together with robust theoretical understanding, was devised. One of them is SoftDrop [19], which we discuss in some detail.

The SoftDrop procedure starts with a standard jet, typically an anti-k_t jet in LHC studies. However, if we want to understand the substructure of this jet, the first thing we should do is to order the constituents in a way that reflects the jet formation history. Since the anti-k_t history does not have this feature, we recluster the jet with a more physical algorithm, namely, Cambridge-Aachen. After this procedure, we have at our disposal

[7] We should mention that in the more challenging pile-up environments of LHC Run 2 and 3, grooming algorithms are not enough to remove pile-up and dedicated pile-up subtraction techniques are applied.

a physically meaningful clustering tree, in which the clustering steps are ordered in angle, e.g., the final node, which corresponds to the first splitting, clusters together two prongs that are far away in the azimuth-rapidity plane. The SoftDrop procedure then performs the following steps:

(1) Break the jet j into two subjets by undoing the last stage of Cambridge-Aachen clustering. Label the resulting two subjets as j_1 and j_2.
(2) If the subjets pass the SoftDrop condition $\frac{\min(p_{t1},p_{t2})}{p_{t1}+p_{t2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R}\right)^{\beta}$, then deem j to be the final SoftDrop jet.
(3) Otherwise, redefine j to be equal to subjet with larger p_t and iterate the procedure.
(4) If j is a singleton and can no longer be declustered, then one can either remove j from consideration ("tagging mode") or leave j as the final SoftDrop jet ("grooming mode").

The difficulty posed by substructure algorithms in general, and SoftDrop in particular, is the presence of new parameters (here the angular exponent β and the energy fraction z_{cut}) that slice the phase space in a non-trivial way, resulting in potentially complicated all-order behavior of the observable at hand. This is exemplified in Fig. 6.8, where we show the

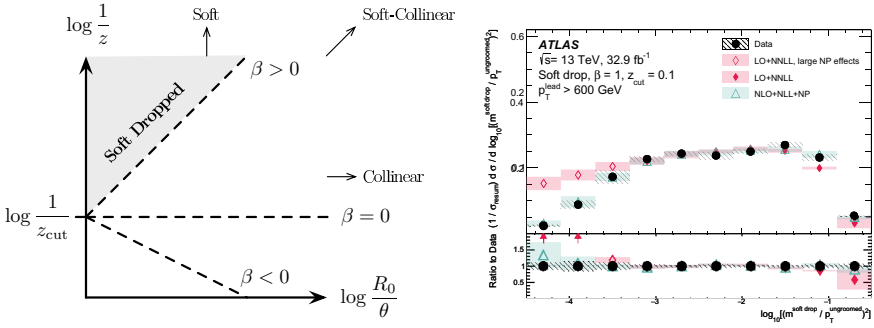

Figure 6.8. On the left, we show the SoftDrop phase space for emissions on the $(\ln \frac{1}{z}, \ln \frac{R}{\theta})$ Lund plane. For $\beta > 0$, soft emissions are vetoed while much of the soft-collinear region is maintained. For $\beta = 0$, both soft and soft-collinear emissions are vetoed. For $\beta < 0$, all (two-prong) singularities are regulated by the SoftDrop procedure. Figure taken from Ref. [19]. On the right, we show a measurement of the normalized SoftDrop jet mass distribution by the ATLAS collaboration. The data are compared to two different high-precision perturbative calculations, showing excellent agreement, across a wide range of the observable.

Source: Figure taken from Ref. [57].

phase space for soft and collinear gluon emission, from a hard parton, in the $(\ln \frac{1}{z}, \ln \frac{R}{\theta})$ plane, where $0 \leq z \leq 1$ is the energy fraction of the emitted gluon with respect to the hard parton initiating the jet, and $0 \leq \theta \leq R$ is the angle of the emission, measured from the hard parton. This representation of the soft and collinear phase space is often called the Lund plane. In the soft and collinear limit, the SoftDrop condition can be written as

$$z > z_{\text{cut}} \left(\frac{\theta}{R}\right)^{\beta} \quad \Rightarrow \quad \ln \frac{1}{z} < \ln \frac{1}{z_{\text{cut}}} + \beta \ln \frac{R}{\theta} \tag{6.17}$$

