# The Hidden Geometry of Particle Collisions

**Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler**

*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

*E-mail:* pkomiske@mit.edu, metodiev@mit.edu, jthaler@mit.edu

ABSTRACT: We establish that many fundamental concepts and techniques in quantum field theory and collider physics can be naturally understood and unified through a simple new geometric language. The idea is to equip the space of collider events with a metric, from which other geometric objects can be rigorously defined. Our analysis is based on the energy mover's distance, which quantifies the "work" required to rearrange one event into another. This metric, which operates purely at the level of observable energy flow information, allows for a clarified definition of infrared and collinear safety and related concepts. A number of well-known collider observables can be exactly cast as the minimum distance between an event and various manifolds in this space. Jet definitions, such as exclusive cone and sequential recombination algorithms, can be directly derived by finding the closest few-particle approximation to the event. Several area- and constituent-based pileup mitigation strategies are naturally expressed in this formalism as well. Finally, we lift our reasoning to develop a precise distance between theories, which are treated as collections of events weighted by cross sections. In all of these various cases, a better understanding of existing methods in our geometric language suggests interesting new ideas and generalizations.

# Contents

# 1 Introduction

Unification of ideas in physics has been an important way of achieving elegance, clarity, and simplicity, which in turn helps inspire meaningful new developments. In this paper, we use the energy mover's distance (EMD) between collider events [1] to provide a natural geometric language that unifies many important concepts and techniques in quantum field theory and collider physics from the past five decades. Furthermore, we introduce and discuss several new ideas inspired by this geometric approach to studying the space of events.

Throughout this paper, we refer to an event and its *energy flow* interchangeably. The energy flow, or distribution of energy, is the kinematic information that is experimentally observable and perturbatively well-defined in quantum field theories with massless particles [2]. As it relates to collider physics, the energy flow has been extensively studied [2–17], and this paper builds on many of these previous concepts. For an event consisting of $M$ particles with positive energies $E_i$ and angular directions $\hat{n}_i$, the energy flow is:

$$\mathcal{E}(\hat{n}) = \sum_{i=1}^{M} E_i \, \delta(\hat{n} - \hat{n}_i). \tag{1.1}$$

Note that the energy flow is insensitive to charge and flavor information. Particles are taken to be massless in the body of this paper, with $n_i^\mu = (1, \hat{n}_i)^\mu = p_i^\mu / E_i$, and the case of massive particles is discussed in App. A. In a hadron collider context, particle transverse momenta $p_{T,i}$ are typically used in place of particle energies, but we focus on energies in this paper to minimize extraneous notation.

The EMD was introduced in Ref. [1] as a metric between events. It is based on the well-known earth mover's distance [18–22], also known as the Wasserstein metric [23, 24]. Intuitively, the EMD between two events is the amount of "work" required to rearrange one event to the other. Its value can be obtained by solving the following optimal transport problem between energy flows $\mathcal{E}$ and $\mathcal{E}'$:

$$\mathrm{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}') = \min_{\{f_{ij} \geq 0\}} \sum_{i=1}^{M} \sum_{j=1}^{M'} f_{ij} \left( \frac{\theta_{ij}}{R} \right)^\beta + \left| \sum_{i=1}^{M} E_i - \sum_{j=1}^{M'} E'_j \right|, \tag{1.2}$$

$$\sum_{i=1}^{M} f_{ij} \leq E'_j, \qquad \sum_{j=1}^{M'} f_{ij} \leq E_i, \qquad \sum_{i=1}^{M} \sum_{j=1}^{M'} f_{ij} = \min \left( \sum_{i=1}^{M} E_i, \sum_{j=1}^{M'} E'_j \right), \tag{1.3}$$

where $\theta_{ij}$ is a pairwise distance between particles known as the *ground metric*, $R > 0$ is a parameter controlling the tradeoff between transporting energy and destroying it, and $\beta > 0$ is an angular weighting exponent.[1] For the angular metric between two massless particles,

---

[1] Strictly speaking, for the case of $\beta > 1$, one must raise the first term in Eq. (1.2) to the $1/\beta$ power for the EMD to be a proper metric satisfying the triangle inequality, in which case it is known as a $p$-Wasserstein metric with $p = \beta$. Additionally, $2R$ should be larger than or equal to the maximum distance in the ground space for the EMD to satisfy the triangle inequality. When written without subscripts, $\mathrm{EMD}(\mathcal{E}, \mathcal{E}')$ refers to the case of $\beta = 1$ and a sufficiently large $R$ to ensure that we have a proper metric. Even if the EMD is not a proper metric, though, it is still a valid optimal transport problem for any positive values of $\beta$ and $R$.
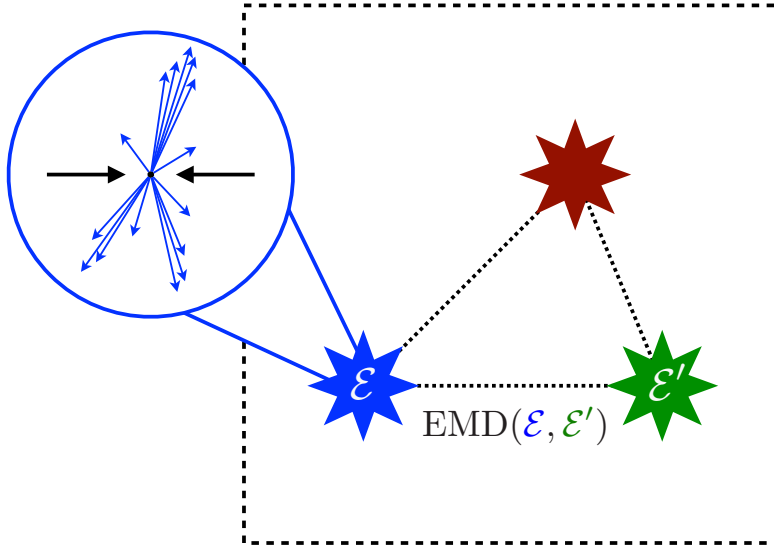
**Figure 1**: An illustration of the space of events. Each point in the space is a collider event consisting of the particles produced in a collision, as indicated by the blue event. The distance between events is quantified by the EMD, giving rise to a metric space. Geometry in this abstract space of events provides a natural language to understand many ideas and developments in quantum field theory and collider physics.

we focus on the case of

$$\theta_{ij} = \sqrt{2n_i^\mu n_{j\mu}} = \sqrt{2(1 - \hat{n}_i \cdot \hat{n}_j)}, \tag{1.4}$$

which reduces to their opening angle in the nearby limit.[2] The first term in Eq. (1.2) quantifies the difference in radiation patterns while the second term, which vanishes in the case of normalized energy flows, allows for the comparison of events with different total energies. The constraints in Eq. (1.3) specify that the amount of energy moved to or from a particle cannot exceed its initial energy, and that as much energy must be moved as possible.

The EMD has previously been used to bound modifications to infrared- and collinear-safe (IRC-safe) observables, distinguish different types of jets, and enable visualizations of the space of events [1]. It has also been used to explore the space of jets and quantify detector effects with CMS Open Data from the Large Hadron Collider (LHC) [25]. Alternative pairwise event distances were considered in Ref. [26] in the context of new physics searches. Here, we demonstrate that the EMD can be used to clarify numerous concepts throughout quantum field theory and collider physics using a unified language of event space geometry.

---

[2]Many modifications to this EMD definition are possible, including alternative angular distances such as strict opening angle or rapidity-azimuth distance as well as alternative notions of energy such as transverse momentum. In addition, energies can be normalized by dividing by their total scalar sum so that energy flows become proper probability distributions. If desired, the EMD in the center-of-mass frame can be phrased in a manifestly Lorentz-invariant way by replacing the particle energies $E_i$ with $p_i^\mu P_\mu / \sqrt{P_\mu P^\mu}$, where $P_\mu$ is the total event four-momentum.

In addition to demonstrating how concepts such as IRC safety, observables, jet finding, and pileup subtraction are related, we will develop new ideas and techniques in each of these areas, which we describe below.

Equipping collider events with a metric allows us to explore interesting geometric and topological ideas in the space of events. Fig. 1 illustrates the space of events with the EMD as a metric. One key construction for relating these concepts is the notion of a manifold in the space of events, which will allow us to define the distance between an event and a manifold, as well as the point of closest approach on a manifold. Since fixed-order perturbation theory works with a definite number of particles, an important type of manifold will be the idealized massless $N$-particle manifold:

$$\mathcal{P}_N = \left\{ \sum_{i=1}^{N} E_i \, \delta(\hat{n} - \hat{n}_i) \ \middle| \ E_i \geq 0 \right\}, \tag{1.5}$$

which, intuitively, is the set of all possible events with $N$ massless particles. Note that $\mathcal{P}_N \supset \mathcal{P}_{N-1} \supset \cdots \mathcal{P}_2 \supset \mathcal{P}_1 \supset \mathcal{P}_0$ via soft and collinear limits, so that the idealized $N$-particle manifold contains each manifold of smaller particle multiplicity.

The key concepts unified in this paper are outlined and summarized in Table 1. In Sec. 2, we discuss observables as functions defined on event radiation patterns and IRC safety as smoothness in the space of energy flows. Colloquially, the label "IRC safe" indicates that an observable should be well-defined and calculable in perturbation theory [27, 28] due to its robustness to long-distance effects (e.g. hadronization in the case of QCD). This "perturbatively accessible" IRC safety is traditionally connected to the observable being "insensitive" to the addition of low energy particles or collinear splittings of particles [29–36]. Here, we refine the definition of IRC safety and clarify when discontinuities in an observable spoil its perturbative calculability. Critical to our formulation is the notion of continuity with respect to the metric topology provided by the EMD:

**Definition 1.** *An observable $\mathcal{O}$ is EMD continuous at an event $\mathcal{E}$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all events $\mathcal{E}'$:*

$$\mathrm{EMD}(\mathcal{E}, \mathcal{E}') < \delta \quad \implies \quad |\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| < \epsilon. \tag{1.6}$$

We argue that IRC safety is EMD continuity everywhere except a negligible set of events, where a negligible set is one that contains no EMD balls of non-zero radius. Using the EMD provides a definition of IRC safety that does not refer to particles directly, which circumvents many pathologies of previous definitions. We argue that observables that are calculable in fixed-order perturbation are exactly those that satisfy a slightly stronger continuity condition known as Hölder continuity [62, 63], which restricts the types of divergences that can appear in the distribution of an observable [32, 36]. Fascinatingly, this framework naturally accommodates Sudakov-safe observables [37–39] as those that are IRC safe but fail to satisfy EMD Hölder continuity on a non-negligible subset of some $\mathcal{P}_N$ (where a non-negligible subset of

| Sec. | Concept | Equation | Illustration |
|------|---------|----------|--------------|
| 2 | **Infrared and Collinear Safety** [27–39] | $\mathrm{EMD}(\mathcal{E}, \mathcal{E}') < \delta \implies$ $\|\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')\| < \epsilon$ |  |
| 3<br>3.1<br>3.2 | **Observables**<br>Event Shapes [40–45]<br>Jet Shapes [46–48] | $\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}')$ |  |
| 4<br>4.1<br>4.2 | **Jets**<br>Cone Finding [49, 50]<br>Seq. Rec. [51–54] | $\mathcal{J}(\mathcal{E}) = \arg\min_{\mathcal{J} \in \mathcal{P}_N} \mathrm{EMD}(\mathcal{E}, \mathcal{J})$ |  |
| 5 | **Pileup Subtraction** [55–61] | $\mathcal{E}_C(\mathcal{E}, \rho) = \arg\min_{\mathcal{E}' \in \Omega} \mathrm{EMD}(\mathcal{E}, \mathcal{E}' + \rho\,\mathcal{U})$ |  |
| 6 | **Theory Space** | $\mathcal{T}(\mathcal{E}) = \sum_{i=1}^{N} \sigma_i\, \delta(\mathcal{E} - \mathcal{E}_i)$ |  |

**Table 1**: Concepts from quantum field theory and collider physics, unified in this paper as geometric and topological constructions in the space of events. In Sec. 2, IRC safety is identified as continuity in this space. In Sec. 3, many classic collider observables are shown to be the shortest distance between the event and a manifold of events. In Sec. 4, popular jet algorithms are derived by projecting the event onto manifolds of $N$-particle events. In Sec. 5, common pileup mitigation strategies are cast as transporting away uniform radiation. In Sec. 6, a space of theories is developed using a distance between event distributions.

$\mathcal{P}_N$ is one that has measure in $\mathcal{P}_N$). This suggests, in agreement with Ref. [39], that Sudakov safe observables are indeed perturbatively calculable once properly regulated.

In Sec. 3, we highlight that many well-known collider observables can be viewed as the distance of closest approach between an event and a manifold of events. Many of the observables we consider can be exactly cast as:

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}'), \tag{1.7}$$

for particular choices of the manifold $\mathcal{M}$ and parameters $\beta$ and $R$. Observables that have the form of Eq. (1.7) include thrust [40, 41], spherocity [42], (recoil-free) broadening [43], and $N$-jettiness [44]. Particularly interesting is the event isotropy, recently proposed in Ref. [45], which was inspired by EMD geometry and is directly based on optimal transport. This geometric framework also includes jet substructure observables such as jet angularities [46] and $N$-subjettiness [47, 48].

In Sec. 4, we demonstrate how jet finding can be phrased in our geometric language. Intuitively, a jet algorithm "approximates" an $M$-particle event with $N < M$ objects called jets. To phrase this geometrically, we are interested in the point of closest approach in $\mathcal{P}_N$ to our event, allowing us to define jets as:

$$\mathcal{J}(\mathcal{E}) = \underset{\mathcal{J} \in \mathcal{P}_N}{\arg\min} \ \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{J}), \tag{1.8}$$

where $\mathcal{J}$ is the collection of $N$ jets corresponding to the event $\mathcal{E}$. Many common jet finding algorithms can be derived in full detail from Eq. (1.8). For instance, we show that jets defined by Eq. (1.8) are precisely those found by XCone [49, 50], where $\beta$ is the angular weighting exponent and $R$ is the jet radius. Also, several popular sequential clustering algorithms and recombination schemes, such as $k_T$ clustering [51, 52] with winner-take-all recombination [43, 53, 54], can be exactly obtained by iterating Eq. (1.8) with $N = M - 1$ for various $\beta$. It is satisfying that a rich diversity of jet algorithms can be concisely encoded using event geometry, and we find that several new schemes not previously appearing in the literature naturally emerge.

In Sec. 5, we connect several pileup mitigation strategies to optimal transport through the EMD. There is a long-established relationship between pileup subtraction and geometric concepts [55–61]. Since pileup is reasonably modeled as uniform contamination in rapidity and azimuth [60], we phrase pileup subtraction as removing a uniform distribution of radiation from the event using optimal transport. Intuitively, pileup mitigation finds the event that, when combined with an amount $\rho$ of uniform radiation $\mathcal{U}$, is closest to the given event:

$$\mathcal{E}_C(\mathcal{E}, \rho) = \underset{\mathcal{E}' \in \Omega}{\arg\min} \ \text{EMD}_{\beta}(\mathcal{E}, \mathcal{E}' + \rho\mathcal{U}), \tag{1.9}$$

yielding the pileup-corrected event $\mathcal{E}_C$. Here, $\Omega$ refers to the space of all possible energy flows and $\text{EMD}_{\beta}$ compares events of equal energy, as described at the beginning of Sec. 3. We demonstrate that Voronoi area subtraction [55, 56] and constituent subtraction [58] can be

|  | Energy Mover's Distance | Cross Section Mover's Distance |
|---|---|---|
| Symbol | EMD | $\Sigma$MD |
| Description | Distance between events | Distance between theories |
| Weight | Particle energies $E_i$ | Event cross sections $\sigma_i$ |
| Ground Metric | Particle distances $\theta_{ij}$ | Event distances $\text{EMD}(\mathcal{E}_i, \mathcal{E}_j)$ |

**Table 2**:  Comparing the constructions of EMD and $\Sigma$MD as optimal transport problems. Events are treated as energy-weighted angular distributions, whereas theories are treated as cross section-weighted event distributions. This connection allows us to bootstrap the EMD as a ground metric for the $\Sigma$MD to develop a rigorous notion of theory space.

phrased exactly as Eq. (1.9) in the small-pileup limit. Generalizing this to the large-pileup limit, we develop two new pileup subtraction schemes, Apollonius subtraction and iterated Voronoi subtraction, and discuss their prospects and potential advantages.

In Sec. 6, we introduce a distance between theories: the cross section mover's distance (stylized as $\Sigma$MD, using the typical greek letter for cross section). Here, a "theory" $\mathcal{T}$ is taken to be a distribution over (or collection of) events $\{\mathcal{E}_i\}$ weighted by cross sections $\{\sigma_i\}$:

$$\mathcal{T}(\mathcal{E}) = \sum_{i=1}^{N} \sigma_i \, \delta(\mathcal{E} - \mathcal{E}_i). \tag{1.10}$$

The $\Sigma$MD is formulated as an optimal transport problem with EMD as the ground metric and cross sections as the weights. The similarity of the constructions of EMD and $\Sigma$MD are highlighted in Table 2. Interestingly, we connect $\Sigma$MD to a recently proposed technique for probing jet modifications due to the quark-gluon plasma by comparing similar sets of events between proton-proton and heavy-ion collisions [64]. We also demonstrate that representative events can be identified by clustering using the $\Sigma$MD, analogously to how particles are clustered into jets. The $\Sigma$MD provides the foundation for a rigorous formulation of "theory space", quantifying how different two theories are based on all of their physically observable quantities simultaneously.

Our conclusions are presented in Sec. 7, where we also highlight the interesting and unique interplay between machine learning and the natural sciences in this story.

## 2 Infrared and collinear safety: Smoothness in the space of events

IRC safety is a central notion in collider physics because it indicates when an observable is robust to long distance effects and hence can be described in perturbation theory [27, 28] using a combination of fixed-order calculations and resummation. This insensitivity is frequently connected to the invariance of an observable under certain modifications of the event, namely soft and collinear splittings [29–31, 33–36].
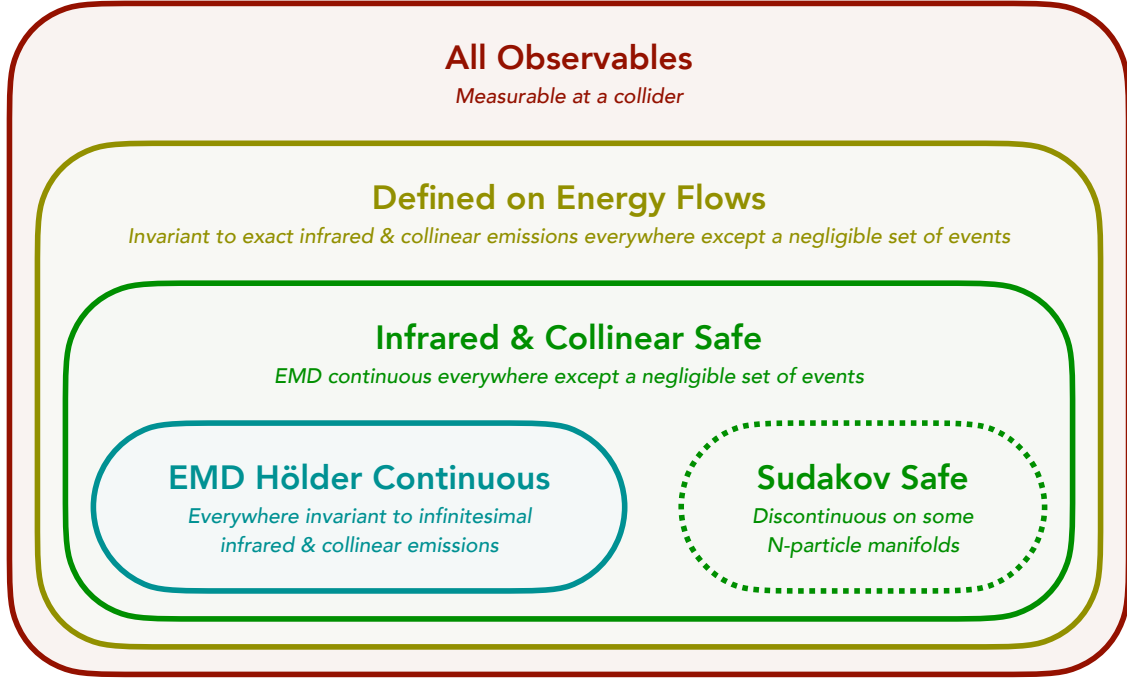
**Figure 2**: An illustration of the set of observables partitioned according to various IRC-invariance properties. Examples of observables in each category are listed in Table 3.

In this section, we review some of the common mathematical statements of this invariance that have appeared in the literature, with the goal of clarifying and categorizing their implications. We arrive at a simple, unified description of IRC safety and related concepts (including Sudakov safety) as statements about continuity in the space of energy flows. In Fig. 2, we show the breakdown of observables into broad classes according to our categorization. A few common examples of each category are given in Table 3.

## 2.1 Review of infrared and collinear invariance

The most straightforward statement of IRC invariance is that an observable $\mathcal{O}$ is unchanged under the addition of an exactly zero energy particle or an exactly collinear splitting [33]:

$$\text{Exact Infrared Invariance:} \quad \mathcal{O}(p_1^\mu, \ldots, p_M^\mu) = \mathcal{O}(0p_0^\mu, p_1^\mu, \ldots, p_M^\mu), \tag{2.1}$$

$$\text{Exact Collinear Invariance:} \quad \mathcal{O}(p_1^\mu, \ldots, p_M^\mu) = \mathcal{O}(\lambda p_1^\mu, (1-\lambda)p_1^\mu, \ldots, p_M^\mu), \tag{2.2}$$

for any soft momentum $p_0^\mu$ and collinear splitting fraction $\lambda \in [0, 1]$. These conditions correctly rule out some observables from having a perturbative description, such as the number of particles in an event, which change by a finite amount under any splitting. Exact IRC invariance, however, is not sufficiently restrictive to guarantee perturbative calculability of an observable. For instance, the number of calorimeter cells with non-zero energy is safe

| All Observables | Comments |
| --- | --- |
| Multiplicity ($\sum_i 1$) | IR unsafe and C unsafe |
| Momentum Dispersion [65] ($\sum_i E_i^2$) | IR safe but C unsafe |
| Sphericity Tensor [66] ($\sum_i p_i^\mu p_i^\nu$) | IR safe but C unsafe |
| Number of Non-Zero Calorimeter Deposits | C safe but IR unsafe |

| **Defined on Energy Flows** | |
| --- | --- |
| Pseudo-Multiplicity ($\min\{N \,|\, \mathcal{T}_N = 0\}$) | Robust to exact IR or C emissions |

| **Infrared & Collinear Safe** | |
| --- | --- |
| Jet Energy ($\sum_i E_i$) | Disc. at jet boundary |
| Heavy Jet Mass [67] | Disc. at hemisphere boundary |
| Soft-Dropped Jet Mass [38, 68] | Disc. at grooming threshold |
| Calorimeter Activity [69] ($N_{95}$) | Disc. at cell boundary |

| *Sudakov Safe* | |
| --- | --- |
| Groomed Momentum Fraction [39] ($z_g$) | Disc. on 1-particle manifold |
| Jet Angularity Ratios [37] | Disc. on 1-particle manifold |
| $N$-subjettiness Ratios [47, 48] ($\tau_{N+1}/\tau_N$) | Disc. on $N$-particle manifold |
| $V$ parameter [36] (Eq. (2.11)) | Hölder disc. on 3-particle manifold |

| **EMD Hölder Continuous Everywhere** | |
| --- | --- |
| Thrust [40, 41] | |
| Spherocity [42] | |
| Angularities [70] | |
| $N$-jettiness [44] ($\mathcal{T}_N$) | |
| $C$ parameter [71–74] | Resummation beneficial at $C = \frac{3}{4}$ |
| Linear Sphericity [72] ($\sum_i E_i n_i^\mu n_i^\nu$) | |
| Energy Correlators [36, 75–77] | |
| Energy Flow Polynomials [15, 17] | |

**Table 3**: Examples of well-known collider observables, along with their classification according to Fig. 2. The observables satisfy the conditions of all bold-faced categories above them in the table. Note that via our classification, Sudakov safe observables are IRC safe, since the discontinuities appear on $N$-particle manifolds which are negligible sets in the full space.

according to Eqs. (2.1) and (2.2), though it is highly sensitive to arbitrarily low-energy effects [69]. Similarly, the pseudo-multiplicity, which we define as the smallest $N$ that yields zero $N$-jettiness (see Sec. 3.2.2 below), is unchanged by exact infrared and collinear emissions,[3] but is highly sensitive to any emissions at finite energy or angle.

Another common statement of IRC invariance refines the concept by invoking the limit as particles become soft or collinear [30, 31, 34, 36]:

$$\text{Near Infrared Invariance:} \quad \mathcal{O}(p_1^\mu, \ldots, p_M^\mu) = \lim_{\epsilon \to 0} \mathcal{O}(\epsilon p_0^\mu, p_1^\mu, \ldots, p_M^\mu), \tag{2.3}$$

$$\text{Near Collinear Invariance:} \quad \mathcal{O}(p_1^\mu, \ldots, p_M^\mu) = \lim_{p_0^\mu \to p_1^\mu} \mathcal{O}(\lambda p_0^\mu, (1-\lambda)p_1^\mu, \ldots, p_M^\mu). \tag{2.4}$$

One issue with this definition is that many reasonable observables that have hard boundaries in phase space are excluded, such as jet kinematics due to sensitivity to particles on a jet boundary. Hybrid definitions mixing exact and near IRC invariance also appear in the literature but they suffer from the same pathologies. Another issue is that Eqs. (2.3) and (2.4) (and also Eqs. (2.1) and (2.2)) do not guarantee insensitivity to multiple soft or collinear splittings.

Several of these issues were previously identified in Ref. [36], which utilized a limit-based statement of IRC invariance, recognized the importance of allowing for multiple soft and collinear emissions, and allowed for exceptions on sets of measure zero. Despite noting that a rigorous mathematical definition of IRC safety would be desirable, Ref. [36] concluded that formulating one without pathologies was challenging and that a satisfactory definition had not yet been obtained. Here, we explore how the geometric picture provided by the EMD yields a natural and elegant way to phrase IRC safety and to control these various subtleties. This builds on the notion of "$C$-continuity" advocated for in Refs. [2, 7], which argue that the perturbative calculability of $C$-continuous observables can be seen by relating the energy flow to the stress-energy tensor of the underlying quantum field theory.

## 2.2 Infrared and collinear safety in the space of events

The EMD provides a natural language for understanding IRC-safe observables as continuous functions on the space of events. To make this precise, we first must understand which observables are well-defined functions of the energy flow.

We can show that observables that are defined on *all* energy flows are precisely those which have exact IRC invariance according to Eqs. (2.1) and (2.2). First, an observable is well defined on the space of energy flows if its value is the same on events that are zero EMD apart. The following lemma establishes the remaining connection to exact IRC invariance.

**Lemma 1.** *Two events are zero EMD apart if and only if they differ by zero energy emissions or exactly collinear splittings.*

*Proof.* Adding a zero energy particle or a collinear splitting to an event manifestly does zero energy moving, proving the forward direction. To prove the reverse direction, suppose that two events are zero EMD apart and take their energy flows to be:

$$\mathcal{E}(\hat{n}) = \sum_{i=1}^{M} E_i\, \delta(\hat{n} - \hat{n}_i), \qquad \mathcal{E}'(\hat{n}) = \sum_{j=1}^{M'} E_j'\, \delta(\hat{n} - \hat{n}_j'). \tag{2.5}$$

Since the EMD is a proper metric between energy flows, the identity of indiscernibles says that $\text{EMD}(\mathcal{E}(\hat{n}), \mathcal{E}'(\hat{n})) = 0$ implies $\mathcal{E}(\hat{n}) = \mathcal{E}'(\hat{n})$. For any direction $\hat{n}$ with at least one particle, either the sums of energies in that direction are equal between the two events or the particle has zero energy. In the first case, the events differ by exactly collinear splittings in that direction, and in the second case they differ by zero energy particles. □

By this lemma we see that exact IRC invariance ensures that we can write $\mathcal{O}(\mathcal{E})$ rather than $\mathcal{O}(p_1^\mu, \cdots, p_M^\mu)$ for an observable. As discussed in Sec. 2.1, exact IRC invariance is insufficient to guarantee IRC safety and we must formulate a stronger condition phrased in the geometric language of the space of events.

We propose that IRC safety is achieved by requiring an observable to be EMD continuous, in the sense of Definition 1, except possibly on a negligible set of events. We define a negligible set to be one that contains no EMD ball. The (open) EMD ball $B_r(\mathcal{E})$ around an event $\mathcal{E}$ is defined as all events within an EMD of $r > 0$:

$$B_r(\mathcal{E}) = \left\{ \mathcal{E}' \in \Omega \,\middle|\, \text{EMD}(\mathcal{E}, \mathcal{E}') < r \right\}, \tag{2.6}$$

where $\Omega$ is the space of all energy flows. Implicit in the above requirement is that an observable must be well defined on energy flows. Concretely, we state IRC safety as the following:

> **Infrared and Collinear Safety.** *An observable is IRC safe if it is EMD continuous for all energy flows, except potentially on a negligible set of events.*

This new formulation of IRC safety has many aspects of existing ideas of safety discussed in Sec. 2.2 wrapped into a concise and rigorous statement. It makes mathematically precise the intuitive notion that small perturbations in the energy flow of the event give rise to small perturbations in the observable. This notion of EMD continuity for IRC safe observables is illustrated in Fig. 3. The exception for negligible sets allows observables to be discontinuous in a way that affords them the opportunity to depend sharply on phase space but does not spoil their calculability. Calculability is a statement about integrability, and removing a negligible set of points from an integral cannot change its value.

To get some familiarity with this definition, consider additive IRC-safe observables, which are ubiquitous structures [17] that take the form $\mathcal{O}(\mathcal{E}) = \sum_{i=1}^{M} E_i f(\hat{n}_i)$ for an angular function
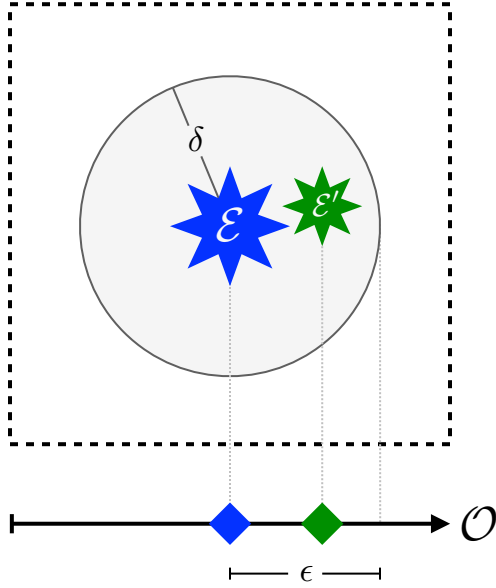
**Figure 3**: An illustration of IRC safety of an observable as continuity in the space of events. As formulated in Eq. (1.6), small perturbations to the event, as measured by EMD, yield small changes in the observable value.

$f$. One can prove that they are Lipschitz continuous in the space of events assuming $f$ is Lipschitz continuous [1], and therefore they naturally satisfy continuity according to the EMD. As a generalization of additive observables, energy flow networks [16] are a machine learning architecture that can approximate any IRC-safe observable through an additive IRC-safe latent space. As long as the activation functions are continuous almost everywhere, then the final energy flow network output will be IRC safe.

There are also observables that fail the criteria of Eqs. (2.3) and (2.4) for small sets of events but are safe according to our definition and are indeed calculable. The energy of a jet is a simple example where emissions on the jet boundary result in discontinuous behavior of the observable, but this discontinuity is integrable in fixed-order perturbation theory. A more complicated example is the invariant mass after soft drop grooming [38, 68]: for events on the threshold of having an emission dropped, tiny perturbations can give rise to discontinuously large changes in the observable. This issue, however, only occurs on a negligible set, satisfying our definition of safety and avoiding serious analytic pathologies [78–81]. Piecewise continuity does, however, complicate analyzing the nonperturbative corrections [82] and detector response [83, 84] of soft-dropped jet mass.

Our definition also includes observables that would sometimes not be called IRC safe since they do not have a well defined Taylor expansion in the small parameter of the theory (e.g. $\alpha_s$ for QCD). These observables are nevertheless perturbatively calculable, though methods beyond fixed-order perturbation theory may be required. The next subsections are devoted

to exploring which IRC-safe observables are calculable in fixed-order perturbation theory and which require additional techniques.

## 2.3 Calculability in fixed-order perturbation theory

IRC safety has long been connected with the notion of calculability order-by-order in perturbative quantum field theory. However, IRC safety according to our Definition 1 includes observables that are not calculable in fixed-order perturbation theory, which we explore further in the next subsection. Here, building off the work in Refs. [32, 36], we formulate the stronger notion of EMD Hölder continuity [62, 63] and argue that it is the appropriate condition to guarantee order-by-order perturbative control:

**Definition 2.** *An observable $\mathcal{O}$ is EMD Hölder continuous with exponent $\alpha \in (0, 1]$ at an event $\mathcal{E}$ if there exists $K > 0$ such that for all $\mathcal{E}'$ in some neighborhood of $\mathcal{E}$:*

$$|\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| \leq K \operatorname{EMD}(\mathcal{E}, \mathcal{E}')^{\alpha}. \tag{2.7}$$

Note that the case of $\alpha = 1$ corresponds to Lipschitz continuity at $\mathcal{E}$, and in general we have containment such that Hölder continuity with exponent $\alpha$ implies Hölder continuity with exponent $\beta$ if $\beta \leq \alpha$. EMD Hölder continuity effectively specifies that the $\delta$ in Definition 1 is no smaller than $\epsilon$ to some power (times a constant) for all points in a neighborhood of $\mathcal{E}$, and thus it is a stronger requirement than plain EMD continuity.

To connect to fixed-order perturbation theory, we state the following conjecture:

---

**Conjecture 1.** *An observable is calculable order-by-order in perturbation theory if it is EMD Hölder continuous on all but a negligible set of events in each $N$-particle manifold.*

---

This relation phrases the ideas of Ref. [32] and "Version 2" of the IRC safety definition of Ref. [36] in our geometric language via the EMD. While these criteria were originally formulated for the calculability of moments of an observable, they appear to also extend to the calculability of distributions of observables [85].

It is possible to demonstrate a precise equivalence between our Conjecture 1 and the following criteria of Ref. [32] regarding when the average value of an observable $\mathcal{O}$ is calculable in fixed-order perturbation theory:

$$\lim_{|\vec{p}_i| \to 0} \frac{\mathcal{O}(\vec{p}_1, \ldots, \vec{p}_i, \ldots, \vec{p}_M) - \mathcal{O}(\vec{p}_1, \ldots, \vec{p}_{i-1}, \vec{p}_{i+1}, \ldots, \vec{p}_M)}{|\vec{p}_i|^a} = 0, \tag{2.8}$$

$$\lim_{\theta_{ij} \to 0} \frac{\mathcal{O}(\vec{p}_1, \ldots, \vec{p}_i, \ldots, \vec{p}_j, \ldots, \vec{p}_M) - \mathcal{O}(\vec{p}_1, \ldots, \vec{p}_i + \vec{p}_j, \ldots, \vec{p}_{j-1}, \vec{p}_{j+1}, \ldots, \vec{p}_M)}{\theta_{ij}^b} = 0, \tag{2.9}$$

where the powers $a$ and $b$ are positive and the choices of $i$ and $j$ are arbitrary. Here, Eq. (2.8) is a statement of Hölder continuity in the energy of particle $i$, which implies ordinary soft

safety. Similarly, Eq. (2.9) is a statement of Hölder continuity in the angular distance between particles $i$ and $j$, which implies ordinary collinear safety. In these soft and collinear limits, $\mathrm{EMD}(\mathcal{E}, \mathcal{E}') \propto E_i$ and $\mathrm{EMD}(\mathcal{E}, \mathcal{E}') \propto \theta_{ij}$ respectively, and so Eqs. (2.8) and (2.9) can be phrased compactly as:

$$\lim_{\mathcal{E}' \to \mathcal{E}} \frac{\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')}{\mathrm{EMD}(\mathcal{E}, \mathcal{E}')^c} = 0. \tag{2.10}$$

for some positive exponent $c$. This is equivalent to the Hölder continuity of the observable $\mathcal{O}$ at $\mathcal{E}$ with some exponent $\alpha \geq c$, connecting the formulation of Ref. [32] to our conjecture.

Our Conjecture 1 also nicely connects to "Version 2" of the IRC safety definition in Ref. [36], which we restate here with a suggestive relabeling of the original notation. The criteria for fixed-order calculability of an observable in Ref. [36] are as follows:

> Ref. [36]: Given almost any fixed set of particles and any value $n$, then for any $\epsilon > 0$, however small, there should exist a $\delta > 0$ such that producing $n$ extra soft or collinear emissions, each emission being at a distance of no more than $\delta$ from the nearest particle, then the value of the observable does not change by more than $\epsilon$. Furthermore, there should exist a positive power $c$ such that for small $\epsilon$, $\delta^c$ can always be taken greater than $\epsilon$.

By equipping the space of events with these topological and geometric structures via EMD, our language provides a natural language to sharply mathematically formulate this discussion. The first sentence can be encoded as EMD continuity of the observable on all but a negligible set of events. The power relation between the $\epsilon$ and $\delta$ parameters is precisely captured by EMD Hölder continuity with some exponent $\alpha > c$, connecting to our Conjecture 1.

A variety of observables are considered in Ref. [36] at the boundary of perturbative calculability, which helpfully illustrate the various requirements in their definition.[4] An observable that is useful to consider is:

$$V(\mathcal{E}) = \mathcal{T}_2(\mathcal{E}) \left(1 + \frac{1}{\ln E(\mathcal{E})/\mathcal{T}_3(\mathcal{E})}\right), \tag{2.11}$$

where $\mathcal{T}_N$ are $N$-jettiness observables [44] discussed further in Sec. 3.1.4, and $E$ is the total energy of the event. We will refer to this observable as the "$V$ parameter". The double logarithmic structure of $\mathcal{T}_3$ spoils the integrability of $V$ at fixed order due to its behavior as $\mathcal{T}_3$ goes to zero [36], which occurs on the three-particle manifold $\mathcal{P}_3$. Nonetheless, this observable can be calculated using techniques beyond fixed-order perturbation theory, such as the Sudakov safety approach discussed in the next section.

The relation between our formalism and fixed-order perturbative calculability is phrased as a conjecture since additional subtleties or nuances about this type of calculability may emerge with future research. Nonetheless, it is very satisfying that our geometric language provides an efficient encapsulation and unification of the existing formulations of Refs. [32, 36]. In future work, it would be interesting to find a geometric phrasing of recursive IRC

---

[4]We thank Gavin Salam for discussions related to this point.

safety [36], which is a more restrictive condition than EMD Hölder continuity and relevant for understanding factorization and resummation. It would also be interesting to find a geometric phrasing of unsafe observables that can be nevertheless be computed with the help of non-perturbative fragmentation functions (see Ref. [86] for a broad class of such observables). We hope that further refinements and developments will benefit from and be enabled by the rigorous geometric and topological constructions we have introduced for the space of events via the EMD.

## 2.4 A refined understanding of Sudakov safety

Sudakov-safe observables [37–39] are an interesting class of observables that are not typically considered IRC safe because divergences may appear order by order in perturbation theory; this issue was originally pointed out in Ref. [57]. Nevertheless, the distribution for a Sudakov-safe observable $\mathcal{O}_s$ can be computed perturbatively by calculating its conditional distribution with an IRC-safe companion observable $\mathcal{O}_c$, resumming the $\mathcal{O}_c$ distribution, and then marginalizing over $\mathcal{O}_c$ to obtain a finite answer [39]:

$$p(\mathcal{O}_s) = \int \mathrm{d}\mathcal{O}_c \, p(\mathcal{O}_s|\mathcal{O}_c) \, p(\mathcal{O}_c). \tag{2.12}$$

The conditional probability $p(\mathcal{O}_s|\mathcal{O}_c)$ can either be computed in fixed-order perturbation theory or it can be further resummed to obtain a more accurate prediction for $p(\mathcal{O}_s)$.

Here, we interpret Sudakov-safe observables as observables that *are* IRC safe according to our definition but may be EMD (Hölder) discontinuous on sets with non-zero measure when restricted to some idealized massless $N$-particle manifold $\mathcal{P}_N$, defined in Eq. (1.5). The relevant manifolds are the $N$-particle manifolds since these contain the infrared singular regions of massless gauge theories, namely configurations that differ by soft and collinear splittings. The IRC safety of an observable according to our definition guarantees that any potentially problematic energy flows are infinitesimally close to energy flows for which the observable is well defined. The strategy in Eq. (2.12) also enables the computation of observables such as the $V$ parameter in Eq. (2.11), which are EMD continuous everywhere but exhibit Hölder discontinuities on sets with non-zero measure in $\mathcal{P}_N$ and are therefore incalculable with fixed-order perturbation theory alone.

It is instructive to make a connection to practical methods of computing Sudakov-safe observables. In a quantum field theory of massless particles, the cross section to produce events with exactly $N$ particles is zero (i.e. the naive $S$-matrix is zero), and such theories ultimately yield smooth predictions in the space of events. Hence, divergences that appear in the calculation of such an observable in a fixed-order expansion can be regulated by a joint, all-orders calculation of the observable and the distance from the problematic manifold $\mathcal{P}_N$. This is precisely the strategy represented by Eq. (2.12), though Ref. [39] did not provide a generic method to identify the companion observable $\mathcal{O}_c$. In Sec. 3, we will establish that the distance from an event to the manifold $\mathcal{P}_N$ is precisely $N$-(sub)jettiness [44, 47, 48],

suggesting that they are universal companion observables for the calculation of Sudakov-safe observables, in a similar spirit to Refs. [87–89].