Thus, vetoed emissions lie above a straight line of slope β on the $(\ln \frac{1}{z}, \ln \frac{R}{\theta})$ plane, as shown in Fig. 6.8. For $\beta > 0$, collinear splittings always satisfy the SoftDrop condition, so a SoftDrop jet still contains all of its collinear radiation. The amount of soft-collinear radiation that satisfies the SoftDrop condition depends on the relative scaling of the energy fraction z to the angle θ. As $\beta \to 0$, more of the soft-collinear radiation of the jet is removed, and in the $\beta = 0$ limit, all soft-collinear radiation is removed. In this limit, SoftDrop essentially coincides with the modified Mass Drop Tagger [53, 54]. In the strict $\beta = 0$ limit, collinear radiation is only maintained if $z > z_{\text{cut}}$. Finally, for $\beta < 0$, the soft-collinear region is removed and a hard splitting is imposed. For example, $\beta = -1$ roughly corresponds to a cut on the relative transverse momentum of the two prongs under scrutiny.

The above understanding can be formalized and precision calculations of observables measured on SoftDrop jets have been performed [58, 59]. Furthermore, while by design SoftDrop reduces the sensitivity to the underlying event and pile-up, it has been shown that this algorithm can also reduce the size of hadronization corrections, although they acquire a more complicated structure [53, 60–62].

Thus, because of their theoretical properties, i.e., good perturbative behavior and reduced sensitivity to non-perturbative physics, SoftDrop jets have emerged as an excellent playground for QCD studies at the LHC. As an example of this, we show on the right-hand side of Fig. 6.8 the comparison between a measurement of the SoftDrop jet mass performed by the ATLAS collaboration [57] (CMS also performed similar measurements, see, for instance, Ref. [63]) to high-precision perturbative calculations by two different groups: LO+NNLL [58] and NLO+NLL+NP [60], where the acronyms denote the accuracy of the calculations is apparent. The agreement is excellent and only in the three lower bins there is need for non-perturbative corrections, which are included in the NLO+NLL+NP calculation. The remarkable theoretical understanding reached for SoftDrop

jets, together with the fine measurements performed by the experiments, has led to studies assessing the use of jet substructure techniques to extract Standard Model parameters, such as the strong coupling [64–66] or the top quark mass [67]. Furthermore, these observables can also be used to stress-test and improve event-simulation tools, such as parton showers and hadronization models.

6.3.2 *Jets in the era of artificial intelligence*

Our journey through jet physics would not be complete without a discussion about new approaches based on artificial intelligence. The rapid development, within and outside academia, of machine-learning techniques is having a profound impact on many aspects of society and fundamental research is not immune to this. In the context of jet physics, this revolution has brought to life a third generation of jet substructure techniques, which are now the gold standard for LHC Run 3 analyzes. However, because of its novelty and ongoing rapid progress, machine learning can still be considered an *ad hoc* field: a multitude of problems can be solved and addressed with different techniques, but some of the basic principles, the underlying structure, and a unified picture are still missing. Thus, we believe that times are not mature yet for a complete and exhaustive description of these techniques in a book.[8] Therefore, in this final section, we limit ourselves to raise a few points about the relation between deep-learning tools and expert-knowledge developed in more than ten years of jet substructure studies.

A bread and butter application of machine learning to particle physics are classification problems, including jet tagging. In this context, classification algorithms are typically trained on a control sample, which could be either Monte Carlo pseudo-data or a high-purity dataset, and then applied to an unknown sample to classify its properties. This is an example of so-called supervised learning. These ideas have been exploited in particle physics for a long time. However, because of limitations on efficiency and computing power, algorithms used to be applied to relatively low-dimensional projections of the full radiation pattern that one wished to classify. Even so, such projections usually corresponded to physically motivated observables, such as the jet mass, and therefore limitation in

[8] We refer the interested readers to Ref. [70].

performance was mitigated with physics understanding. Current developments in machine learning allows us to move away from low-dimensional projections and exploit deep neural networks to perform classification. This opens up the door to almost limitless possibilities that go far beyond supervised learning. Just to mention a few examples, unsupervised learning has led to the design of algorithms, which can be applied, for instance, to anomaly detection in new physics searches. Furthermore, neural network can be used not only for classification but also for simulations (e.g., parton showers, hadron formation, and detector responses) in a fast and faithful way — the particle physics equivalent of deepfake.

The most successful innovations in machine learning are coming from outside high energy physics (and chiefly from the industry giants). However, particle physics provides us with one of the few examples of a big-data system with a deep scientific understanding of the underlying model, potentially allowing us to get more insight into the broader machine-learning field. In this context, an interesting debate to mention has to do with the choice of inputs and architecture to use when building a neural network for a specific physics case. Should we be as agnostic as possible and provide a complex network with raw data from the experiments? Or should we build on our understanding of the physical processes and use physically motivated observables as input to (possibly simpler) machine learning algorithms? The former approach has the advantage of being unbiased, while following the second one we may hope, for instance, to understand what kind of information the network is learning from the data.