It is worth mentioning that, even if an observable is EMD Hölder continuous everywhere, resummation along the lines of Eq. (2.12) may still be beneficial for making reliable predictions. The $C$-parameter [71–73] is an example of an EMD Hölder continuous observable, yet its fixed-order perturbative distribution exhibits discontinuous behavior at $C = \frac{3}{4}$ [74]. This perturbative discontinuity can be smoothed through soft-gluon resummation, and such techniques are relevant for other observables that exhibit Sudakov shoulder behavior [90]. This is different, however, from Sudakov-safe observables, where the observable itself (and not just its distribution) is ill-defined on some $\mathcal{P}_N$.

To summarize, our definition of IRC safety does includes Sudakov-safe observables, but we argue that this is appropriate since such observables are indeed perturbatively accessible via regulation with $N$-(sub)jettiness. This motivates the following conjecture:

---

**Conjecture 2.** *An observable is perturbatively calculable, using a combination of fixed-order and resummation techniques, if it is IRC safe according to the definition in Sec. 2.2.*

---

Proving this conjecture, or finding a counterexample, would shed considerable light on the structure of perturbative quantum field theory. Of course, even if an observable is perturbatively calculable, it may suffer from large non-perturbative or detector corrections, and it may be helpful to use the $K$ and $\alpha$ parameters in Eq. (2.7) to assess the sensitivity of observables to long-distance effects.

## 3 Observables: Distances between events and manifolds

In this section, we show that a number of event-level and jet substructure observables can be identified as geometric quantities in the space of events. Broadly speaking, the observables we consider take the general form of a distance between an event and a manifold, as in Eq. (1.7). The illustration in Fig. 4 shows an observable as a distance between geometric objects in the space of events. While not all IRC-safe observables can be written in this way, a remarkably large family of classic observables take precisely this geometric form. We will work with unnormalized observables here, but normalized versions can be obtained by dividing by the total energy (or transverse momentum in the hadronic case).

We begin by discussing thrust and spherocity, where the manifold is the set of all back-to-back two-particle events. To understand (recoil-free) broadening, we expand the manifold to all two-particle events, beyond just back-to-back configurations. Then, to connect to $N$-jettiness, we utilize the idealized $N$-particle manifold defined in Eq. (1.5). Our geometric language gives clear and intuitive explanations of what physics these observables probe and why they take the forms that they do. While these EMD formulations do not necessarily lead
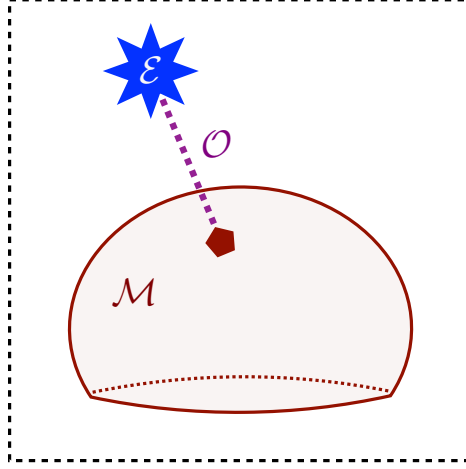
**Figure 4**: An illustration of an observable $\mathcal{O}$ as the distance of closest approach, as measured by the EMD, between the event $\mathcal{E}$ and a manifold $\mathcal{M}$ of events. Many classic collider observables fit into this precise form, stated in Eq. (1.7), with particular choices of manifold.

to practical computational improvements, we do highlight ways to speed up the numerical evaluation of event isotropy using techniques from the optimal transport literature. Finally, we identify jet angularities and $N$-subjettiness as jet substructure observables obeying similar principles at the level of jets.

For most of the observables in this section, the $R$ parameter is not needed, in which case we define a notion of EMD relevant for comparing events with equal energies:

$$\text{EMD}_\beta(\mathcal{E}, \mathcal{E}') = \lim_{R \to \infty} R^\beta \, \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}'). \tag{3.1}$$

This only has a finite limit if $\mathcal{E}$ and $\mathcal{E}'$ have the same total energy, which is a useful property to simplify our analysis. Explicitly, when comparing events with equal energy, this EMD simplifies to:

$$\text{EMD}_\beta(\mathcal{E}, \mathcal{E}') = \min_{\{f_{ij} \geq 0\}} \sum_{i=1}^{M} \sum_{j=1}^{M'} f_{ij} \theta_{ij}^\beta, \tag{3.2}$$

$$\sum_{i=1}^{M} f_{ij} = E_j', \qquad \sum_{j=1}^{M'} f_{ij} = E_i, \qquad \sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij} = \sum_{i=1}^{M} E_i = \sum_{j=1}^{M'} E_j'. \tag{3.3}$$

This will be the precise notion of EMD we use when the $R$ subscript is suppressed.

In Table 4, we summarize some of the observables considered below and their geometric interpretations. In Fig. 5, we illustrate the geometric construction of many of these observables, which we will explore in detail below.

| Name | | $\beta$ | Manifold $\mathcal{M}$ |
|---|---|---|---|
| $\mathcal{O}(\mathcal{E}) = \min\limits_{\mathcal{E}'\in\mathcal{M}} \mathrm{EMD}_\beta(\mathcal{E},\mathcal{E}')$ | | | |
| Thrust | $t(\mathcal{E})$ | 2 | $\mathcal{P}_2^{\mathrm{BB}}$: 2-particle events, back to back |
| Spherocity | $\sqrt{s(\mathcal{E})}$ | 1 | $\mathcal{P}_2^{\mathrm{BB}}$: 2-particle events, back to back |
| Broadening | $b(\mathcal{E})$ | 1 | $\mathcal{P}_2$: 2-particle events |
| $N$-jettiness | $\mathcal{T}_N^{(\beta)}(\mathcal{E})$ | $\beta$ | $\mathcal{P}_N$: $N$-particle events |
| Isotropy | $\mathcal{I}^{(\beta)}(\mathcal{E})$ | $\beta$ | $\mathcal{M}_{\mathcal{U}}$: Uniform events |
| Jet Angularities | $\lambda_\beta(\mathcal{J})$ | $\beta$ | $\mathcal{P}_1$: 1-particle jets |
| $N$-subjettiness | $\tau_N^{(\beta)}(\mathcal{J})$ | $\beta$ | $\mathcal{P}_N$: $N$-particle jets |

**Table 4**: Observables as the EMD between the event $\mathcal{E}$ and a manifold $\mathcal{M}$, using the EMD definition in Eq. (3.2). Several of these observables are illustrated in Fig. 5. Here, we consider only the "recoil-free" versions of these observables.



(a)   (b)
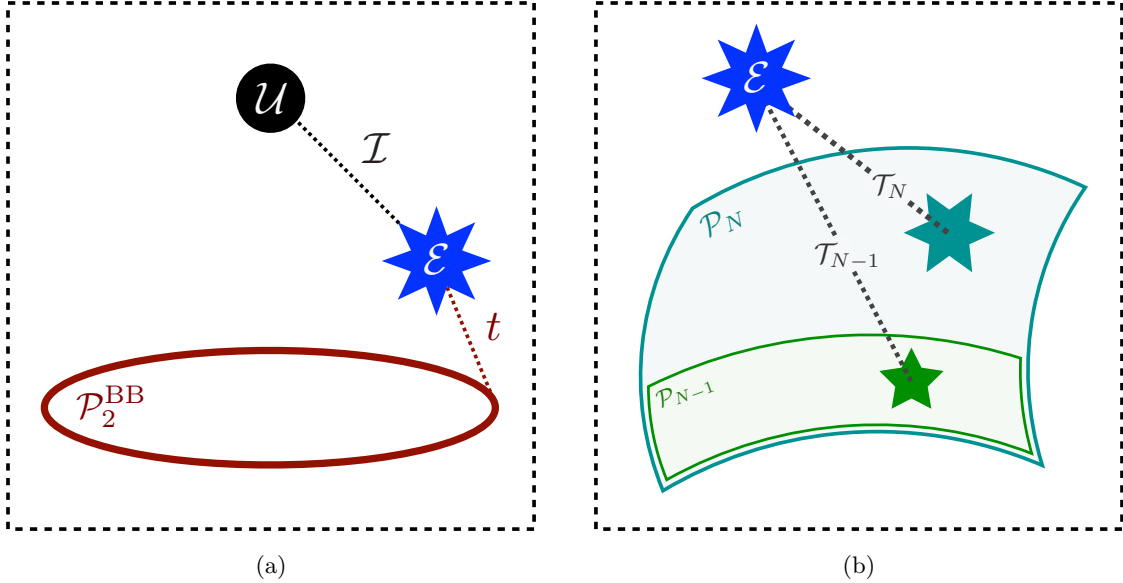
**Figure 5**: An illustration of a variety of observables as distances between an event $\mathcal{E}$ and various manifolds in the space of events, as summarized in Table 4. (a) Thrust $t$ is the smallest distance from the event to the manifold $\mathcal{P}_2^{\mathrm{BB}}$ of two-particle back-to-back events, while event isotropy $\mathcal{I}$ is the distance to the uniform event $\mathcal{U}$. (b) $N$-jettiness observables $\mathcal{T}_N$ are the smallest distances from the event to the $N$-particle manifolds $\mathcal{P}_N$.

### 3.1 Event-level observables

#### 3.1.1 Thrust

Thrust is an observable that quantifies the degree to which an event is pencil-like [40, 41, 91]. It has been experimentally measured [92–104] and theoretically calculated [105–110] in detail for electron-positron collisions. Thrust seeks to find an axis $\hat{n}$ (the "thrust axis") such that most of the radiation lies in the direction of either $\hat{n}$ or $-\hat{n}$; i.e. it maximizes the amount of radiation *longitudinal* to the thrust axis. While a variety of conventions for defining thrust exist, here we use the following dimensionful definition:

$$t(\mathcal{E}) = 2 \min_{\hat{n}} \sum_{i=1}^{M} |\vec{p}_i|(1 - |\vec{n}_i \cdot \hat{n}|), \tag{3.4}$$

where $\hat{n}_i = \vec{p}_i/|\vec{p}_i|$ and other definitions follow by simple rescalings. A thrust value of zero corresponds to an event consisting of two back-to-back prongs, while its maximum value of the total energy corresponds to a perfectly spherical event.

Interestingly, the value of thrust in Eq. (3.4) is equivalent to the cost of an optimal transport problem. This connection will allow us to cast thrust as a simple geometric quantity written in terms of the EMD. Using $E_i = |\vec{p}_i|$ for massless particles and writing out the absolute value, we can cast Eq. (3.4) as:

$$t(\mathcal{E}) = 2 \min_{\hat{n}} \sum_{i=1}^{M} E_i \min(1 - \hat{n}_i \cdot \hat{n}, \, 1 + \hat{n}_i \cdot \hat{n}). \tag{3.5}$$

For a fixed $\hat{n}$, the summand in Eq. (3.5) is the transportation cost to move particle $i$ to the closer of $\hat{n}$ or $-\hat{n}$ with an angular measure of $\theta_{ij}^2 = 2n_i^\mu n_{j\mu} = 2(1 - \hat{n}_i \cdot \hat{n}_j)$. The sum is then the EMD between the event and a two-particle event consisting of back-to-back particles directed along $\hat{n}$, where the energy of each of the two particles is equal to the total energy in the corresponding hemisphere. The minimization over $\hat{n}$ is equivalent to a minimization over all such two-particle events.

Thus, thrust is our first example of an observable that can be cast in the form of Eq. (1.7). First, we define the manifold of back-to-back two-particle events:

$$\mathcal{P}_2^{\text{BB}} = \left\{ \sum_{i=1}^{2} E_i \, \delta(\hat{n} - \hat{n}_i) \; \middle| \; E_i \geq 0, \; \hat{n}_1 = -\hat{n}_2 \right\}. \tag{3.6}$$

Then, using the notation of Eq. (3.2) with $\beta = 2$,[5] thrust is the smallest EMD from the event to the $\mathcal{P}_2^{\text{BB}}$ manifold:

$$\boxed{t(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_2^{\text{BB}}} \text{EMD}_2(\mathcal{E}, \mathcal{E}'),} \tag{3.7}$$

---

[5]As mentioned in footnote 1, strictly speaking only the square root of $\text{EMD}_2$ is a proper metric. Because the square root is a monotonic function, though, this has no impact on the interpretation of thrust as an optimal transport problem.

where the minimization is carried out over all back-to-back two-particle configurations.

Because of the $R \to \infty$ limit in Eq. (3.1), the optimal back-to-back configuration is guaranteed to have the same total energy as the event $\mathcal{E}$, as desired. Note that even if this analysis is carried out in the center-of-mass frame, the optimal back-to-back configuration will generically not be at rest, since it involves two massless particles with different energies.[6] This suggests a possible variant of thrust where one restricts the two-particle manifold to only include events that are physically accessible, either by forcing $E_1 = E_2$ or by considering massive particles as in App. A.

### 3.1.2 Spherocity

Spherocity is an observable that also probes the jetty nature of events [42]. It seeks to find an axis that minimizes the amount of radiation in the event *transverse* to it according to the following criterion:

$$s(\mathcal{E}) = \min_{\hat{n}} \left( \sum_{i=1}^{M} E_i |\vec{n}_i \times \hat{n}| \right)^2 , \tag{3.8}$$

where the original definition of spherocity is related to this by an overall rescaling. In the small $s$ limit, where the event configurations are back to back, we can write $|\vec{n}_i \times \hat{n}| \simeq \sqrt{2(1 - |\hat{n}_i \cdot \hat{n}|)}$ and obtain:

$$s(\mathcal{E}) \simeq \min_{\hat{n}} \left( \sum_{i=1}^{M} E_i \sqrt{2(1 - |\hat{n}_i \cdot \hat{n}|)} \right)^2 . \tag{3.9}$$

We focus on this limiting form for the following discussion.

Similar to the case of thrust, we can identify the spherocity expression to be minimized as an optimal transport problem. For a fixed $\hat{n}$, the summand in Eq. (3.9) is the cost to transport particle $i$ to the closer of $\hat{n}$ or $-\hat{n}$ with an angular measure of $\theta_{ij} = \sqrt{2 n_i^\mu n_{j\mu}}$.[7] The sum is once again the EMD from the event to the manifold of back-to-back events, with the minimization over $\hat{n}$ interpreted as a minimization over the manifold.

Spherocity, in the appropriate limit, is therefore the square of the smallest EMD (with $\beta = 1$) from the event to the manifold $\mathcal{P}_2^{\mathrm{BB}}$ from Eq. (3.6):

$$\boxed{\sqrt{s(\mathcal{E})} = \min_{\mathcal{E}' \in \mathcal{P}_2^{\mathrm{BB}}} \mathrm{EMD}_1(\mathcal{E}, \mathcal{E}').} \tag{3.10}$$

Through this lens, spherocity differs from thrust (besides the overall exponent) solely in the angular weighting factor: $\beta = 1$ for spherocity and $\beta = 2$ for thrust. One could continue in this direction, defining the distance of closest approach for general $\beta$. (This is related to the event shape angularities [70], with a key difference being that angularities are traditionally measured with respect to the thrust axis.) Instead, we now turn towards enlarging the manifold itself.

---

[6]We thank Samuel Alipour-fard for discussions related to this point.

[7]In fact, Eq. (3.8) is already an optimal transport problem, using $\theta_{ij} = \sin \Omega_{ij}$, where $\Omega_{ij}$ is the opening angle between particles $i$ and $j$. This has the same small angle behavior as $\theta_{ij} = 2 \sin \frac{\Omega_{ij}}{2}$ from Eq. (1.4).

### 3.1.3 Broadening

Recoil-free broadening [43] is an observable that is sensitive to two-pronged events that are not precisely back-to-back jets. Here we focus on recoil-free broadening, to be distinguished from the original jet broadening [111–113] which is defined in terms of the thrust axis.[8] It differs from spherocity only in that it minimizes the same quantity over two "kinked" axes that need not be antipodal. Though subtle, this difference gives rise to very important theoretical differences between broadening and spherocity in the treatment of soft recoil effects [114], as discussed extensively in Ref. [43].

Here, we use the following definition of broadening:

$$b(\mathcal{E}) = \min_{\hat{n}_L, \hat{n}_R} \sum_{i=1}^{M} E_i \min(\theta_{iL}, \theta_{iR}), \qquad (3.11)$$

where $\theta_{iL}$ and $\theta_{iR}$ are the angular distances between particle $i$ and $\hat{n}_L$ and $\hat{n}_R$, respectively. The fact that $\hat{n}_L$ and $\hat{n}_R$ are minimized separately (rather than $\hat{n}_L = -\hat{n}_R$) is the key distinction between recoil-free broadening and previous observables. For a fixed $\hat{n}_L$ and $\hat{n}_R$, the summand in Eq. (3.11) is the cost to transport particle $i$ to the closer of $\hat{n}_L$ or $\hat{n}_R$ with an angular measure of $\theta_{ij} = \sqrt{2n_i^\mu n_{j\mu}}$. The sum is then the EMD from the event to the manifold of all two-particle events, which need not be back-to-back, namely $\mathcal{P}_2$ from Eq. (1.5). The minimization over $\hat{n}_L$ and $\hat{n}_R$ is then interpreted as a minimization over this manifold.

Thus, broadening is the smallest EMD with $\beta = 1$ from the event to $\mathcal{P}_2$:

$$\boxed{b(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_2} \mathrm{EMD}_1(\mathcal{E}, \mathcal{E}').} \qquad (3.12)$$

The geometrical formulation of broadening in Eq. (3.12) differs from that of spherocity in Eq. (3.10) only in that it does not restrict the manifold to back-to-back configurations.This distinction is important to extend these ideas beyond the two-particle manifold.

### 3.1.4 $N$-jettiness

$N$-jettiness [44] (see also Ref. [115]) is an observable that partitions an event into $N$ jet regions and, for hadronic collisions, a beam region. Without a beam region, it is defined based on a minimization procedure over $N$ axes:

$$\mathcal{T}_N^{(\beta)} = \min_{\hat{n}_1, \cdots, \hat{n}_N} \sum_{i=1}^{M} E_i \min\left(\theta_{i1}^\beta, \theta_{i2}^\beta, \cdots, \theta_{iN}^\beta\right), \qquad (3.13)$$

where $\theta_{i1}$ through $\theta_{iN}$ are the angular distances between particle $i$ and axes $\hat{n}_1$ through $\hat{n}_N$, respectively.

---

[8]There is an EMD-based definition of the original jet broadening, using the thrust axis defined by $\mathcal{E}_t = \arg\min_{\mathcal{E}' \in \mathcal{P}_2^{\mathrm{BB}}} \mathrm{EMD}_2(\mathcal{E}, \mathcal{E}')$. With modified angular measure and normalization, the original jet broadening with respect to the thrust axis is $b_t(\mathcal{E}) = \mathrm{EMD}_1(\mathcal{E}, \mathcal{E}_t)$. Note the two different values of $\beta$ in these expressions.

We immediately identify the summand as the cost of transporting particle $i$ to the nearest axis. For fixed $\hat{n}_1$ through $\hat{n}_N$, assigning the energy transported to each axis as the energy of that axis gives rise to an $N$-particle event. The expression to be minimized is then the EMD between the original event and that $N$-particle event. The minimization over $\hat{n}_1$ through $\hat{n}_N$ is interpreted as a minimization over all such $N$-particle events.

Therefore, $N$-jettiness is the smallest distance between the event and the manifold $\mathcal{P}_N$ of $N$-particle events. Equivalently, one can view it as the EMD to the best $N$-particle approximation of the event, and we return to this interpretation in Sec. 4.1. Thus, we have:

$$\boxed{\mathcal{T}_N^{(\beta)} = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_\beta(\mathcal{E}, \mathcal{E}').}$$
(3.14)

We see that $N$-jettiness generalizes the geometric interpretation of broadening to a general $N$-particle manifold and a general angular weighting exponent $\beta$.

For hadronic collisions, initial state radiation and underlying event activity require the introduction of a "beam" (or out-of-jet) region [44, 116, 117]. This can be accomplished via the introduction of a beam distance $\theta_{i,\text{beam}}$ into the minimization of Eq. (3.13). There are many possible beam measures [49, 118], including ones that involve optimizing over two beam axes $\hat{n}_a$ and $\hat{n}_b$. For simplicity, we focus on $\theta_{i,\text{beam}} = R^\beta$ which makes no explicit reference to the beam directions [48]. Dividing by an overall factor of $R^\beta$, this modified version of $N$-jettiness can be written as:

$$\mathcal{T}_N^{(\beta,R)} = \min_{\hat{n}_1,\cdots,\hat{n}_N} \sum_{i=1}^M E_i \min\left(1, \frac{\theta_{i1}^\beta}{R^\beta}, \frac{\theta_{i2}^\beta}{R^\beta}, \cdots, \frac{\theta_{iN}^\beta}{R^\beta}\right).$$
(3.15)

This definition of $N$-jettiness is similar to Eq. (3.13), though now a particle can be closer to the beam than to any axis. In this case, we say that the particle is transported to the beam and removed for a cost $E_i$. The summand is then the cost to transport the event to an $N$-particle event plus the cost of removing any particles beyond $R$ from any axes.

Remarkably, this precisely corresponds to the EMD when formulated for events of different total energy. Namely, $N$-jettiness with this beam region is simply the smallest distance between the event and the manifold of $N$-particle events, with $R$ smaller than the radius of the space:

$$\boxed{\mathcal{T}_N^{(\beta,R)} = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}').}$$
(3.16)

Particles removed by the optimal transport procedure are interpreted as being part of the beam region. This fact will also be relevant in Sec. 4.2 for understanding sequential recombination jet clustering algorithms as geometric constructions in the space of events.

### 3.1.5  Event isotropy

Our new geometric phrasing of these classic collider observables highlights the types of configurations that they are designed to probe. Specifically, Eq. (1.7) can be interpreted

as how similar an event is to the class of events on the manifold $\mathcal{M}$. This framework also suggests regions of phase space that are poorly resolved by existing observables and provides a prescription for developing new observables by identifying new manifolds of interest.

Event isotropy [45] is a recently-proposed observable that provides a clear example of this strategy. It is based on the insight that distances from the $N$-particle manifolds (such as thrust and $N$-jettiness) are not well-suited for resolving isotropic events with uniform radiation patterns. Having observables with sensitivity to isotropic events can, for instance, improve new physics searches for microscopic black holes or strongly-coupled scenarios. This motivates event isotropy, which is the distance between the event $\mathcal{E}$ and an isotropic event $\mathcal{U}$ of the same total energy:

$$\mathcal{I}^{(\beta)}(\mathcal{E}) = \mathrm{EMD}_\beta(\mathcal{E}, \mathcal{U}). \tag{3.17}$$

Since $\mathcal{E}$ and $\mathcal{U}$ have the same total energy by construction, it is natural to normalize event isotropy by the total energy to make it dimensionless. The analysis in Ref. [45] focused primarily on $\beta = 2$, though this approach can be extended to a general angular exponent. For practical applications, it is convenient to consider a manifold of quasi-isotropic events of the same total energy and then estimate event isotropy as the average EMD between an event and this manifold.

We can cast Eq. (3.17) into the form of Eq. (1.7) by introducing a manifold $\mathcal{M}_\mathcal{U}$ of uniform events with varying total energies:

$$\boxed{\mathcal{I}^{(\beta)}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}_\mathcal{U}} \mathrm{EMD}_\beta(\mathcal{E}, \mathcal{E}').} \tag{3.18}$$

The $R \to \infty$ limit in Eq. (3.1) enforces that the optimal isotropic approximation $\mathcal{U}$ has the same total energy as $\mathcal{E}$, as in the original event isotropy definition.

The particular notion of a uniform distribution depends on the collider context—spherical for electron-positron collisions and cylindrical or ring-like for hadronic collisions—with corresponding choices for the energy and angular measures. The case of ring-like isotropy at a hadron collider is particularly interesting, since there are known simplifications for one-dimensional circular optimal transport problems. For $\beta = 1$, ring-like event isotropy can be computed in $\mathcal{O}(M)$ runtime [119] and there are fast approximations for any $\beta \geq 1$ [119]. This is much faster than the generic $\mathcal{O}(M^3 \log M)$ expectation for EMD computations, motivating further studies of these one-dimensional geometries.

## 3.2 Jet substructure observables

### 3.2.1 Jet angularities

Jet angularities are the energy-weighted angular moments of radiation within a jet [46] (see also Refs. [43, 120, 121]). Here, we use the following definition of a recoil-free jet angularity:

$$\lambda_\beta(\mathcal{J}) = \min_{\hat{n}} \sum_{i=1}^{M} E_i \, \theta_i^\beta, \tag{3.19}$$

where $\theta_i$ is the angular distance between particle $i$ and an axis $\hat{n}$. The summand of an angularity is the EMD from the jet to the axis, so we can follow the analogous logic from our previous discussions of event shapes to reframe this observable in our geometric language. Specifically, the recoil-free angularities are the closest distance between the jet and the 1-particle manifold $\mathcal{P}_1$:

$$\lambda_\beta(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_1} \mathrm{EMD}_\beta(\mathcal{J}, \mathcal{J}'). \qquad (3.20)$$

One can alternatively consider a definition of angularities where $\theta_i$ is computed with respect to a fixed jet axis. In that case, the angularities are the EMD from the jet to a 1-particle configuration where the total energy of the jet is placed at the position of the desired axis.

### 3.2.2  $N$-subjettiness

$N$-subjettiness is a jet substructure observable that applies the ideas of $N$-jettiness at the level of jet substructure [47, 48]. $N$ axes are placed within the jet, with a penalty for having energy far away from any axis, and then the positions of the axes are optimized. The (dimensionful) $N$-subjettiness of a jet can be defined as follows:

$$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\hat{n}_1, \cdots, \hat{n}_N} \sum_{i=1}^{M} E_i \min\left(\theta_{i1}^\beta, \theta_{i2}^\beta, \cdots, \theta_{iN}^\beta\right), \qquad (3.21)$$

where $\theta_{i1}$ through $\theta_{iN}$ are the angular distances between particle $i$ and axes $\hat{n}_1$ through $\hat{n}_N$. The beam region is absent due to the fact that these observables are only defined using the particles already within an identified jet.

We can find a geometric interpretation for $N$-subjettiness by using the analogous discussion from $N$-jettiness in Sec. 3.1.4. $N$-subjettiness is the distance between the jet and the manifold of all $N$-particle jets:

$$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_N} \mathrm{EMD}_\beta(\mathcal{J}, \mathcal{J}'). \qquad (3.22)$$

As a limiting case, $N = 1$ corresponds to the jet angularities in Eq. (3.20).

In this way, we can view $N$-subjettiness values as "coordinates" for the space of jets, defined as distances from each of the $N$-particle manifolds, illustrated in Fig. 6. The $N$-subjettiness ratios $\tau_N/\tau_{N-1}$, used ubiquitously for jet substructure studies [122–124], are then the relative distances between the manifolds $\mathcal{P}_N$ and $\mathcal{P}_{N-1}$. This is also an interesting way to interpret existing constructions of observable bases using $N$-(sub)jettiness [125–127]; the fact that multiple $\beta$ values are typically needed for these constructions emphasizes that the choice of ground metric affects the geometry of the space induced by the EMD.
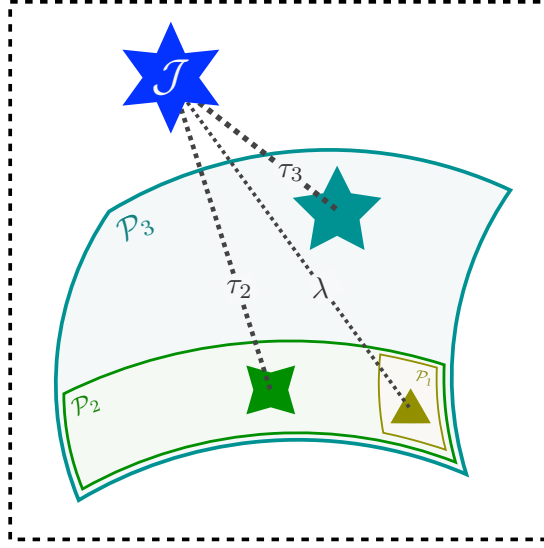
**Figure 6**: An illustration of $N$-subjettiness values as the smallest distances, as measured by EMD, between the event $\mathcal{E}$ and each of the $N$-particle manifolds $\mathcal{P}_N$. The jet angularities are the distances to the 1-particle manifold $\mathcal{P}_1$. These observables form a set of "coordinates" for the space.



(a)

(b)
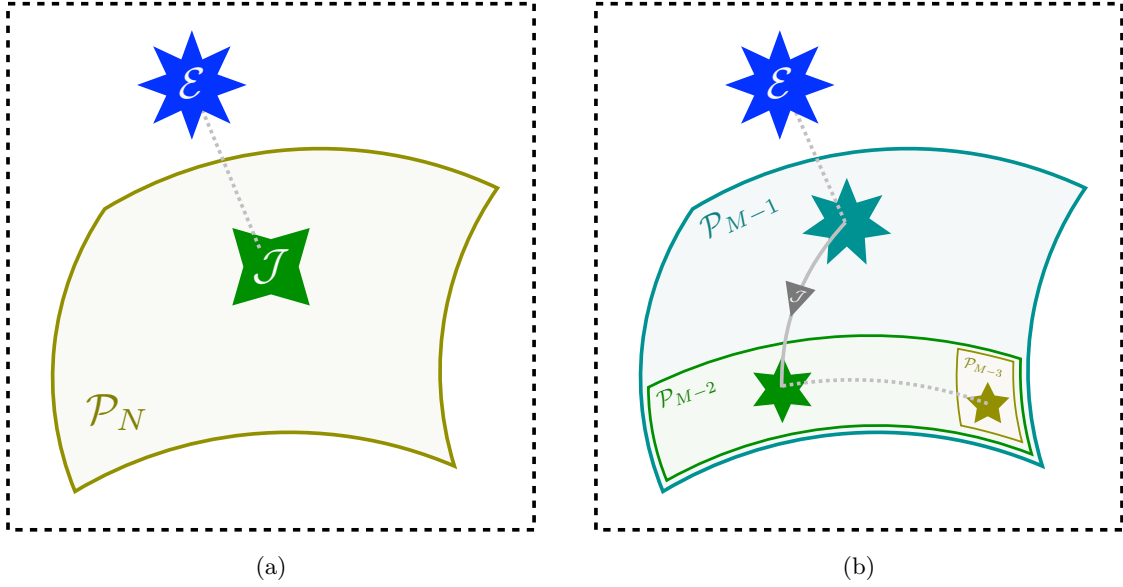
**Figure 7**: An illustration of jet clustering algorithms as projections to $N$-particle manifolds $\mathcal{P}_N$ in the space of events. (a) Exclusive cone finding algorithms yield $N$ jets as the closest $N$-particle approximation to the event, as measured by the EMD. (b) Sequential recombination algorithms iteratively find the best $(M-1)$-particle approximation to the $M$-particle event, either (dashed) merging two particles or (solid) removing a particle and calling it a jet.

# 4 Jets: The closest $N$-particle description of an $M$-particle event

In this section, we turn our attention to how jets are defined. We interpret two of the most common classes of jet algorithms as simple geometric constructions in the space of events. Intuitively, we find that jets are the best $N$-particle approximation to an $M$-particle event. Many existing techniques naturally emerge from this simple principle in fascinating ways.

First, we discuss exclusive cone finding, as this technique corresponds exactly to the intuition above that jets approximate the energy flow of an event using a smaller number of particles. Next, we show that many sequential recombination algorithms can be derived by iteratively approximating an $M$-particle event using $M-1$ particles. These jet-finding strategies are illustrated in Fig. 7 as projections to $N$-particle manifolds in the space of events.

## 4.1 General $N$: Exclusive cone finding

XCone [49, 50] is an exclusive cone finding algorithm that seeks to find jets by minimizing $N$-jettiness. It returns a fixed number of jets based on the parameters $N$ and $R$, in the same spirit as the exclusive version of the $k_t$ sequential recombination algorithm [51]. XCone proceeds by finding the $N$ axes that minimize $N$-jettiness as defined in Eq. (3.15):

$$\underset{\hat{n}_1,\cdots,\hat{n}_N}{\operatorname{argmin}} \sum_{i=1}^{M} E_i \min\left(1, \frac{\theta_{i1}^{\beta}}{R^{\beta}}, \frac{\theta_{i2}^{\beta}}{R^{\beta}}, \cdots, \frac{\theta_{iN}^{\beta}}{R^{\beta}}\right). \tag{4.1}$$

Together with the energy assigned to those axes, or equivalently the set of particles mapped to each axis, the $N$ axes from Eq. (4.1) define $N$ jets. The jet radius parameter $R$ controls which particles are not assigned to any jet (i.e. assigned to the beam region). Following the discussion in Sec. 3.1.4, Eq. (4.1) can be interpreted as finding the $N$-particle configuration that best approximates the event of interest.

In our geometric language, we can cast XCone as identifying the point of closest approach between an event $\mathcal{E}$ and the $N$-particle manifold $\mathcal{P}_N$:

$$\boxed{\mathcal{J}_{N,\beta,R}^{\text{XCone}}(\mathcal{E}) = \underset{\mathcal{J}\in\mathcal{P}_N}{\operatorname{argmin}} \operatorname{EMD}_{\beta,R}(\mathcal{E},\mathcal{J}).} \tag{4.2}$$

Different variants of XCone correspond to different choices for the energy weight $E_i$ and the angular measure $\theta_{ij}$ [49, 118], which in turn correspond to different choices for what defines the "best" $N$-particle approximation to an event.

As discussed in Ref. [128], there is a close relationship between exclusive cone finding algorithms, stable cone algorithms [129–131], and jet maximization algorithms [132–136]. For the choice of $\beta = 2$, the jet axis aligns with the jet momentum direction, which is known as the stable cone criterion [129, 130]. For $N = 1$, one can relate the optimization problem in Eq. (4.2) to maximizing a "jet function" over all possible partitions of an event into one in-jet region and one out-of-jet region [132]. Iteratively applying the $N = 1$ procedure is related to the SISCone algorithm with progressive jet removal [131]. All of these various algorithms can now be interpreted in our geometric picture as different ways to "project" the event $\mathcal{E}$ onto the $N$-particle manifold $\mathcal{P}_N$.

## 4.2   $N = M - 1$: Sequential recombination

Sequential recombination algorithms are a class of jet clustering algorithms that have seen tremendous use at colliders, particularly the anti-$k_t$ algorithm [137] which is the current default jet algorithm at the LHC. These methods utilize an interparticle distance $d_{ij}$, a particle-beam distance $d_{iB}$, and a recombination scheme for merging two particles. The algorithm proceeds iteratively by finding the smallest distance, combining particle $i$ and $j$ if it is a $d_{ij}$, or calling $i$ a jet and removing it from further clustering if it is a $d_{iB}$.

There are a variety of distance measures and recombination schemes that appear in the literature, many of which are implemented in the FASTJET library [138]. The most commonly used distance measures take the form:

$$d_{ij} = \min\left(E_i^{2p}, E_j^{2p}\right) \frac{\theta_{ij}^2}{R^2}, \qquad d_{iB} = E_i^{2p}, \tag{4.3}$$

where $p$ is an energy weighting exponent and $R$ is the jet radius. The exponent $p = 1$ corresponds to $k_t$ jet clustering [51, 52], $p = 0$ corresponds to Cambridge/Aachen (C/A) clustering [139, 140], and $p = -1$ corresponds to anti-$k_t$ clustering [137]. The recombination scheme determines the energy $E_c$ and direction $\hat{n}_c$ of the combined particle and typically takes the form:

$$E_c = E_i + E_j, \qquad \hat{n}_c = \frac{E_i^\kappa \, \hat{n}_i + E_j^\kappa \, \hat{n}_j}{E_i^\kappa + E_j^\kappa}, \tag{4.4}$$

where $\kappa = 1$ corresponds the $E$-scheme (most typically used), $\kappa = 2$ is the $E^2$-scheme [51, 141], and $\kappa \to \infty$ is the winner-take-all scheme [43, 53, 54]. In the $E$-scheme, the four-momenta of the two particles are simply added.[9] In the winner-take-all scheme, the direction is determined by the more energetic particle.

The conceptual and algorithmic richness of these different distance measures and recombination schemes arose from decades of phenomenological studies. Remarkably, many of these techniques naturally emerge from event space geometry, as finding the point on the $(M-1)$-particle manifold $\mathcal{P}_{M-1}$ that is closest to configuration $\mathcal{E}$ with $M$ particles. Note that the sequential recombination algorithms in Eqs. (4.3) and (4.4) depend on the two parameters $p$ and $\kappa$, whereas Eq. (4.6) depends only on $\beta$, so the logic below will only identify a one-dimensional family of jet algorithms, as summarized in Table 5.

To derive this connection between event geometry and sequential recombination, we need the following simple yet profound lemma, using the suggestive notation of $d_{iB}$ and $d_{ij}$ to refer to the EMD cost of rearrangement.

---

[9]One has to be a bit careful about the interpretation of jet masses in the $E$-scheme. In the discussion below, the combined particle is interpreted as a massless four-vector. For the angular distance in Eq. (1.4), the direction $\hat{n}_i$ is the same for massless and massive particles, so one can consistently assign the mass of the jet to be the invariant mass of the summed jet constituents. For the rapidity-azimuth distance typically used at hadron colliders, though, the rapidity of a particle depends on its mass, so one has to be careful about whether one is talking about a light-like jet axis or a massive jet when discussing the $E$-scheme. See further discussion in App. A.

| EMD$_{\beta,R}$ | Name | Measure $d_{ij}$ | $d_{iB}$ | Name | Scheme $\lambda^*$ |
|---|---|---|---|---|---|
| $0 < \beta < 1$ | Gen. $k_t$ | $\min(E_i, E_j)\frac{\theta_{ij}^\beta}{R^\beta}$ | $E_i$ | Winner-take-all | $\mathrm{argmin}(E_i, E_j)$ |
| $\beta = 1$ | $k_T$ | $\min(E_i, E_j)\frac{\theta_{ij}}{R}$ | $E_i$ | Winner-take-all | $\mathrm{argmin}(E_i, E_j)$ |
| $\beta = \frac{3}{2}$ | ? | $\frac{E_i E_j}{\sqrt{E_i^2 + E_j^2}}\frac{\theta_{ij}^{\frac{3}{2}}}{R^{\frac{3}{2}}}$ | $E_i$ | $E^2$-scheme | $\frac{E_j^2}{E_i^2 + E_j^2}$ |
| $\beta = 2$ | ? | $\frac{E_i E_j}{E_i + E_j}\frac{\theta_{ij}^2}{R^2}$ | $E_i$ | $E$-scheme | $\frac{E_j}{E_i + E_j}$ |
| $\beta \to \infty$ | C/A | $\frac{\theta_{ij}}{2R}$ | $1$ | ? | $\frac{1}{2}$ |

**Table 5**: Different sequential recombination measures $d_{ij}$ and recombination schemes $\lambda^*$ that emerge from an EMD formulation. A question mark indicates a method that, to our knowledge, does not yet appear in the literature. The traditional definitions of generalized $k_t$ and C/A require squaring $d_{ij}$ and $d_{iB}$. Note the factor of 2 in the C/A effective jet radius parameter.

**Lemma 2.** *As measured by the EMD, the closest $(M-1)$-particle event to an $M$-particle event has, without loss of generality, either:*

*(a) Two of the particles in the event merged together.*

*(b) One of the particles in the event removed.*

*Proof.* Removing a particle from the event has some EMD cost $d_{iB}$ and merging a pair of particles has a some EMD cost $d_{ij}$. To reduce the number of particles in the event by one, one can either remove a particle or merge two particles. Altering more than two particles by (re)moving fractions of additional particles always incurs additional EMD costs. If there are multiple pairs that are zero distance apart, then we can without loss of generality always choose to only merge one pair. $\qquad\square$

The two options in this lemma correspond precisely to the two possible actions at each stage of a sequential recombination algorithm. The EMD cost of removing a particle is always

$$d_{iB} = E_i. \tag{4.5}$$

If this is less than the cost of merging two particles together, then particle $i$ can be identified as a jet. For one step of a sequential recombination (SR) procedure applied to an event $\mathcal{E}$ with $M$ particles, we can express this step mathematically as:

$$\boxed{\mathcal{J}_{\beta,R}^{\mathrm{SR}}(\mathcal{E}) = \mathcal{E} - \underset{\mathcal{E}' \in \mathcal{P}_{M-1}}{\mathrm{argmin}}\ \mathrm{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}').} \tag{4.6}$$

In our geometric picture, if the $M$ particle event is "far away" from the $(M-1)$-particle manifold $\mathcal{P}_{M-1}$, then the projected difference is a jet.

On the other hand, if the cost of merging two particles is less than any of the particle energies, then the event is "close" to the $(M-1)$-particle manifold. Consider a pair of particles with energies $E_i$ and $E_j$ separated by a distance $\theta_{ij}$. To find the best $(M-1)$-particle approximation, we want to merge these two particles into one combined particle with energy $E_i + E_j$. Because the EMD is a metric, the optimal transportation plan must occur along a "geodesic" connecting the particles, with particle $i$ moving a distance $\lambda\,\theta_{ij}$ and particle $j$ moving a distance $(1-\lambda)\,\theta_{ij}$ for some $\lambda \in [0,1]$.[10] Minimizing this cost with respect to $\lambda$ yields both the cost of merging those two particles as well as the optimal recombination scheme with which to merge them. Because no energy is removed in this process, Eq. (4.6) yields a zero energy jet, which we can interpret as no jet being found at this step of the sequential recombination.