We close this discussion with a comparison between these two philosophies. In order to do that, we go back to our electroweak boson tagging problem. We can view a particle detector, and in particular the hadronic calorimeter, as a huge camera, taking pictures of particle collisions and, using the information from the calorimeter cells, we can build jet images [68, 71]. After appropriate averaging and pre-processing, the jet images can be input to machine-learning algorithms that are appropriate for pattern recognition, such as convolutional neural networks. Alternatively, we can build a picture of the jets based on our understanding of QCD. This is provided by the (primary) Lund jet plane [69]. The Lund jet plane is constructed by parsing backward the clustering history of a jet's Cambridge-Aachen tree, similar to the SoftDrop procedure previously described. At each step, the kinematics of the splitting, e.g., the distance between the two branches in the azimuth-rapidity plane Δ and the relative transverse momentum k_t, is recorded. The set of values that we obtain always following

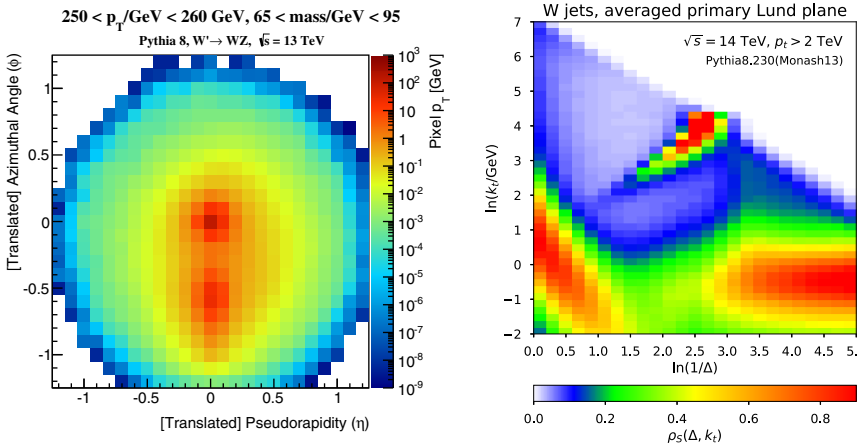

Figure 6.9. Two pictures of W-initiated jets. On the left, the average calorimetric image, after pre-processing, where pixel colors represent the energy deposited in a calorimeter cell (figure taken from Ref. [68]). On the right, the average primary Lund plane density, where colors represent the density of recorded splittings in a given $(-\log\Delta, \log k_t)$ cell (figure taken from Ref. [69]). Note that the presence of the initiating W boson appears as a two-pronged structure, on the left, and as a hot spot at $\log\frac{1}{\Delta} \simeq 0.4$ and $\log k_t \simeq 4$, on the right. Detailed comparisons between the two plots should be taken with a grain of salt because of the rather different transverse momentum selections.

the harder branch constitutes the primary Lund jet plane. Considering many jets, we can construct the density of the primary plane. Examples of a jet image and a primary Lund plane image for W jets are shown in Fig. 6.9.

6.4 Closing Remarks

We conclude this chapter by stressing once again that the key aspect that repeatedly appears in the context of jet physics is the design of algorithms that can be meaningfully used by both theory and experimental communities. Very often this implies the necessity of a tradeoff between performance and robustness. In the 1990s, one of the reason for preferring cone algorithms over sequential recombination ones was the issue of speed, an example of performance. However, as it turned out, the algorithms used at the Tevatron were not robust because they lacked IRC safety.

In the context of jet substructure studies, by performance, we usually mean the discriminating power of a tool when extracting a given signal from the QCD background, and by robustness we mean the ability to describe

the tool using perturbative QCD, i.e., being as little sensitive as possible to model-dependent effects such as hadronization, the underlying event, pile-up, or detector effects, all of which likely translate into systematic uncertainties in an experimental analysis.

We can apply similar considerations to the latest-generation machine-learning tools. On the one hand, these algorithms augment performance so much that they have become standard tools for collider physics. On the other hand, they are sometimes treated as black-boxes and, more often than not, their robustness is difficult to assess with standard technologies. It is an exciting challenge for particle theorists and experimentalists to find new ways to study these tools, assess their systematics, and, ultimately, find the best metric to measure their robustness.

Acknowledgments

I am indebted to Matteo Cacciari, Gavin Salam, and Gregory Soyez for many useful discussions about jet physics and for suggesting ideas and material for this chapter. I would also like to thank Andrea Coccaro, Marc Leblanc, and Jennifer Roloff for a critical reading of this manuscript. Finally, I am grateful to four graduate students at the University of Genova, Simone Caletti, Andrea Ghira, Anna Rinaudo, and Martino Tanasini, for reading this chapter and for providing me with very valuable feedback.

References

[1] G. Soyez. Pileup mitigation at the LHC: A theorist's view. *Physics Reports*, 803:1–158, 2019. DOI: 10.1016/j.physrep.2019.01.007.

[2] A. J. Larkoski, D. Neill, and J. Thaler. Jet shapes with the broadening axis. *JHEP*, 04:017, 2014. DOI: 10.1007/JHEP04(2014)017.

[3] G. P. Salam. Towards jetography. *European Physical Journal C*, 67:637–686, 2010. DOI: 10.1140/epjc/s10052-010-1314-6.

[4] S. Marzani, G. Soyez, and M. Spannowsky. *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-Object Phenomenology*, Vol. 958. Springer, 2019. DOI: 10.1007/978-3-030-15709-8.

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

[6] G. Aad, *et al.* Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions. *European Physical Journal C*, 81(4):334, 2021. DOI: 10.1140/epjc/s10052-021-09054-3.

[7] M. Cacciari and G. P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Physics Letters B*, 641:57–61, 2006. DOI: 10.1016/j.physletb.2006.08.037.

[8] M. Cacciari, G. P. Salam, and G. Soyez. FastJet user manual. *European Physical Journal C*, 72:1896, 2012. DOI: 10.1140/epjc/s10052-012-1896-2.

[9] J. M. Butterworth, J. P. Couchman, B. E. Cox, and B. M. Waugh. KtJet: A C++ implementation of the K-perpendicular clustering algorithm. *Computer Physics Communications*, 153:85–96, 2003. DOI: 10.1016/S0010-4655(03)00156-5.

[10] F. Bloch and A. Nordsieck. Note on the radiation field of the electron. *Physical Review*, 52:54–59, 1937. DOI: 10.1103/PhysRev.52.54.

[11] T. Kinoshita. Mass singularities of Feynman amplitudes. *Journal of Mathematical Physics*, 3:650–677, 1962. DOI: 10.1063/1.1724268.

[12] T. D. Lee and M. Nauenberg. Degenerate systems and mass singularities. *Physical Review*, 133:B1549–B1562, 1964. DOI: 10.1103/PhysRev.133.B1549.

[13] G. F. Sterman and S. Weinberg. Jets from quantum chromodynamics. *Physical Review Letters*, 39:1436, 1977. DOI: 10.1103/PhysRevLett.39.1436.

[14] G. F. Sterman. Mass divergences in annihilation processes. 1. Origin and nature of divergences in cut vacuum polarization diagrams. *Physical Review D*, 17:2773, 1978. DOI: 10.1103/PhysRevD.17.2773.

[15] G. F. Sterman. Mass divergences in annihilation processes. 2. Cancellation of divergences in cut vacuum polarization diagrams. *Physical Review D*, 17:2789, 1978. DOI: 10.1103/PhysRevD.17.2789.

[16] G. F. Sterman. Zero mass limit for a class of jet related cross-sections. *Physical Review D*, 19:3135, 1979. DOI: 10.1103/PhysRevD.19.3135.

[17] A. Banfi, G. P. Salam, and G. Zanderighi. Principles of general final-state resummation and automated implementation. *JHEP*, 03:073, 2005. DOI: 10.1088/1126-6708/2005/03/073.

[18] A. J. Larkoski and J. Thaler. Unsafe but calculable: Ratios of angularities in perturbative QCD. *JHEP*, 09:137, 2013. DOI: 10.1007/JHEP09(2013)137.

[19] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler. Soft drop. *JHEP*. 05:146, 2014. DOI: 10.1007/JHEP05(2014)146.

[20] A. J. Larkoski, S. Marzani, and J. Thaler. Sudakov safety in perturbative QCD. *Physical Review D*, 91(11):111501, 2015. DOI: 10.1103/PhysRevD.91.111501.

[21] P. T. Komiske, E. M. Metodiev, and J. Thaler. The hidden geometry of particle collisions. *JHEP*, 07:006, 2020. DOI: 10.1007/JHEP07(2020)006.