The cost of merging particles $i$ and $j$ depends on the jet radius parameter $R$ and angular exponent $\beta$:

$$d_{ij} = \min_{\lambda} \left[ E_i \left( \lambda \frac{\theta_{ij}}{R} \right)^{\beta} + E_j \left( (1-\lambda) \frac{\theta_{ij}}{R} \right)^{\beta} \right]. \tag{4.7}$$

For $\beta \leq 1$, the cost in Eq. (4.7) is minimized at the endpoints. This corresponds to moving the less energetic particle the entire distance $\theta_{ij}$ to the more energetic particle, which is the precisely behavior of the winner-take-all recombination scheme. For $\beta > 1$, the optimal value $\lambda^*$ can be found by differentiating Eq. (4.7) with respect to $\lambda$ and setting the result equal to zero. In general, the optimal recombination scheme has:

$$\begin{aligned}
0 < \beta \leq 1: \quad & \lambda^* = 1 \text{ if } E_i < E_j, \text{ else } 0, \\
\beta > 1: \quad & \lambda^* = \frac{1}{1 + \left( \frac{E_i}{E_j} \right)^{\frac{1}{\beta-1}}}.
\end{aligned} \tag{4.8}$$

To determine the actual cost, we substitute this $\lambda^*$ back into Eq. (4.7):

$$\begin{aligned}
\beta \leq 1: \quad & d_{ij} = \min(E_i, E_j) \frac{\theta_{ij}^{\beta}}{R^{\beta}}, \\
\beta > 1: \quad & d_{ij} = \frac{E_i E_j^{\frac{\beta}{\beta-1}} + E_i^{\frac{\beta}{\beta-1}} E_j}{\left( E_i^{\frac{1}{\beta-1}} + E_j^{\frac{1}{\beta-1}} \right)^{\beta}} \frac{\theta_{ij}^{\beta}}{R^{\beta}}.
\end{aligned} \tag{4.9}$$

If all $d_{ij}$ values in Eq. (4.9) are smaller than all particle energies in Eq. (4.5), then the optimal transportation plan is to merge particles $i$ and $j$.

In this way, Eq. (4.6) takes an $M$-particle event and returns a jet (with zero energy if no actual jet is found) plus the remaining $(M-1)$-particle approximation. This corresponds

---

[10]This linear decomposition of the distance does not hold for a general ground metric. However, it does hold when using the rapidity-azimuth distance, the opening-angle on the sphere, the small angle limit of Eq. (1.4), or the improved $\theta_{ij}$ distance with particle masses in App. A.

exactly to one step of a sequential clustering procedure. Iterating this procedure until $M = 1$, we derive a sequential recombination jet algorithm, where the jets correspond to all of the positive energy configurations obtained from Eq. (4.6).

Many existing methods reside within the simple framework of Eq. (4.6). For instance, $\beta = 1$ corresponds to $k_t$ jet clustering with winner-take-all recombination. The recombination scheme for $\beta = 2$ is the $E$-scheme, whereas for $\beta = \frac{3}{2}$ it is the $E^2$-scheme. Raising the distance measures to the $1/\beta$ power and taking the $\beta \to \infty$ limit, we obtain the C/A clustering metric, albeit with an effective jet radius that is twice the $R$ parameter. There are also a number of methods, indicated as question marks in Table 5, that emerge from this reasoning yet do not presently appear in the literature. Exploring these new methods is an interesting avenue for future work.

Intriguingly, in this geometric picture, the distance measure $d_{ij}$ and the recombination scheme $\lambda^*$ are paired by the $\beta$ parameter. A similar pairing was noted in Refs. [49, 142] in the context of choosing approximate axes for computing $N$-(sub)jettiness, and it would be interesting to explore the phenomenological implications of these paired choices for jet clustering. One sequential combination algorithm that does not appear is anti-$k_t$. Given that anti-$k_t$ is a kind of hybrid between sequential recombination and cone algorithms, there may be a way to combine the logic of Secs. 4.1 and 4.2 to find a geometric phrasing of anti-$k_t$. If successful, such a geometric construction would likely illuminate the difference between exclusive jet algorithms like XCone that find a fixed number of jets $N$ and inclusive jet algorithms like anti-$k_t$ that determine $N$ dynamically.
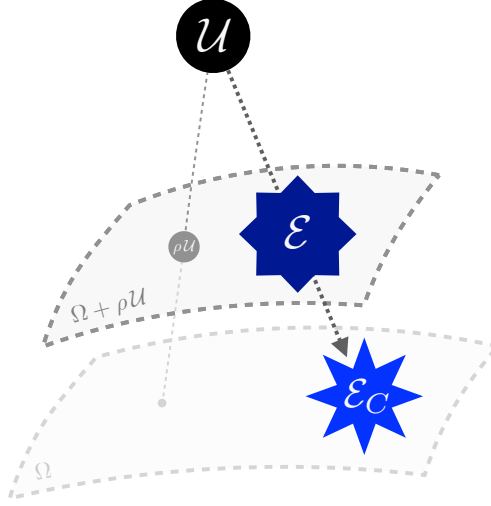
## 5   Pileup subtraction: Moving away from uniform events

The LHC era has brought with it new collider data analysis challenges. One notable example is pileup mitigation [60], removing the diffuse soft contamination from additional uncorrelated proton-proton collisions. The radiation from pileup interactions is approximately uniform in the rapidity-azimuth plane, and several existing pileup mitigation strategies seek to remove this uniform distribution of energy from the event [55, 58, 59, 143–148].

In this section, inspired by the approximate uniformity of pileup, we consider a class of pileup removal procedures that can be described as "subtracting" a uniform distribution of energy with density $\rho$, denoted $\rho \mathcal{U}$, from a given event. We take the pileup density per unit area $\rho$ to be given, for instance, by the area-median approach [55]. Given an event flow $\mathcal{E}$, the subtracted distribution $\mathcal{E} - \rho \mathcal{U}$ is typically not a valid energy flow, since the local energy density can go negative. Therefore, to implement this principle at the level of energy distributions, we turn this logic around and declare the corrected event $\mathcal{E}_C$ to be one that is as close as possible to the given event $\mathcal{E}$ when uniform radiation $\rho \mathcal{U}$ is added to it:

$$\mathcal{E}_C(\mathcal{E}, \rho) = \underset{\mathcal{E}' \in \Omega}{\arg \min} \ \mathrm{EMD}_\beta(\mathcal{E}, \mathcal{E}' + \rho \mathcal{U}). \tag{5.1}$$

Here, $\Omega$ refers to the complete space of energy flows, and the $R \to \infty$ limit of the EMD from Eq. (3.1) enforces that the corrected distribution $\mathcal{E}_C$ has the correct total energy.

**Figure 8**: A visualization of pileup subtraction in the space of events as moving away from uniform radiation. This proceeds by finding the event $\mathcal{E}_C$ that, when combined with uniform contamination $\rho\mathcal{U}$, is most similar to the given event $\mathcal{E}$. Different pileup mitigation strategies implement this removal in different ways. In the figure above, $\Omega$ refers to the space of all energy flows and $\Omega + \rho U$ is a subset of that space obtained by adding uniform contamination to every event configuration (shown as a separate manifold for ease of visualization).

As illustrated in Fig. 8, one can visualize Eq. (5.1) as a procedure that subtracts a uniform component from the energy flow. To make contact with existing techniques, we show that area-based Voronoi subtraction [55, 56, 138] and ghost-based constituent subtraction [58] can be cast in the form of Eq. (5.1) in the low-pileup limit. We then develop two new pileup mitigation techniques that have optimal transport interpretations even away from the low-pileup limit: Apollonius subtraction, which corresponds to exactly implementing Eq. (5.1) for $\beta = 1$, and iterated Voronoi subtraction, which repeatedly applies Eq. (5.1) with an infinitesimal $\rho$. Since pileup is characteristic of a hadron collider, throughout this section we compute the EMD using particle transverse momenta $p_{T,i}$ and rapidity-azimuth coordinates $\hat{n}_i = (y_i, \phi_i)$, with $\theta_{ij}$ being the rapidity-azimuth distance. Typically, pileup is taken to be uniform in a bounded region of the plane (e.g. $|y| < y_{\max}$), though the specifics will not significantly affect our analysis. First, though, we establish an important lemma that justifies why the corrected distribution $\mathcal{E}_C$ has a particle-like interpretation.

## 5.1 A property of semi-discrete optimal transport

There is a direct connection between pileup subtraction in Eq. (5.1) and semi-discrete optimal transport [149]. Semi-discrete means that we are comparing a discrete energy flow (i.e. one composed of individual particles) to a smooth distribution (i.e. uniform pileup contamination).

Importantly, if $\mathcal{E}$ is discrete, then the corrected distribution $\mathcal{E}_C$ will also be discrete. This can be proved via the following lemma.

**Lemma 3.** *$\mathcal{E}_C$ defined according to Eq. (5.1) is strictly contained in $\mathcal{E}$, where containment here means that $\mathcal{E} - \mathcal{E}_C$ is a valid distribution with non-negative particle transverse momenta.*

*Proof.* Suppose for the sake of contradiction that $\mathcal{E}_C$ is defined according to Eq. (5.1) has some support where $\mathcal{E}$ does not. Let $\tilde{\mathcal{E}}$ be the distribution that $\mathcal{E}_C$ flows to when $\mathcal{E}_C + \rho \mathcal{U}$ is optimally transported to $\mathcal{E}$, noting that by definition, $\tilde{\mathcal{E}}$ must be contained in $\mathcal{E}$. By the linear sum structure of Eq. (3.2) [150], we have the following relation:

$$\mathrm{EMD}_\beta(\mathcal{E}, \mathcal{E}_C + \rho \mathcal{U}) = \mathrm{EMD}_\beta(\tilde{\mathcal{E}}, \mathcal{E}_C) + \mathrm{EMD}_\beta(\mathcal{E} - \tilde{\mathcal{E}}, \rho \mathcal{U}). \tag{5.2}$$

Now using the following property of $\mathrm{EMD}_\beta$ inherited from Wasserstein distances [149]:

$$\mathrm{EMD}_\beta(\mathcal{E}, \mathcal{F}) \geq \mathrm{EMD}_\beta(\mathcal{E} + \mathcal{G}, \mathcal{F} + \mathcal{G}), \tag{5.3}$$

with equality if $\beta = 1$ and the ground metric is Euclidean, we add $\tilde{\mathcal{E}}$ to both arguments of the last term in Eq. (5.2) and apply Eq. (5.3) to find:

$$\mathrm{EMD}_\beta(\mathcal{E}, \mathcal{E}_C + \rho \mathcal{U}) \geq \mathrm{EMD}_\beta(\tilde{\mathcal{E}}, \mathcal{E}_C) + \mathrm{EMD}_\beta(\mathcal{E}, \tilde{\mathcal{E}} + \rho \mathcal{U}). \tag{5.4}$$

Now using that $\mathrm{EMD}_\beta(\tilde{\mathcal{E}}, \mathcal{E}_C) > 0$ by the assumption that they have different supports as well as the non-negativity of the EMD, we find:

$$\mathrm{EMD}_\beta(\mathcal{E}, \mathcal{E}_C + \rho \mathcal{U}) > \mathrm{EMD}_\beta(\mathcal{E}, \tilde{\mathcal{E}} + \rho \mathcal{U}), \tag{5.5}$$

which contradicts the assumption that $\mathcal{E}_C$ is found according to Eq. (5.1). Thus, we conclude that $\mathcal{E}_C$ has no support outside of the support of $\mathcal{E}$, verifying the claim. $\qquad\square$

This lemma establishes that pileup mitigation strategies defined by Eq. (5.1) act by scaling the energies of the particles in the original event $\mathcal{E}$, not by producing new particles. Indeed, this is a desirable feature of many popular pileup mitigations schemes, including two well-known methods that we describe next.

## 5.2 Voronoi area subtraction

Voronoi area subtraction [55, 56, 138] is a pileup mitigation technique that estimates a particle's pileup contamination by associating it with an area determined by its corresponding *Voronoi region*, or the set of points in the plane closer to that particle than any other [151]. Letting $A_i^{\mathrm{Vor.}}$ be the area of the Voronoi region of particle $i$, Voronoi subtraction then simply removes $\rho A_i^{\mathrm{Vor.}}$ from each particle's transverse momentum, without letting the particle $p_T$ become negative. If $\rho A_i^{\mathrm{Vor.}} \geq p_{T,i}$ then the particle is removed entirely. In Fig. 9a, we show the Voronoi regions for an example jet recorded by the CMS detector [25, 152].

(a) Voronoi            (b) Constituent Subtraction            (c) Apollonius
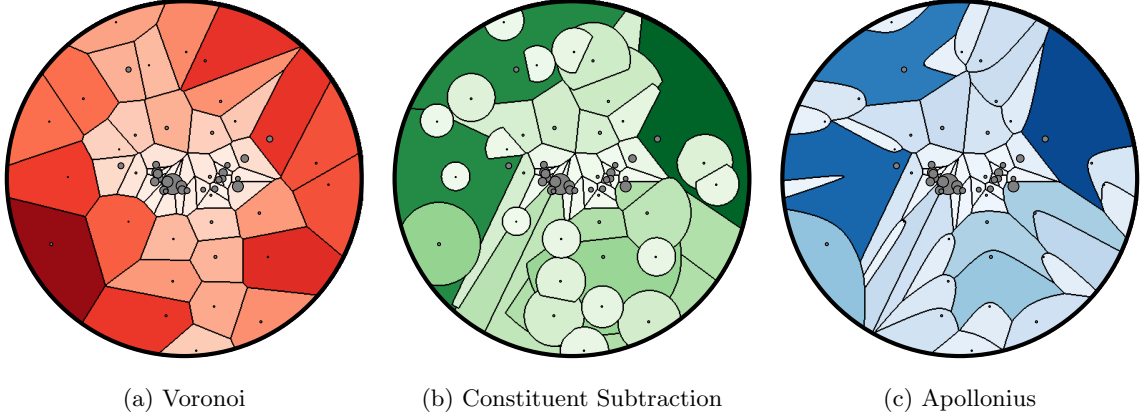
**Figure 9**:  Three pileup subtraction procedures shown for a jet from the 2011A CMS Jet Primary Dataset.  The jet constituents are shown as gray disks at their locations in the rapidity-azimuth plane with sizes proportional to their transverse momenta.  The boundary is at a distance $R = 0.5$ from the jet axis at the center.  The color intensity of each region is proportional to its area, which determines the size of the pileup correction.  The Voronoi diagram (a) is independent of $\rho$.  The constituent subtraction (b) and Apollonius (c) diagrams are determined using a $\rho$ that corresponds to subtracting one-tenth of the total scalar $p_T$ of the jet.

Voronoi area subtraction (VAS) can be thought of as carving up the uniform event $\rho \mathcal{U}$ according to the original event's Voronoi diagram and transporting this energy to the location of the corresponding particle, yielding the corrected energy flow:

$$\mathcal{E}_{\text{VAS}}(\hat{n}) = \sum_{i=1}^{M} \max\left(p_{T,i} - \rho A_i^{\text{Vor.}}, 0\right) \delta(\hat{n} - \hat{n}_i). \tag{5.6}$$

Strictly speaking, Voronoi area subtraction does not satisfy exact IRC invariance (see Eqs. (2.1) and (2.2)) and thus it cannot in general be written as operating on energy flows. The reason is that an exact IRC splitting changes the number of Voronoi regions as well as their areas. In order for Eq. (5.6) to be valid, we therefore assume that particles with exactly zero transverse momentum are removed and exactly coincident particles are combined before applying the Voronoi area subtraction procedure.

In the limit that $\rho \leq p_{T,i}/A_i^{\text{Vor.}}$ for all particles $i$, the max in Eq. (5.6) evaluates to just its first argument. In this case, since no particle is assigned a larger correction than its own transverse momentum, the Voronoi diagram gives the optimal transportation plan that minimizes the EMD of moving the uniform event with density $\rho$ onto the event of interest:

$$\boxed{\rho \leq \min\left\{p_{T,i}/A_i^{\text{Vor.}}\right\}: \quad \mathcal{E}_{\text{VAS}} = \underset{\mathcal{E}' \in \Omega}{\arg\min}\, \text{EMD}_1(\mathcal{E}, \mathcal{E}' + \rho\mathcal{U}).} \tag{5.7}$$

Thus, in this small-pileup limit, Eq. (5.6) agrees with Eq. (5.1) with $\beta = 1$. Despite this attractive geometric interpretation, Voronoi area subtraction beyond this limit is sensitive to arbitrarily soft particles: the amount that is subtracted depends only on particle positions, through their Voronoi areas, and not their transverse momenta.

## 5.3 Constituent subtraction

Constituent subtraction [58] is another pileup mitigation method that resolves several pathologies of Voronoi area subtraction by correcting the particles in a manner that depends on both their positions and their transverse momenta.[11] This comes at the cost of requiring a fine grid of low energy "ghost" particles with $p_T^g = \rho A^{\text{ghost}}$, where $A^{\text{ghost}}$ is the area assigned to each ghost, as a proxy for the pileup contamination. The algorithm is applied by considering the geometrically closest ghost-particle pair $k, i$ and modifying them via:

$$p_{T,i} \to \max(p_{T,i} - p_{T,k}^g, 0), \qquad p_{T,k}^g \to \max(p_{T,k}^g - p_{T,i}, 0), \tag{5.8}$$

continuing until all such pairs have been considered. Since the number of ghosts is typically large in order to have fine angular granularity, this iteration through all ghost-particle pairs can be computationally expensive.

Constituent subtraction (CS) in the continuum ghost limit can be geometrically described by placing circles around each point in the rapidity-azimuth plane and simultaneously increasing their radii. Each point in the plane is assigned to the particle whose circle reaches it first. Circles stop growing when $A_i^{\text{CS}}$, the area assigned to particle $i$, grows larger than $p_{T,i}/\rho$. We can write the resulting distribution as:

$$\mathcal{E}_{\text{CS}}(\hat{n}) = \sum_{i=1}^{M} \left(p_{T,i} - \rho A_i^{\text{CS}}\right) \delta(\hat{n} - \hat{n}_i). \tag{5.9}$$

Unlike naive Voronoi area subtraction, continuum constituent subtraction satisfies exact IRC invariance, since a zero energy particle has zero $A^{\text{CS}}$ and an exact collinear splitting yields two areas that sum to the original $A^{\text{CS}}$. Constituent subtraction is also better suited for intermediate values of $\rho$, where particles can be fully removed, since further corrections are distributed to the next closest particle instead of being ignored as in Voronoi area subtraction.

Due to the complicated shapes of the corresponding regions, it is difficult to describe the areas $A_i^{\text{CS}}$ analytically and in practice they need to be estimated using numerical ghosts. An example of constituent subtraction is shown in Fig. 9b, where it can be seen that some region boundaries are straight and thus contained in the Voronoi diagram of Fig. 9a. Indeed, growing circles from a set of points and assigning points in the plane according to which circle reaches them first is another way of describing the construction of a Voronoi diagram. Regions with circular boundaries correspond to softer particles that are fully subtracted by the constituent subtraction procedure.

---

[11]In this discussion, we focus on the $\alpha = 0$ case of constituent subtraction, as recommended by Ref. [58].

When $\rho$ is sufficiently small such that no particle's region has a circular boundary (i.e. no circle stops growing), constituent subtraction is exactly equivalent to Voronoi area subtraction. Constituent subtraction in the low-pileup limit is then also equivalent to optimally transporting the uniform event with density $\rho$ to the event of interest and subtracting accordingly, again in line with Eq. (5.1) with $\beta = 1$:

$$\rho \leq \min \left\{ p_{T,i} / A_i^{\text{Vor.}} \right\} : \quad \mathcal{E}_{\text{CS}} = \arg \min_{\mathcal{E}' \in \Omega} \text{EMD}_1(\mathcal{E}, \mathcal{E}' + \rho \mathcal{U}). \tag{5.10}$$

Constituent subtraction can also be extended with a $\Delta R^{\text{max}}$ parameter to restrict ghosts from affecting distant particles. Our geometric formalism can also encompass this locality by re-introducing the $R$-parameter to the EMD in Eq. (5.10) with $R = \Delta R^{\text{max}}$.

## 5.4 Apollonius subtraction

Voronoi area subtraction and constituent subtraction both make contact with Eq. (5.1) in the small-$\rho$ limit, but we would like to explore pileup subtraction based on optimal transport for all values of $\rho$. By Lemma 3, we know that the corrected event is contained in the original event, and by the decomposition properties of the EMD in Eq. (5.2), we only need to consider the transport of $\rho \mathcal{U}$ to $\mathcal{E}$. Since the total transverse momenta of $\rho \mathcal{U}$ and $\mathcal{E}$ are generally different, this is now an example of a semi-discrete, *unbalanced* optimal transport problem [153, 154].