[22] F. Abe, *et al.* The topology of three jet events in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV. *Physical Review D*, 45:1448–1458, 1992. DOI: 10.1103/PhysRevD.45.1448.

[23] G. C. Blazey, *et al.* Run II jet physics. In: *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*, May 2000, pp. 47–77.

[24] V. M. Abazov, *et al.* Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review D*, 85:052006, 2012. DOI: 10.1103/PhysRevD.85.052006.

[25] G. P. Salam and G. Soyez. A practical seedless infrared-safe cone jet algorithm. *JHEP*, 05:086, 2007. DOI: 10.1088/1126-6708/2007/05/086.

[26] W. Bartel, *et al.* Experimental studies on multi-jet production in e+ e- annihilation at PETRA energies. *Zeitschrift für Physik*, 33:23, 1986. DOI: 10.1007/BF01410449.

[27] S. Bethke, *et al.* Experimental investigation of the energy dependence of the strong coupling strength. *Physics Letters B*, 213:235–241, 1988. DOI: 10.1016/0370-2693(88)91032-5.

[28] N. Brown and W. J. Stirling. Jet cross-sections at leading double logarithm in e+ e- annihilation. *Physics Letters B*, 252:657–662, 1990. DOI: 10.1016/0370-2693(90)90502-W.

[29] S. Catani. Jet topology and new jet counting algorithms. *Ettore Majorana International Science Series Physical Sciences*, 60:21–41, 1992. DOI: 10.1007/978-1-4615-3440-2_2.

[30] G. Leder. Jet fractions in e+ e- annihilation. *Nuclear Physics B*, 497:334–344, 1997. DOI: 10.1016/S0550-3213(97)00240-X.

[31] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. *Nuclear Physics B*, 406:187–224, 1993. DOI: 10.1016/0550-3213(93)90166-M.

[32] S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Physical Review D*, 48:3160–3166, 1993. DOI: 10.1103/PhysRevD.48.3160.

[33] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997. DOI: 10.1088/1126-6708/1997/08/001.

[34] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In: *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, April 1998, pp. 270–279.

[35] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. DOI: 10.1088/1126-6708/2008/04/063.

[36] M. Cacciari, G. P. Salam, and G. Soyez. The catchment area of jets. *JHEP*, 04:005, 2008. DOI: 10.1088/1126-6708/2008/04/005.

[37] M. H. Seymour. Searches for new particles using cone and cluster jet algorithms: A comparative study. *Zeitschrift für Physik*, 62:127–138, 1994. DOI: 10.1007/BF01559532.

[38] J. M. Butterworth, B. E. Cox, and J. R. Forshaw. WW scattering at the CERN LHC. *Physical Review D*, 65:096014, 2002. DOI: 10.1103/PhysRevD.65.096014.

[39] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Physical Review Letters*, 100:242001, 2008. DOI: 10.1103/PhysRevLett.100.242001.

[40] A. Abdesselam, *et al.* Boosted objects: A probe of beyond the standard model physics. *European Physical Journal C*, 71:1661, 2011. DOI: 10.1140/epjc/s10052-011-1661-y.

[41] A. Altheimer, *et al.* Jet substructure at the tevatron and LHC: New results, new tools, new benchmarks. *Journal of Physics G*, 39:063001, 2012. DOI: 10.1088/0954-3899/39/6/063001.

[42] A. Altheimer, *et al.* Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012. *The European Physical Journal C*, 74(3):2792, 2014. DOI: 10.1140/epjc/s10052-014-2792-8.

[43] D. Adams, *et al.* Towards an understanding of the correlations in jet substructure. *European Physical Journal C*, 75(9):409, 2015. DOI: 10.1140/epjc/s10052-015-3587-2.

[44] A. J. Larkoski, I. Moult, and B. Nachman. Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning. *Physics Reports*, 841:1–63, 2020. DOI: 10.1016/j.physrep.2019.11.001.

[45] R. Kogler, *et al.* Jet substructure at the Large Hadron Collider: Experimental review. *Reviews of Modern Physics*, 91(4):045003, 2019. DOI: 10.1103/RevModPhys.91.045003.

[46] B. Nachman, *et al.* Jets and jet substructure at future colliders. *Frontiers in Physics*, 10:897719, 2022. DOI: 10.3389/fphy.2022.897719.

[47] D. Krohn, J. Thaler, and L.-T. Wang. Jet trimming. *JHEP*, 02:084, 2010. DOI: 10.1007/JHEP02(2010)084.