The problem of minimizing the EMD between a uniform distribution and an event is solved, for general $\beta$, by a *generalized Laguerre diagram* [154]. For the special case of $\beta = 1$, which we focus on here, this is also known as the *Apollonius diagram* (or additively weighted Voronoi diagram) [149, 155, 156], and for $\beta = 2$ it is a *power diagram* [157]. An Apollonius diagram in the plane is constructed from a set of points $\hat{n}_i$ that each carry a non-negative weight $w_i$ that is the $i^{\text{th}}$ component of a vector $\mathbf{w} \in \mathbb{R}_+^M$. In the two-dimensional Euclidean plane, the Apollonius region associated to particle $i$ depending on $\mathbf{w}$ is:

$$R_i^{\text{Apoll.}}(\mathbf{w}) = \left\{ \hat{n} \in \mathbb{R}^2 \,\middle|\, \|\hat{n} - \hat{n}_i\| - w_i \leq \|\hat{n} - \hat{n}_j\| - w_j, \ \forall \, j \neq i \right\}, \tag{5.11}$$

where particle indices $i, j = 1, \ldots, M$ and $\| \cdot \|$ is the Euclidean norm. One interpretation of Eq. (5.11) is that region $i$ is all points closer to a circle of radius $w_i$ centered at $\hat{n}_i$ than to the corresponding circle for any other particle. The boundaries of the Apollonius regions are contained in the set $\{\hat{n} \in \mathbb{R}^2 \,|\, \|\hat{n} - \hat{n}_i\| - \|\hat{n} - \hat{n}_j\| = w_i - w_j\}$, which is a union of hyperbolic segments. Note that adding the same constant to all of the weights does not change the resulting Apollonius diagram. Hence, if all the weights are equal, they can equivalently be set to zero and we attain the Voronoi diagram as a limiting case of an Apollonius diagram.

We can now specify the action of Apollonius subtraction on an event using the areas of

the Apollonius regions subject to the minimal EMD requirement:

$$\mathcal{E}_{\text{Apoll.}}(\hat{n}) = \sum_{i=1}^{M} \left( p_{T,i} - \rho A_i^{\text{Apoll.}}(\mathbf{w}^*) \right) \delta(\hat{n} - \hat{n}_i), \tag{5.12}$$

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathbb{R}_+^M}{\arg\min} \sum_{i=1}^{M} \text{EMD}_{1,R} \left( \{p_{T,i}, \hat{n}_i\}, \rho R_i^{\text{Apoll.}}(\mathbf{w}) \right), \tag{5.13}$$

treating $\rho R_i^{\text{Apoll.}}(\mathbf{w})$ as an event with uniform energy density $\rho$ in that Apollonius region. Here, Eq. (5.12) is analogous to Eqs. (5.6) and (5.9), and Eq. (5.13) implements the requirement that the EMD of the subtraction is minimal. Note that the $R$ parameter in Eq. (5.13) serves only to guarantee that it is more efficient to transport energy rather than create/destroy it. As long as $2R$ is greater than the diameter of the space, $R$ has no impact on the solution other than to guarantee that $\rho A_i^{\text{Apoll.}}(\mathbf{w}^*)$ does not exceed $p_{T,i}$, as this would be less efficient than transporting the excess energy elsewhere. An example of an Apollonius diagram is shown in Fig. 9c, where hyperbolic boundaries of the Apollonius regions are clearly seen in the outer part of the jet and straight boundaries, matching those of the Voronoi diagram, are seen near the core.

In this way, Apollonius subtraction generalizes Voronoi area and constituent subtraction beyond the small-pileup limit, directly implementing Eq. (5.1) for $\beta = 1$ for all values of $\rho$:

$$\boxed{\mathcal{E}_{\text{Apoll.}} = \underset{\mathcal{E}' \in \Omega}{\arg\min} \, \text{EMD}_1(\mathcal{E}, \mathcal{E}' + \rho \mathcal{U}).} \tag{5.14}$$

While the optimal solution in Eq. (5.13) is based on an unbalanced optimal transport problem, the restatement in Eq. (5.14) corresponds to balanced transport. This same connection underpins Lemma 3, guaranteeing that the corrected event in Eq. (5.12) involves the same $M$ directions as the original event, just with different weights.

To turn Eq. (5.14) into a practical algorithm, we would need an efficient way to compute the weights according to Eq. (5.13). While Refs. [153, 154] have developed the theoretical framework of semi-discrete, unbalanced optimal transport needed to solve this convex minimization problem, they stop short of describing easily-implementable algorithms to attain practical solutions. In order to create Fig. 9c, we were limited to using numerical ghosts to directly solve for the transport plan that minimizes the EMD cost of subtracting the uniform energy component from the event, which is too computationally costly for LHC applications.

If the target areas $A_i^{\text{Apoll.}}$ are previously specified, then the solution to Eq. (5.13) simplifies [149]. Given that the areas depend nontrivially on the resulting weight vector, though, the only case where we know them ahead of time is when $\rho$ is such that *all* of the energy will be exactly subtracted, in which case $A_i^{\text{Apoll.}} = p_{T,i}/\rho$. Though this is not so useful for pileup, where we typically want to subtract an amount of energy less than the total, it does indicate that an Apollonius diagram can be found and used to compute the event isotropy from Sec. 3.1.5 without the use of numerical ghosts. We leave the implementation of such a procedure to future work, though we note that Ref. [149] has already built an implementation

that relies on numerical ghosts to estimate the areas of the Apollonius regions rather than solving for them analytically.

## 5.5 Iterated Voronoi subtraction

Given the difficulty of analytically solving Eq. (5.13) and thus implementing Apollonius subtraction, we now develop an alternative method called iterated Voronoi subtraction that gives up a global notion of minimizing EMD but retains a local one. In all three methods described above, the difficulty comes when a particle is removed in the course of subtracting pileup. Otherwise, the above methods all reduce to subtracting transverse momentum according to the Voronoi areas of the regions corresponding to the particles, as in Eq. (5.7). This suggests a procedure in which pileup is subtracted according to Eq. (5.1) an infinitesimal amount at a time, thus ensuring that Eq. (5.7) can be used at every stage of the procedure.

The area of the Voronoi cell of particle $i$ is now a function of the total amount of energy density that has been subtracted thus far, a quantity that starts at zero and will be integrated up to the target $\rho^{\text{tot}}$ over the course of the procedure. When a particle loses all of its transverse momentum, it is removed from the Voronoi diagram and is considered to have zero area associated to it. The removal of a particle from the diagram changes the Voronoi regions of all of its neighbors, and their areas are updated accordingly. Denoting the area associated to particle $i$ after $\rho$ worth of energy density has been subtracted as $A_i^{\text{IVS}}(\rho)$, we can write the corrected distribution for iterated Voronoi subtraction (IVS) as:

$$\mathcal{E}_{\text{IVS}}(\hat{n}) = \sum_{i=1}^{M} \left( p_{T,i} - \int_0^{\rho^{\text{tot}}} d\rho \, A_i^{\text{IVS}}(\rho) \right) \delta(\hat{n} - \hat{n}_i). \tag{5.15}$$

Unlike Eqs. (5.12) and (5.13), Eq. (5.15) naturally lends itself to a simple and efficient implementation. We can iteratively solve for $A_i^{\text{IVS}}(\rho)$ using the fact that the areas correspond to Voronoi regions, and furthermore that these regions change only when a particle is removed. Let $\mathcal{E}^{(0)}$ be the initial event consisting of particles with transverse momenta $p_{T,i}^{(0)}$ in Voronoi regions with area $A_i^{(0)}$. We subtract a total energy density $\rho^{\text{tot}}$ by breaking up the integral in Eq. (5.15) starting with $\rho^{(0)} = 0$ and determining the boundaries from:

$$\rho^{(n)} = \max \left\{ \rho \,\middle|\, 0 \le \rho \le \rho^{\text{tot}} - \sum_{i=0}^{n-1} \rho^{(i)} \text{ s.t. } \rho A_j^{(n-1)} \le p_{T,j}^{(n-1)} \, \forall \, j \right\}, \tag{5.16}$$

where $n$ starts at 1 and goes up to at most $M$. The values of $\rho^{(n)}$ can be expressed simply as:

$$\rho^{(n)} = \min \left( \min \left\{ \frac{p_{T,j}^{(n-1)}}{A_j^{(n-1)}} \right\}, \rho^{\text{tot}} - \sum_{i=0}^{n-1} \rho^{(i)} \right), \tag{5.17}$$

where the inner minimum is taken over all remaining particles with $p_{T,j}^{(n-1)} > 0$.

The updated particle momenta from each piece of the integral in Eq. (5.15) are then:

$$p_{T,i}^{(n)} = p_{T,i}^{(n-1)} - \rho^{(n)} A_i^{(n-1)}, \tag{5.18}$$

and a particle is considered removed if its transverse momentum is zero, in which case it is also considered to have zero area. The areas $A_i^{(n)}$ are determined by the Voronoi diagram of $\mathcal{E}^{(n)}$. The above procedure terminates either when the total amount of energy density removed is equal to $\rho^{\text{tot}}$ or there are no more particles left in the event. Thus, iterated Voronoi subtraction makes contact with the geometric perspective of Eq. (5.1), applying it in infinitesimal increments, resulting in the discrete steps:

$$\boxed{\mathcal{E}_{\text{IVS}}^{(n)} = \underset{\mathcal{E}' \in \Omega}{\arg\min}\, \text{EMD}_1(\mathcal{E}^{(n-1)}, \mathcal{E}' + \rho^{(n)}\mathcal{U}).} \tag{5.19}$$

Said another way, this is simply a repeated application of Voronoi area subtraction: subtract until a particle reaches zero momentum, and repeat until the desired energy density has been removed.

Iterated Voronoi subtraction is made even more attractive computationally when one considers that the Voronoi diagram of $\mathcal{E}^{(n)}$ does not need to be recomputed from scratch. Rather, it can be obtained from the Voronoi diagram of $\mathcal{E}^{(n-1)}$ by removing a site and updating only the neighboring regions. Thus, we only need to construct the Voronoi diagram of $\mathcal{E}^{(0)}$ and each removal can be done in constant (amortized) time as the average number of neighbors of any cell is no more than 6 [151]. We have constructed an implementation of iterated Voronoi subtraction that interfaces with FastJet and will explore its phenomenological properties in future work.

# 6    Theory space

When do two theories give rise to similar signatures? In this section, we seek to generalize the intuition behind the EMD to obtain a metric between theories using their predicted cross sections in energy flow space. A construction of such a distance and the induced "theory space" is conceptually useful and, in fact, naturally underpins several recently introduced techniques for collider physics.

We introduce the cross section mover's distance ($\Sigma$MD) as a metric for the space of theories. Here, we treat a "theory" as an ensemble of event energy flows with corresponding cross sections, encompassing both the predictions of quantum field theories as well as the structure of collider datasets. To accomplish this, we again make use of an EMD-like construction, except the $\Sigma$MD uses the EMD itself as the "angles" and the event cross sections as the "energies", as mentioned in Table 2. The resulting space of theories with the $\Sigma$MD as a metric is illustrated in Fig. 10. Interestingly, Ref. [158] also put a metric on theory space by using the Fisher information matrix.

## 6.1    Introducing a distance between theories

A "theory" $\mathcal{T}$ is taken to be a (finite, for now) set of events with associated cross sections $\{(\mathcal{E}_i, \sigma_i)\}_{i=1}^N$. We can equivalently view $\mathcal{T}$ as a distribution over the space of event configura-
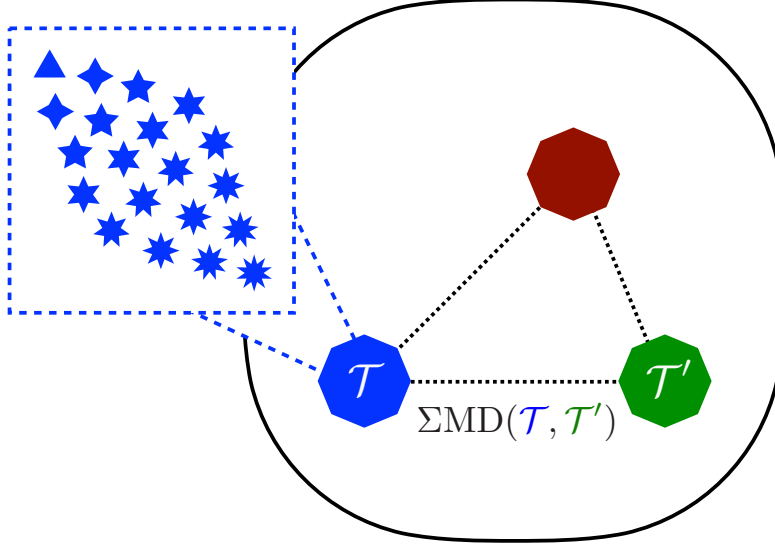
**Figure 10**: An illustration of the space of theories. Each point in the space is a "theory": a distribution over (or collection of) events, as indicated by the blue point. The distance between theories is quantified by the $\Sigma$MD, giving rise to a metric space.

tions $\mathcal{E}$:

$$\mathcal{T}(\mathcal{E}) = \sum_{i=1}^{N} \sigma_i \, \delta(\mathcal{E} - \mathcal{E}_i). \tag{6.1}$$

In the case of unweighted events, the cross sections are simply $\sigma_i = 1/\mathcal{L}$, where $\mathcal{L}$ is the total integrated luminosity. One can think of these $N$ events as being produced by a perfect event generator or measured by a perfect collider experiment in the context of this theory. In the $\mathcal{L} \to \infty$ or $N \to \infty$ limit, Eq. (6.1) becomes a smooth distribution.

The $\Sigma$MD is the minimum "work" required to rearrange one theory $\mathcal{T}$ into another $\mathcal{T}'$ by moving cross section $\mathcal{F}_{ij}$ from event $i$ in one theory to event $j$ in the other:

$$\Sigma\mathrm{MD}_{\gamma,S;\beta,R}(\mathcal{T}, \mathcal{T}') = \min_{\{\mathcal{F}_{ij} \geq 0\}} \sum_{i=1}^{N} \sum_{j=1}^{N'} \mathcal{F}_{ij} \left( \frac{\mathrm{EMD}_{\beta,R}(\mathcal{E}_i, \mathcal{E}'_j)}{S} \right)^{\gamma} + \left| \sum_{i=1}^{N} \sigma_i - \sum_{j=1}^{N'} \sigma'_j \right|, \tag{6.2}$$

$$\sum_{i=1}^{N} \mathcal{F}_{ij} \leq \sigma'_j, \quad \sum_{j=1}^{N'} \mathcal{F}_{ij} \leq \sigma_i, \quad \sum_{i=1}^{N} \sum_{j=1}^{N'} \mathcal{F}_{ij} = \min \left( \sum_{i=1}^{N} \sigma_i, \sum_{j=1}^{N'} \sigma'_j \right). \tag{6.3}$$

Here, $i$ and $j$ index the events in theories $\mathcal{T}$ and $\mathcal{T}'$, respectively. The parameter $S$, which has the same units as the EMD, controls the relative importance of the two terms, analogous to the jet radius parameter $R$ in the EMD. We have also introduced a possible $\gamma$ exponent, analogous to the $\beta$ angular exponent in the EMD.

The $\Sigma$MD has dimensions of cross section, where the first term quantifies the difference in event distributions and the second term accounts for the creation or destruction of cross

section. For $\gamma > 0$, it is a true metric as long as the underlying EMD is a metric and $2S$ is larger than the largest attainable EMD between two events.[12] In the limit as $S \to \infty$, the $\Sigma$MD reduces simply to the difference in total cross section between the two theories. The natural continuum notion of Eq. (6.2) can be used whenever such an analysis is analytically tractable.

The $\Sigma$MD from a theory to itself is zero in the continuum limit or with infinite data. Further, two theories that differ in their Lagrangians yet give rise to identical scattering cross sections for all energy flows will have a $\Sigma$MD of zero. This includes, for instance, theories that are equivalent up to field redefinitions [159] or rearrangements of the asymptotic states [160–162]. Finally, if two theories have any observable differences in energy flow, then the $\Sigma$MD between them will be non-zero. Note that the $\Sigma$MD inherits the flavor and charge insensitivity of the EMD, but it is interesting to consider extending the $\Sigma$MD to account for additional quantum numbers that particles may carry.

## 6.2 Jet quenching via quantile matching

Quantile matching [64] is an analysis strategy to study the modification of jets as they traverse the quark-gluon plasma in heavy-ion collisions. We now show that, surprisingly, this technique can be cast naturally in a $\Sigma$MD formulation. The optimal "theory moving" transport between two otherwise-equivalent datasets provides a proxy for the jet modification by the quark-gluon plasma.

Intuitively, the idea is to select a set of statistically equivalent jets from both proton-proton (pp) collisions and heavy-ion (AA) collisions. This gives a snapshot of jets and their energies before and after modification by the quark-gluon plasma, respectively. Such a selection can be achieved by selecting jets with the same upper cumulative effective cross section, after appropriately normalizing the AA cross section to account for the average number of nucleon-nucleon collisions.

With such a selection, a quantile matching can be used to specify $p_T^{\text{quant}}$ for a given heavy-ion jet with reconstructed transverse momentum $p_T^{\text{AA}}$:

$$\Sigma_{\text{pp}}^{\text{eff}}(p_T^{\text{quant}}) \equiv \Sigma_{\text{AA}}^{\text{eff}}(p_T^{\text{AA}}), \tag{6.4}$$

where $p_T^{\text{quant}}$ gives a proxy for the jet $p_T$ prior to modification by the quark-gluon plasma. The ratio between the heavy-ion and proton-proton jet transverse momentum in the same quantile then gives a physically-motivated quantification of the medium jet modification.

$$Q_{\text{AA}}(p_T^{\text{quant}}) = \frac{p_T^{\text{AA}}}{p_T^{\text{quant}}}. \tag{6.5}$$

We now turn to explaining the intriguing connection between this quantile matching procedure and the $\Sigma$MD through optimal transport. We use transverse momenta in place of energies and take $R \to \infty$ in Eq. (1.2), where the EMD becomes simply the difference in

---

[12]The analogous discussion to footnote 1 holds for $\gamma > 1$.

jet transverse momenta. Further, we set $\gamma = 1$ and $S = 1$ in Eq. (6.2) and note that the normalization of the cross sections makes the second term in that equation vanish.

The theory moving problem now becomes a simple one-dimensional optimal transport problem of moving the pp jet $p_T$ distribution to the AA jet $p_T$ distribution. Remarkably, this is mathematically equivalent to quantile matching. We use the notation TM to represent the optimal theory movement $\mathcal{F}^*$ in the $\Sigma$MD. Letting $\mathcal{T}_{\mathrm{AA}}$ be the set of heavy-ion jets and $\mathcal{T}_{\mathrm{pp}}$ be the set of proton-proton jets, we have that:

$$p_T^{\mathrm{quant}} = \mathrm{TM}\left(\mathcal{T}_{\mathrm{AA}}, \mathcal{T}_{\mathrm{pp}}\right)[p_T^{\mathrm{AA}}], \tag{6.6}$$

where we can define this formally using a "ghost" heavy-ion jet with transverse momentum $p_T^{\mathrm{AA}}$ and infinitesimal cross section $\sigma \sim 0$.

Quantile matching can therefore be seen as a matching induced by the optimal theory movement between the heavy-ion and proton-proton jets. In this sense, it operationally defines the modification by the quark-gluon plasma in terms of the theory-movement of the jet transverse momentum spectrum. It would be interesting to follow this connection further and explore this procedure using the full EMD beyond the $R \to \infty$ limit to study the medium modification as a function of the jet substructure.

## 6.3 Event clustering and coresets

One of the essential unsupervised methods for probing a dataset is to analyze its complexity. A method to do this for collider physics datasets is that of $k$-eventiness, recently introduced in Ref. [25]. Here, one seeks to find $k$ representative events that minimize the EMD from each event in the dataset to the nearest representative event:

$$\mathcal{V}_k^{(\gamma)} = \min_{\mathcal{K}_1,\ldots,\mathcal{K}_k} \sum_{i=1}^{N} \sigma_i \min\{\mathrm{EMD}(\mathcal{E}_i, \mathcal{K}_1)^\gamma, \mathrm{EMD}(\mathcal{E}_i, \mathcal{K}_2)^\gamma, \ldots, \mathrm{EMD}(\mathcal{E}_i, \mathcal{K}_k)^\gamma\}, \tag{6.7}$$

where we have dropped the $\beta$ and $R$ subscripts on EMD for compactness. The value of $\mathcal{V}_k$ probes how well the dataset is approximated by the $k$ events. This gives rise to the notion of $\mathcal{V}_k$ as the "$k$-eventiness" of the dataset, in analogy with $N$-(sub)jettiness, where smaller values of $\mathcal{V}_k$ indicate better approximations.

From a geometric perspective, $\mathcal{V}_k$ is the smallest $\Sigma$MD to the manifold of $k$-event datasets. Analogous to Eq. (3.2), we can introduce the $S \to \infty$ version of the $\Sigma$MD:

$$\Sigma\mathrm{MD}_\gamma(\mathcal{T}, \mathcal{T}') = \lim_{S \to \infty} S^\gamma \, \Sigma\mathrm{MD}_{\gamma,S}(\mathcal{T}, \mathcal{T}'), \tag{6.8}$$

which yields an infinite distances between theories of different total cross sections. Following the identical logic to Secs. 3.1.4 and 3.2.2, we have:

$$\mathcal{V}_k^{(\gamma)} = \min_{|\mathcal{T}'|=k} \Sigma\mathrm{MD}_\gamma(\mathcal{T}, \mathcal{T}'). \tag{6.9}$$

Here, we use the $|\cdot|$ notation to count the number of events in $\mathcal{T}'$. Just like for $N$-(sub)jettiness, different values of $\gamma$ highlight different aspects of theory space geometry.

Following the logic in Sec. 4.1 of lifting the $N$-jettiness observable into the XCone jet algorithm, we can lift $k$-eventiness into an event clustering algorithm. The representative events $\mathcal{K}$ (i.e. the point of closest approach on the $k$-event manifold), has the interpretation of the "$k$-geometric-medians" for $\gamma = 1$ or "$k$-means" for $\gamma = 2$. For practical applications, it is often convenient to restrict the representative events to be within the dataset $\mathcal{T}$, i.e. the "$k$-medoids", giving only an approximate value of $\mathcal{V}_k$. While the full problem of finding the representative jets may be computationally intractable, fast approximations to find the medoids exist and have been explored in Ref. [25].

Inspired by Sec. 4.2, one might consider implementing sequential clustering algorithms by iterating $\Sigma$MD computations to approximate $M$ events with $M - 1$ events and so forth. Such a clustering may be helpful for rigorous data compression of large collider datasets or, if implemented efficiently, for tasks such as triggering. These ideas are closely related to the notion of finding a coreset (see Ref. [163] for a recent review), for which techniques from quantum information and quantum computation may also find use [164]. Additionally, Ref. [165] uses the Wasserstein metric to construct "measure coresets" that take into account the underlying data distribution and which may prove useful for high-energy physics applications. We leave further exploration of theory geometry and theory space algorithms to future work.

## 7   Conclusions

In this paper, we have explored the metric space of collider events from a theoretical perspective. Beginning from the EMD between final states, namely the "work" required to rearrange one into another, we have cast a multitude of diverse collider algorithms and analysis techniques in a geometric language. First, we connected this metric to the fundamental notion of IRC safety in massless quantum field theories, with the EMD providing a sharp language to define IRC safety and even Sudakov safety. We extended this connection by highlighting that a wide variety of collider observables, including thrust and $N$-jettiness, can be cast as distances between events and manifolds in this space. Further, we demonstrated that many jet clustering algorithms, such as exclusive cone finding and sequential clustering, can be exactly derived in full detail from the simple principle that jets are the best $N$-particle approximation to the event. Even pileup mitigation techniques developed to face the LHC-era challenge of high luminosity running can be cast in the language of subtracting a uniform radiation pattern, connecting this field to semidiscrete unbalanced optimal transport. Finally, we generalized our reasoning to define a distance between "theories" as sets of events with cross sections, proving a new lens to understand several existing techniques and a roadmap for future developments in the geometry of theory space.

From the perspective of massless quantum field theories, our metric space of events is the natural space for understanding observables, as the only truly observable quantities are IRC safe. More speculatively, it would be interesting to circumvent the (unphysical) particle-

level stage of calculations and make theoretical predictions directly in the space of events. Understanding and expanding in this direction would require natural notions of volume and integration in this space, perhaps aided by recent developments in Wasserstein spaces [166–169], though we leave this fascinating exploration to future work. Nonetheless, it has already been established that the energy flows themselves obey factorization theorems in effective field theory contexts [170], and give rise to rich behavior in correlators [5, 171–173]. Going directly from first principles and symmetries to observables (i.e. energy flows, not particles) suggests a natural extension to the philosophy driving the present scattering amplitudes program (see Refs. [159, 174, 175] for reviews). It is also interesting to extend this logic to massive quantum field theories where observable quantities can depend on flavor and charge. One promising strategy is to treat events as collections of objects (jets, electrons, muons, etc.) and to use a ground distance that penalizes converting one type of object into another [176]. Alternatively, it may be possible to find an EMD variant that mimics the flavor-sensitive clustering behavior of Ref. [177].

It is also useful to discuss these developments in the broader context of machine learning and the physical sciences [122, 178, 179]. Typically, problems in the natural sciences can be cast as machine learning problems such as classification and regression, whereby the relevant tools from machine learning can be applied to achieve improved performance on those tasks. It is far rarer for machine learning to enhance our theoretical or conceptual understanding of physics directly. This story provides an interesting case where new insights and questions exposed by machine learning have impacted purely theoretical and phenomenological collider physics. The question of when two collider events are similar, for which the EMD was introduced, was originally motivated by unsupervised learning methods and autoencoders [180–183], which require a distance matrix or reconstruction loss. By providing an answer to this simple question, which itself involved familiar machine learning tools such as optimal transport, we uncovered a new mathematical formalism to better understand and express concepts in quantum field theory and collider physics. We hope that this will be just one example of many future profound insights into the natural sciences facilitated by this perspective.

### Acknowledgments

## A  Energy moving with massive particles

In this appendix, we briefly explore an alternative definition of energy flow appropriate for massive particles, with a corresponding change in the measures used to define the EMD. The energy flow in Eq. (1.1) treats events as sets of particles that have energy-like weights $\{E_i\}$ and geometric directions $\{\hat{n}_i\}$. The EMD in Eq. (1.2) is based on pairwise distances $\{\theta_{ij}\}$ that are only functions of the $\hat{n}_i$ and $\hat{n}_j$ directions. The exact definitions of $E_i$, $\hat{n}_i$, and $\theta_{ij}$ may vary depending on the collider context and other choices. For massless final-state particles in $e^+e^-$ collisions, it is typical to take the energy $E$ to be equal to the total momentum $|\vec{p}|$, and the geometric direction $\hat{n}$ to be equal to the unit vector $\vec{p}/E$. For massless particles in hadronic collisions, it is natural to use transverse momentum $p_T$ and a geometric direction based on azimuth $\phi$ and pseudorapidity $\eta$.

It is straightforward to adapt the energy flow to massive particles (see related discussion in Ref. [13]). For the energy measure, the natural choices are energy in the $e^+e^-$ case and transverse energy in the hadronic case:

$$E_i = \sqrt{|\vec{p}_i|^2 + m_i^2}, \qquad E_{Ti} = \sqrt{p_{Ti}^2 + m_i^2}. \tag{A.1}$$

Both of these reduce nicely to the expected expressions in the $m_i \to 0$ limit. For the geometric direction, the natural choices are velocity and transverse velocity, written in four-vector notation:

$$n_i^\mu = \frac{p_i^\mu}{E_i} = (1, \vec{v})^\mu, \qquad n_{Ti}^\mu = \frac{p_i^\mu}{E_{Ti}} = (\cosh y_i, \vec{v}_{Ti}, \sinh y_i)^\mu, \tag{A.2}$$

where $\vec{v} = \vec{p}_i/E_i$ is the particle three-velocity, $\vec{v}_{Ti} = \vec{p}_{Ti}/E_{Ti}$ is the particle transverse two-velocity, and $y_i$ is the particle rapidity. Again, these have the expected behavior in the $m_i \to 0$ limit, and for finite mass, the velocities are bounded as $|\vec{v}| \in [0,1]$ and $|\vec{v}_T| \in [0,1]$.

To define the EMD, we choose the following pairwise angular distance:

$$\theta_{ij} = \sqrt{-(n_i^\mu - n_j^\mu)^2}, \tag{A.3}$$

where one replaces $n^\mu$ with $n_T^\mu$ in the hadronic case. The first minus sign is needed because the difference between two time-like vectors with $n^2 \in [0,1]$ is space-like. This expression reduces to the usual expression $\theta_{ij} = \sqrt{2n_i^\mu n_{j\mu}}$ in the massless limit.

To gain intuition for this geometric distance between massive particles, it is instructive to expand out Eq. (A.3) in the $e^+e^-$ case:

$$\theta_{ij} = \sqrt{(\vec{v}_i - \vec{v}_j)^2} = \sqrt{v_i^2 + v_j^2 - 2\,v_i\,v_j \cos \Omega_{ij}}, \tag{A.4}$$

where $\Omega_{ij} = \arccos \hat{n}_i \cdot \hat{n}_j$ is the purely geometric angle between particles $i$ and $j$. We see that the velocity magnitude $v = |\vec{v}|$ acts as a radial coordinate on the sphere, and the pairwise distances $\theta_{ij}$ are just the Euclidean distances between two points in the unit ball, with distances $v_i$ and $v_j$ from the origin and angle $\Omega_{ij}$ between them. Massless particles live
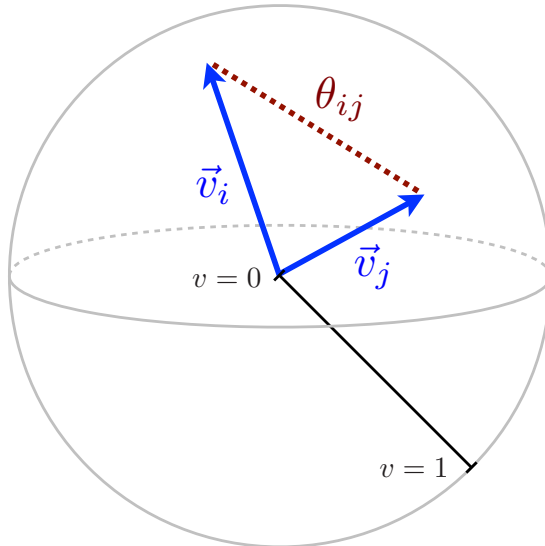
**Figure 11**: The space of (massive) particle kinematics, with pairwise distances corresponding to Euclidean distances in this space. Massless particles with $v = 1$ live on the boundary and particles at rest with $v = 0$ are at the origin. One can interpret this figure as a snapshot of the event taken at a time $t$ after the collision, when the particles have traveled a distance $vt$.

entirely on the boundary with $v = 1$ and massive particles live inside the ball with $0 \leq v < 1$. An illustration of this massive particle phase space is shown in Fig. 11.

The use of this massive distance measure has an interesting interplay with some of the studies in the body of the paper. For example, the analysis of thrust in Sec. 3.1.1 involved finding the EMD to the manifold of back-to-back massless particle configurations of potentially unequal energy. Using the massive particle distance, one could consider finding the EMD to the manifold of all possible two-particle configurations, including massive particles. For $\beta = 2$, this is equivalent to partitioning the event into two halves with masses $M_A$ and $M_B$ and corresponding energies $E_A$ and $E_B$, and minimizing the quantity $M_A^2/E_A + M_B^2/E_B$. A nice feature of this approach is that the optimal two particle configuration has the same energies and velocities as one would get from clustering the particles in each half. Note that this approach is closely related to (but not identical to) the original definition of heavy jet mass in Ref. [67] based on minimizing $M_1^2 + M_2^2$.

The idea of optimizing jet regions based on $M^2/E$ also appears in the jet maximization approach [132]. In fact, using the massive distance measure in Eq. (4.2) with $\beta = 2$ and $N = 1$, and repeating the logic in Ref. [128], we recover precisely the algorithm in Ref. [132], where the parameter $R$ controls the size of the resulting jet region. The EMD approach yields a natural way to extend the jet maximization algorithm to $N > 1$, and also allows for an alternative definition of $N$-jettiness from Eq. (3.16) based on time-like axes.

As another example, the analysis of recombination schemes in Sec. 4.2 involved minimiz-

ing the transportation cost to merge two particles into one. For $\beta = 2$, this was equivalent to the $E$-scheme, namely $\kappa = 1$ in Eq. (4.4), up to the subtlety noted in footnote 9. Using the massive particle distance, the merged particle in the $E$-scheme has the energy and direction:

$$E_c = E_i + E_j, \qquad n_c^\mu = \frac{E_i n_i^\mu + E_j n_j^\mu}{E_i + E_j}, \tag{A.5}$$

where appropriate $T$ subscripts should be included in the hadronic case. The combined four-vector is

$$p_c^\mu = E_c n_c^\mu = p_i^\mu + p_j^\mu, \tag{A.6}$$

which is a valid expression in both the $e^+e^-$ and hadronic cases. Thus, the combined four-vector is just the sum of the two particles, which is indeed the desired $E$-scheme behavior. Note, however, that the interpretation of the jet radius is very different if one uses the massive particle distance, since clustering happens in velocity space. We leave further studies of the massive particle distance to future work.

# References

[1] P. T. Komiske, E. M. Metodiev and J. Thaler, *Metric Space of Collider Events*, *Phys. Rev. Lett.* **123** (2019) 041801, [1902.02346].

[2] F. V. Tkachov, *Measuring multijet structure of hadronic energy flow or What is a jet?*, *Int. J. Mod. Phys.* **A12** (1997) 5411–5529, [hep-ph/9601308].

[3] N. A. Sveshnikov and F. V. Tkachov, *Jets and quantum field theory*, *Phys. Lett.* **B382** (1996) 403–408, [hep-ph/9512370].

[4] G. P. Korchemsky, G. Oderda and G. F. Sterman, *Power corrections and nonlocal operators*, *AIP Conf. Proc.* **407** (1997) 988, [hep-ph/9708346].

[5] C. L. Basham, L. S. Brown, S. D. Ellis and S. T. Love, *Energy Correlations in electron-Positron Annihilation in Quantum Chromodynamics: Asymptotically Free Perturbation Theory*, *Phys. Rev.* **D19** (1979) 2018.

[6] P. S. Cherzor and N. A. Sveshnikov, *Jet observables and energy momentum tensor*, in *Quantum field theory and high-energy physics. Proceedings, Workshop, QFTHEP'97, Samara, Russia, September 4-10, 1997*, pp. 402–407, 1997, hep-ph/9710349, http://theory.sinp.msu.su/ qfthep/97/pro97/cherzor.ps.gz.

[7] F. V. Tkachov, *A Theory of jet definition*, *Int. J. Mod. Phys.* **A17** (2002) 2783–2884, [hep-ph/9901444].

[8] G. P. Korchemsky and G. F. Sterman, *Power corrections to event shapes and factorization*, *Nucl. Phys.* **B555** (1999) 335–351, [hep-ph/9902341].

[9] A. V. Belitsky, G. P. Korchemsky and G. F. Sterman, *Energy flow in QCD and event shape functions*, *Phys. Lett.* **B515** (2001) 297–307, [hep-ph/0106308].

[10] C. F. Berger et al., *Snowmass 2001: Jet energy flow project*, *eConf* **C010630** (2001) P512, [hep-ph/0202207].

[11] C. W. Bauer, S. P. Fleming, C. Lee and G. F. Sterman, *Factorization of e+e- Event Shape Distributions with Hadronic Final States in Soft Collinear Effective Theory*, *Phys. Rev.* **D78** (2008) 034027, [0801.4569].

[12] D. M. Hofman and J. Maldacena, *Conformal collider physics: Energy and charge correlations*, *JHEP* **05** (2008) 012, [0803.1467].

[13] V. Mateu, I. W. Stewart and J. Thaler, *Power Corrections to Event Shapes with Mass-Dependent Operators*, *Phys. Rev.* **D87** (2013) 014025, [1209.3781].

[14] A. V. Belitsky, S. Hohenegger, G. P. Korchemsky, E. Sokatchev and A. Zhiboedov, *From correlation functions to event shapes*, *Nucl. Phys.* **B884** (2014) 305–343, [1309.0769].

[15] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy flow polynomials: A complete linear basis for jet substructure*, *JHEP* **04** (2018) 013, [1712.07124].

[16] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121, [1810.05165].

[17] P. T. Komiske, E. M. Metodiev and J. Thaler, *Cutting Multiparticle Correlators Down to Size*, *Phys. Rev. D* **101** (2020) 036019, [1911.04491].

[18] S. Peleg, M. Werman and H. Rom, *A unified approach to the change of resolution: Space and gray-level*, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (1989) 739–742.

[19] Y. Rubner, C. Tomasi and L. J. Guibas, *A metric for distributions with applications to image databases*, in *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, (Washington, DC, USA), pp. 59–66, IEEE Computer Society, 1998, http://dl.acm.org/citation.cfm?id=938978.939133.

[20] Y. Rubner, C. Tomasi and L. J. Guibas, *The earth mover's distance as a metric for image retrieval*, *Int. J. Comput. Vision* **40** (Nov., 2000) 99–121.

[21] O. Pele and M. Werman, *A linear time histogram metric for improved SIFT matching*, in *Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III*, pp. 495–508, 2008, DOI.

[22] O. Pele and B. Taskar, *The tangent earth mover's distance*, in *Geometric Science of Information - First International Conference, GSI 2013, Paris, France, August 28-30, 2013. Proceedings*, pp. 397–404, 2013, DOI.

[23] L. N. Wasserstein, *Markov processes over denumerable products of spaces describing large systems of automata*, *Problems of Information Transmission* **5** (1969) 47–52.

[24] R. L. Dobrushin, *Prescribing a system of random variables by conditional distributions*, *Theory of Probability & Its Applications* **15** (1970) 458–486.

[25] P. T. Komiske, R. Mastandrea, E. M. Metodiev, P. Naik and J. Thaler, *Exploring the Space of Jets with CMS Open Data*, *Phys. Rev.* **D101** (2020) 034009, [1908.08542].

[26] A. Mullin, H. Pacey, M. Parker, M. White and S. Williams, *Does SUSY have friends? A new approach for LHC event analysis*, 1912.10625.

[27] T. Kinoshita, *Mass singularities of Feynman amplitudes*, *J. Math. Phys.* **3** (1962) 650–677.

[28] T. D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, *Phys. Rev.* **133** (1964) B1549–B1562.

[29] G. F. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, *Phys. Rev. Lett.* **39** (1977) 1436.

[30] G. F. Sterman, *Mass Divergences in Annihilation Processes. 1. Origin and Nature of Divergences in Cut Vacuum Polarization Diagrams*, *Phys. Rev.* **D17** (1978) 2773.

[31] G. F. Sterman, *Mass Divergences in Annihilation Processes. 2. Cancellation of Divergences in Cut Vacuum Polarization Diagrams*, *Phys. Rev.* **D17** (1978) 2789.

[32] G. F. Sterman, *Zero Mass Limit for a Class of Jet Related Cross-sections*, *Phys. Rev. D* **19** (1979) 3135.

[33] G. Sterman, J. Smith, J. C. Collins, J. Whitmore, R. Brock, J. Huston et al., *Handbook of perturbative qcd*, *Reviews of Modern Physics* **67** (1995) 157.

[34] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations.* Cambridge University Press, 2005.

[35] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and collider physics*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1–435.

[36] A. Banfi, G. P. Salam and G. Zanderighi, *Principles of general final-state resummation and automated implementation*, *JHEP* **03** (2005) 073, [`hep-ph/0407286`].

[37] A. J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **09** (2013) 137, [`1307.1699`].

[38] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [`1402.2657`].

[39] A. J. Larkoski, S. Marzani and J. Thaler, *Sudakov Safety in Perturbative QCD*, *Phys. Rev.* **D91** (2015) 111501, [`1502.01719`].

[40] S. Brandt, C. Peyrou, R. Sosnowski and A. Wroblewski, *The Principal axis of jets. An Attempt to analyze high-energy collisions as two-body processes*, *Phys. Lett.* **12** (1964) 57–61.

[41] E. Farhi, *A QCD Test for Jets*, *Phys. Rev. Lett.* **39** (1977) 1587–1588.

[42] H. Georgi and M. Machacek, *A Simple QCD Prediction of Jet Structure in e+ e- Annihilation*, *Phys. Rev. Lett.* **39** (1977) 1237.

[43] A. J. Larkoski, D. Neill and J. Thaler, *Jet Shapes with the Broadening Axis*, *JHEP* **04** (2014) 017, [`1401.2158`].

[44] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002, [`1004.2489`].

[45] C. Cesarotti and J. Thaler, *A Robust Measure of Event Isotropy at Colliders*, `2004.06125`.

[46] S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig and C. Lee, *Jet Shapes and Jet Algorithms in SCET*, *JHEP* **11** (2010) 101, [`1001.0014`].

[47] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [`1011.2268`].

[48] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093, [`1108.2701`].

[49] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion and T. F. Wilkason, *XCone: N-jettiness as an Exclusive Cone Jet Algorithm*, *JHEP* **11** (2015) 072, [`1508.01516`].

[50] J. Thaler and T. F. Wilkason, *Resolving Boosted Jets with XCone*, *JHEP* **12** (2015) 051, [`1508.01518`].

[51] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Longitudinally invariant $K_t$ clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B406** (1993) 187–224.

[52] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D48** (1993) 3160–3166, [`hep-ph/9305266`].

[53] D. Bertolini, T. Chan and J. Thaler, *Jet Observables Without Jet Algorithms*, *JHEP* **04** (2014) 013, [`1310.7584`].

[54] G. Salam, $E_t^\infty$ *Scheme*, *Unpublished* .

[55] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* **B659** (2008) 119–126, [`0707.1378`].

[56] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005, [`0802.1188`].

[57] G. Soyez, G. P. Salam, J. Kim, S. Dutta and M. Cacciari, *Pileup subtraction for jet shapes*, *Phys. Rev. Lett.* **110** (2013) 162001, [`1211.2811`].

[58] P. Berta, M. Spousta, D. W. Miller and R. Leitner, *Particle-level pileup subtraction for jets and jet shapes*, *JHEP* **06** (2014) 092, [`1403.3108`].

[59] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup Per Particle Identification*, *JHEP* **10** (2014) 059, [`1407.6013`].