[48] G. Aad, *et al.* Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 09:076, 2013. DOI: 10.1007/JHEP09(2013)076.

[49] J. Thaler and K. Van Tilburg. Identifying boosted objects with N-subjettiness. *JHEP*, 03:015, 2011. DOI: 10.1007/JHEP03(2011)015.

[50] J. Thaler and K. Van Tilburg. Maximizing boosted top identification by minimizing N-subjettiness. *JHEP*, 02:093, 2012. DOI: 10.1007/JHEP02(2012)093.

[51] S. D. Ellis, C. K. Vermilion, and J. R. Walsh. Techniques for improved heavy particle searches with jet substructure. *Physical Review D*, 80:051501, 2009. DOI: 10.1103/PhysRevD.80.051501.

[52] S. D. Ellis, C. K. Vermilion, and J. R. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Physical Review D*, 81:094023, 2010. DOI: 10.1103/PhysRevD.81.094023.

[53] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013. DOI: 10.1007/JHEP09(2013)029.

[54] M. Dasgupta, A. Fregoso, S. Marzani, and A. Powling. Jet substructure with analytical methods. *European Physical Journal C*, 73(11):2623, 2013. DOI: 10.1140/epjc/s10052-013-2623-3.

[55] I. Feige, M. D. Schwartz, I. W. Stewart, and J. Thaler. Precision jet substructure from boosted event shapes. *Physical Review Letters*, 109:092001, 2012. DOI: 10.1103/PhysRevLett.109.092001.

[56] M. Dasgupta, A. Powling, and A. Siodmok. On jet substructure methods for signal jets. *JHEP*, 08:079, 2015. DOI: 10.1007/JHEP08(2015)079.

[57] M. Aaboud, *et al.* Measurement of the soft-drop jet mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physical Review Letters*, 121(9): 092001, 2018. DOI: 10.1103/PhysRevLett.121.092001.

[58] C. Frye, A. J. Larkoski, M. D. Schwartz, and K. Yan. Factorization for groomed jet substructure beyond the next-to-leading logarithm. *JHEP*, 07: 064, 2016. DOI: 10.1007/JHEP07(2016)064.

[59] A. Kardos, A. J. Larkoski, and Z. Trócsányi. Groomed jet mass at high precision. *Physics Letters B*, 809:135704, 2020. DOI: 10.1016/j.physletb. 2020.135704.

[60] S. Marzani, L. Schunk, and G. Soyez. The jet mass distribution after soft drop. *European Physical Journal C*, 78(2):96, 2018. DOI: 10.1140/epjc/ s10052-018-5579-5.

[61] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart. Nonperturbative corrections to soft drop jet mass. *JHEP*, 12:002, 2019. DOI: 10.1007/ JHEP12(2019)002.

[62] A. Pathak, I. W. Stewart, V. Vaidya, and L. Zoppi. EFT for soft drop double differential cross section. *JHEP*, 04:032, 2021. DOI: 10.1007/JHEP04(2021) 032.

[63] A. M. Sirunyan, *et al.* Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:113, 2018. DOI: 10.1007/JHEP11(2018)113.

[64] J. Baron, S. Marzani, and V. Theeuwes. Soft-drop thrust. *JHEP*, 08:105, 2018. DOI: 10.1007/JHEP08(2018)105. [Erratum: *JHEP*, 05:056, 2019].

[65] S. Marzani, D. Reichelt, S. Schumann, G. Soyez, and V. Theeuwes. Fitting the strong coupling constant with soft-drop thrust. *JHEP*, 11:179, 2019. DOI: 10.1007/JHEP11(2019)179.

[66] H. S. Hannesdottir, A. Pathak, M. D. Schwartz, and I. W. Stewart. Prospects for strong coupling measurement at hadron colliders using soft-drop jet mass, October 2022.

[67] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart. Extracting a short distance top mass with light grooming. *Physical Review D*, 100(7), 074021, 2019. DOI: 10.1103/PhysRevD.100.074021.

[68] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images — deep learning edition. *JHEP*, 07:069, 2016. DOI: 10.1007/ JHEP07(2016)069.

[69] F. A. Dreyer, G. P. Salam, and G. Soyez. The Lund jet plane. *JHEP*. 12: 064, 2018. DOI: 10.1007/JHEP12(2018)064.

[70] M. Feickert and B. Nachman. A living review of machine learning for particle physics, February 2021.

[71] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman. Jet-images: Computer vision inspired techniques for jet tagging. *JHEP*, 02:118, 2015. DOI: 10. 1007/JHEP02(2015)118.