[60] G. Soyez, *Pileup mitigation at the LHC: A theorist's view*, *Phys. Rept.* **803** (2019) 1–158, [`1801.09721`].

[61] P. Berta, L. Masetti, D. W. Miller and M. Spousta, *Pileup and Underlying Event Mitigation with Iterative Constituent Subtraction*, *JHEP* **08** (2019) 175, [`1905.03470`].

[62] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial and Applied Mathematics, USA, 10.1137/1.9780898719468.ch3.

[63] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, 10.1007/978-3-642-61798-0_4.

[64] J. Brewer, J. G. Milhano and J. Thaler, *Sorting out quenched jets*, *Phys. Rev. Lett.* **122** (2019) 222301, [`1812.05111`].

[65] CMS collaboration, C. Collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, .

[66] J. D. Bjorken and S. J. Brodsky, *Statistical Model for electron-Positron Annihilation Into Hadrons*, *Phys. Rev.* **D1** (1970) 1416–1420.

[67] L. Clavelli and D. Wyler, *Kinematical Bounds on Jet Variables and the Heavy Jet Mass Distribution*, *Phys. Lett.* **103B** (1981) 383–387.

[68] M. Dasgupta, A. Fregoso, S. Marzani and G. P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029, [`1307.0007`].

[69] J. Pumplin, *How to tell quark jets from gluon jets*, *Phys. Rev.* **D44** (1991) 2025–2032.

[70] C. F. Berger, T. Kucs and G. F. Sterman, *Event shape / energy flow correlations*, *Phys. Rev.* **D68** (2003) 014012, [`hep-ph/0303051`].

[71] G. Parisi, *Super Inclusive Cross-Sections*, *Phys. Lett. B* **74** (1978) 65–67.

[72] J. F. Donoghue, F. E. Low and S.-Y. Pi, *Tensor Analysis of Hadronic Jets in Quantum Chromodynamics*, *Phys. Rev.* **D20** (1979) 2759.

[73] R. Ellis, D. Ross and A. Terrano, *The Perturbative Calculation of Jet Structure in e+ e- Annihilation*, *Nucl. Phys. B* **178** (1981) 421–456.

[74] S. Catani and B. R. Webber, *Infrared safe but infinite: Soft gluon divergences inside the physical region*, *JHEP* **10** (1997) 005, [`hep-ph/9710333`].

[75] A. J. Larkoski, G. P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108, [`1305.0007`].

[76] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009, [`1409.6298`].

[77] I. Moult, L. Necib and J. Thaler, *New Angles on Energy Correlation Functions*, *JHEP* **12** (2016) 153, [`1609.07483`].

[78] C. Frye, A. J. Larkoski, M. D. Schwartz and K. Yan, *Precision physics with pile-up insensitive observables*, `1603.06375`.

[79] C. Frye, A. J. Larkoski, M. D. Schwartz and K. Yan, *Factorization for groomed jet substructure beyond the next-to-leading logarithm*, *JHEP* **07** (2016) 064, [`1603.09338`].

[80] S. Marzani, L. Schunk and G. Soyez, *A study of jet mass distributions with grooming*, *JHEP* **07** (2017) 132, [`1704.02210`].

[81] S. Marzani, L. Schunk and G. Soyez, *The jet mass distribution after Soft Drop*, *Eur. Phys. J.* **C78** (2018) 96, [`1712.05105`].

[82] A. H. Hoang, S. Mantry, A. Pathak and I. W. Stewart, *Nonperturbative Corrections to Soft Drop Jet Mass*, *JHEP* **12** (2019) 002, [`1906.11843`].

[83] ATLAS COLLABORATION collaboration, *Impact of Alternative Inputs and Jet Grooming on Large-R Jet Performance*, Tech. Rep. ATL-PHYS-PUB-2019-027, CERN, Geneva, Jul, 2019.

[84] ATLAS collaboration, G. Aad et al., *Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s}$ =13 TeV*, *Phys. Rev. D* **101** (2020) 052007, [`1912.09837`].

[85] G. F. Sterman, *Review of theoretical status: The Long and short of high energy Jets*, *eConf* **C0601121** (2006) R001, [`hep-ph/0606032`].

[86] B. T. Elder, M. Procura, J. Thaler, W. J. Waalewijn and K. Zhou, *Generalized Fragmentation Functions for Fractal Jet Observables*, *JHEP* **06** (2017) 085, [`1704.05456`].

[87] S. Alioli, C. W. Bauer, C. J. Berggren, A. Hornig, F. J. Tackmann, C. K. Vermilion et al., *Combining Higher-Order Resummation with Multiple NLO Calculations and Parton Showers in GENEVA*, *JHEP* **09** (2013) 120, [`1211.7049`].

[88] S. Alioli, C. W. Bauer, C. Berggren, F. J. Tackmann, J. R. Walsh and S. Zuberi, *Matching Fully Differential NNLO Calculations and Parton Showers*, *JHEP* **06** (2014) 089, [`1311.0286`].

[89] S. Alioli, C. W. Bauer, C. Berggren, F. J. Tackmann and J. R. Walsh, *Drell-Yan production at NNLL'+NNLO matched to parton showers*, *Phys. Rev.* **D92** (2015) 094020, [`1508.01475`].

[90] A. J. Larkoski and I. Moult, *The Singular Behavior of Jet Substructure Observables*, *Phys. Rev.* **D93** (2016) 014017, [`1510.08459`].

[91] A. De Rujula, J. R. Ellis, E. G. Floratos and M. K. Gaillard, *QCD Predictions for Hadronic Final States in e+ e- Annihilation*, *Nucl. Phys.* **B138** (1978) 387–429.

[92] D. P. Barber et al., *Tests of Quantum Chromodynamics and a Direct Measurement of the Strong Coupling Constant $\alpha_S$ at $\sqrt{s} = 30$-GeV*, *Phys. Lett.* **89B** (1979) 139–144.

[93] JADE collaboration, W. Bartel et al., *Observation of Planar Three Jet Events in e+ e- Annihilation and Evidence for Gluon Bremsstrahlung*, *Phys. Lett.* **91B** (1980) 142–147.

[94] TASSO collaboration, M. Althoff et al., *Jet Production and Fragmentation in e+ e- Annihilation at 12-GeV to 43-GeV*, *Z. Phys.* **C22** (1984) 307–340.

[95] D. Bender et al., *Study of Quark Fragmentation at 29-GeV: Global Jet Parameters and Single Particle Distributions*, *Phys. Rev.* **D31** (1985) 1.

[96] MARK-II collaboration, G. S. Abrams et al., *First Measurements of Hadronic Decays of the Z Boson*, *Phys. Rev. Lett.* **63** (1989) 1558.

[97] AMY collaboration, Y. K. Li et al., *Multi - hadron event properties in $e^+e^-$ annihilation at $\sqrt{s} = 52$ GeV to 57-GeV*, *Phys. Rev.* **D41** (1990) 2675.

[98] ALEPH collaboration, D. Decamp et al., *Measurement of alpha-s from the structure of particle clusters produced in hadronic Z decays*, *Phys. Lett.* **B257** (1991) 479–491.

[99] TASSO collaboration, W. Braunschweig et al., *Global Jet Properties at 14-GeV to 44-GeV Center-of-mass Energy in $e^+e^-$ Annihilation*, *Z. Phys.* **C47** (1990) 187–198.

[100] SLD collaboration, K. Abe et al., *Measurement of alpha-s (M(Z)\*\*2) from hadronic event observables at the Z0 resonance*, *Phys. Rev.* **D51** (1995) 962–984, [`hep-ex/9501003`].

[101] ALEPH collaboration, A. Heister et al., *Studies of QCD at e+ e- centre-of-mass energies between 91-GeV and 209-GeV*, *Eur. Phys. J.* **C35** (2004) 457–486.

[102] DELPHI collaboration, J. Abdallah et al., *A Study of the energy evolution of event shape distributions and their means with the DELPHI detector at LEP*, *Eur. Phys. J.* **C29** (2003) 285–312, [`hep-ex/0307048`].

[103] L3 collaboration, P. Achard et al., *Studies of hadronic event structure in $e^+e^-$ annihilation from 30-GeV to 209-GeV with the L3 detector*, *Phys. Rept.* **399** (2004) 71–174, [`hep-ex/0406049`].

[104] OPAL collaboration, G. Abbiendi et al., *Measurement of event shape distributions and moments in e+ e- —¿ hadrons at 91-GeV - 209-GeV and a determination of alpha(s)*, *Eur. Phys. J.* **C40** (2005) 287–316, [`hep-ex/0503051`].

[105] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover and G. Heinrich, *Second-order QCD corrections to the thrust distribution*, *Phys. Rev. Lett.* **99** (2007) 132002, [`0707.1285`].

[106] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover and G. Heinrich, *NNLO corrections to event shapes in e+ e- annihilation*, *JHEP* **12** (2007) 094, [0711.4711].

[107] T. Becher and M. D. Schwartz, *A precise determination of $\alpha_s$ from LEP thrust data using effective field theory*, *JHEP* **07** (2008) 034, [0803.0342].

[108] S. Weinzierl, *Event shapes and jet rates in electron-positron annihilation at NNLO*, *JHEP* **06** (2009) 041, [0904.1077].

[109] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu and I. W. Stewart, *Thrust at $N^3LL$ with Power Corrections and a Precision Global Fit for alphas(mZ)*, *Phys. Rev.* **D83** (2011) 074021, [1006.3080].

[110] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu and I. W. Stewart, *Precision Thrust Cumulant Moments at $N^3LL$*, *Phys. Rev.* **D86** (2012) 094002, [1204.5746].

[111] P. E. L. Rakow and B. R. Webber, *Transverse Momentum Moments of Hadron Distributions in QCD Jets*, *Nucl. Phys.* **B191** (1981) 63–74.

[112] R. K. Ellis and B. R. Webber, *QCD Jet Broadening in Hadron Hadron Collisions*, *Conf. Proc.* **C860623** (1986) 74.

[113] S. Catani, G. Turnock and B. R. Webber, *Jet broadening measures in $e^+e^-$ annihilation*, *Phys. Lett.* **B295** (1992) 269–276.

[114] Y. L. Dokshitzer, A. Lucenti, G. Marchesini and G. P. Salam, *On the QCD analysis of jet broadening*, *JHEP* **01** (1998) 011, [hep-ph/9801324].

[115] S. Brandt and H. D. Dahmen, *Axes and Scalar Measures of Two-Jet and Three-Jet Events*, *Z. Phys.* **C1** (1979) 61.

[116] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Factorization at the LHC: From PDFs to Initial State Jets*, *Phys. Rev.* **D81** (2010) 094035, [0910.0467].

[117] C. F. Berger, C. Marcantonini, I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Higgs Production with a Central Jet Veto at NNLL+NNLO*, *JHEP* **04** (2011) 092, [1012.4480].

[118] T. T. Jouttenus, I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *Jet mass spectra in Higgs boson plus one jet at next-to-next-to-leading logarithmic order*, *Phys. Rev.* **D88** (2013) 054031, [1302.0846].

[119] J. Rabin, J. Delon and Y. Gousseau, *Transportation distances on the circle*, *Journal of Mathematical Imaging and Vision* **41** (2011) 147–167.

[120] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung and J. Virzi, *Substructure of high-pT Jets at the LHC*, *Phys. Rev.* **D79** (2009) 074017, [0807.0234].

[121] A. J. Larkoski, J. Thaler and W. J. Waalewijn, *Gaining (Mutual) Information about Quark/Gluon Discrimination*, *JHEP* **11** (2014) 129, [1408.3122].

[122] 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*, *Phys. Rept.* **841** (2020) 1–63, [1709.04464].

[123] R. Kogler et al., *Jet Substructure at the Large Hadron Collider: Experimental Review*, *Rev. Mod. Phys.* **91** (2019) 045003, [1803.06991].

[124] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, vol. 958. Springer, 2019, 10.1007/978-3-030-15709-8.

[125] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073, [1704.08249].

[126] K. Datta and A. J. Larkoski, *Novel Jet Observables from Machine Learning*, *JHEP* **03** (2018) 086, [1710.01305].

[127] A. J. Larkoski and E. M. Metodiev, *A Theory of Quark vs. Gluon Discrimination*, *JHEP* **10** (2019) 014, [1906.01639].

[128] J. Thaler, *Jet maximization, axis minimization, and stable cone finding*, *Phys. Rev.* **D92** (2015) 074001, [1506.07876].

[129] G. C. Blazey et al., *Run II jet physics*, in *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*.

[130] S. D. Ellis, J. Huston and M. Tonnesmann, *On building better cone jet algorithms*, *eConf* **C010630** (2001) 513, [hep-ph/0111434].

[131] G. P. Salam and G. Soyez, *A Practical Seedless Infrared-Safe Cone jet algorithm*, *JHEP* **05** (2007) 086, [0704.0292].

[132] H. Georgi, *A Simple Alternative to Jet-Clustering Algorithms*, 1408.1161.

[133] S.-F. Ge, *The Georgi Algorithms of Jet Clustering*, *JHEP* **05** (2015) 066, [1408.3823].

[134] Y. Bai, Z. Han and R. Lu, $J_{E_T}$: *a global jet finding algorithm*, *JHEP* **03** (2015) 102, [1411.3705].

[135] Y. Bai, Z. Han and R. Lu, $J_{E_T}^{\mathrm{II}}$: *A Two-prong Jet Finding Algorithm*, 1509.07522.

[136] A. Y. Wei, P. Naik, A. W. Harrow and J. Thaler, *Quantum Algorithms for Jet Clustering*, 1908.08949.

[137] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].

[138] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896, [1111.6097].

[139] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[140] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279, 1998, hep-ph/9907280.

[141] J. M. Butterworth, J. P. Couchman, B. E. Cox and B. M. Waugh, *KtJet: A C++ implementation of the K-perpendicular clustering algorithm*, *Comput. Phys. Commun.* **153** (2003) 85–96, [hep-ph/0210022].

[142] M. Dasgupta, L. Schunk and G. Soyez, *Jet shapes for boosted jet two-prong decays from first-principles*, *JHEP* **04** (2016) 166, [1512.00516].

[143] D. Krohn, M. D. Schwartz, M. Low and L.-T. Wang, *Jet Cleansing: Pileup Removal at High Luminosity*, *Phys. Rev.* **D90** (2014) 065020, [1309.4777].

[144] M. Cacciari, G. P. Salam and G. Soyez, *Use of charged-track information to subtract neutral pileup*, *Phys. Rev.* **D92** (2015) 014003, [1404.7353].

[145] M. Cacciari, G. P. Salam and G. Soyez, *SoftKiller, a particle-level pileup removal method*, *Eur. Phys. J.* **C75** (2015) 59, [1407.0408].

[146] P. T. Komiske, E. M. Metodiev, B. Nachman and M. D. Schwartz, *Pileup Mitigation with Machine Learning (PUMML)*, *JHEP* **12** (2017) 051, [1707.08600].

[147] P. Hansen, J. W. Monk and C. Wiglesworth, *A Wavelet Based Pile-Up Mitigation Method for the LHC Upgrade*, 1812.07412.

[148] J. Arjona Martínez, O. Cerri, M. Pierini, M. Spiropulu and J.-R. Vlimant, *Pileup mitigation at the Large Hadron Collider with graph neural networks*, *Eur. Phys. J. Plus* **134** (2019) 333, [1810.07988].

[149] V. Hartmann and D. Schuhmacher, *Semi-discrete optimal transport-the case p= 1*, 1706.07650.

[150] L. Ambrosio, N. Gigli and G. Savare, *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Birkhäuser Basel, 1 ed., 2005, 10.1007/b137080.

[151] F. Aurenhammer, R. Klein and D.-T. Lee, *Voronoi Diagrams and Delaunay Triangulations.* World Scientific, 2013, 10.1142/8685.

[152] C. Collaboration, *Jet primary dataset in AOD format from RunA of 2011 (/Jet/Run2011A-12Oct2013-v1/AOD)*, *CERN Open Data Portal* (2016) .

[153] M. Liero, A. Mielke and G. Savaré, *Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures*, *Inventiones mathematicae* **211** (2018) 969–1117.

[154] D. P. Bourne, B. Schmitzer and B. Wirth, *Semi-discrete unbalanced optimal transport and quantization*, 1808.01962.

[155] M. I. Karavelas and M. Yvinec, *Dynamic additively weighted voronoi diagrams in 2d*, in *Algorithms - ESA 2002, 10th Annual European Symposium, Rome, Italy, September 17-21, 2002, Proceedings* (R. H. Möhring and R. Raman, eds.), vol. 2461 of *Lecture Notes in Computer Science*, pp. 586–598, Springer, 2002, DOI.

[156] D. Geiß, R. Klein, R. Penninger and G. Rote, *Optimally solving a transportation problem using voronoi diagrams*, *Comput. Geom.* **46** (2013) 1009–1016.

[157] S. Xin, B. Lévy, Z. Chen, L. Chu, Y. Yu, C. Tu et al., *Centroidal power diagrams with capacity constraints: computation, applications, and extension*, *ACM Trans. Graph.* **35** (2016) 244:1–244:12.

[158] J. Erdmenger, K. T. Grosvenor and R. Jefferson, *Information geometry in quantum field theory: lessons from simple examples*, 2001.02683.

[159] C. Cheung, *TASI Lectures on Scattering Amplitudes*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle*

*Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*, pp. 571–623, 2018, 1708.03872, DOI.

[160] C. Frye, H. Hannesdottir, N. Paul, M. D. Schwartz and K. Yan, *Infrared Finiteness and Forward Scattering*, *Phys. Rev.* **D99** (2019) 056015, [1810.10022].

[161] H. Hannesdottir and M. D. Schwartz, *A Finite S-Matrix*, 1906.03271.

[162] H. Hannesdottir and M. D. Schwartz, *An S-Matrix for Massless Particles*, 1911.06821.

[163] I. Jubran, A. Maalouf and D. Feldman, *Introduction to Coresets: Accurate Coresets*, *arXiv e-prints* (Oct., 2019) arXiv:1910.08707, [1910.08707].

[164] A. W. Harrow, *Small quantum computers and large classical data sets*, 2004.00026.

[165] S. Claici and J. Solomon, *Wasserstein coresets for lipschitz costs*, *CoRR* (2018) , [1805.07412].

[166] S. Bianchini and A. Brancolini, *Estimates on path functionals over wasserstein spaces*, *SIAM J. Math. Analysis* **42** (2010) 1179–1217.

[167] M. Agueh and G. Carlier, *Barycenters in the wasserstein space*, *SIAM J. Math. Analysis* **43** (2011) 904–924.

[168] J. Bertrand and B. R. Kloeckner, *A geometric study of wasserstein spaces: An addendum on the boundary*, in *Geometric Science of Information - First International Conference, GSI 2013, Paris, France, August 28-30, 2013. Proceedings*, pp. 405–412, 2013, DOI.

[169] T. L. Gouic and J. Loubes, *Barycenter in wasserstein spaces: Existence and consistency*, in *Geometric Science of Information - Second International Conference, GSI 2015, Palaiseau, France, October 28-30, 2015, Proceedings*, pp. 104–108, 2015, DOI.

[170] C. W. Bauer, A. Hornig and F. J. Tackmann, *Factorization for generic jet production*, *Phys. Rev.* **D79** (2009) 114013, [0808.2191].

[171] A. V. Belitsky, S. Hohenegger, G. P. Korchemsky, E. Sokatchev and A. Zhiboedov, *Energy-Energy Correlations in N=4 Supersymmetric Yang-Mills Theory*, *Phys. Rev. Lett.* **112** (2014) 071601, [1311.6800].

[172] L. J. Dixon, I. Moult and H. X. Zhu, *Collinear limit of the energy-energy correlator*, *Phys. Rev.* **D100** (2019) 014009, [1905.01310].

[173] H. Chen, M.-X. Luo, I. Moult, T.-Z. Yang, X. Zhang and H. X. Zhu, *Three Point Energy Correlators in the Collinear Limit: Symmetries, Dualities and Analytic Results*, 1912.11050.

[174] H. Elvang and Y.-t. Huang, *Scattering Amplitudes*, 1308.1697.

[175] J. J. M. Carrasco, *Gauge and Gravity Amplitude Relations*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pp. 477–557, WSP, WSP, 2015, 1506.00974, DOI.

[176] M. C. Romao, N. Castro, J. Milhano, R. Pedro and T. Vale, *Use of a Generalized Energy Mover's Distance in the Search for Rare Phenomena at Colliders*, 2004.09360.

[177] A. Banfi, G. P. Salam and G. Zanderighi, *Infrared safe definition of jet flavor*, *Eur. Phys. J. C* **47** (2006) 113–124, [hep-ph/0601139].

[178] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel et al., *Machine learning at the energy and intensity frontiers of particle physics*, *Nature* **560** (08, 2018) .

[179] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby et al., *Machine learning and the physical sciences*, *Rev. Mod. Phys.* **91** (2019) 045002, [1903.10563].

[180] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty Detection Meets Collider Physics*, 1807.10261.

[181] T. Heimel, G. Kasieczka, T. Plehn and J. M. Thompson, *QCD or What?*, *SciPost Phys.* **6** (2019) 030, [1808.08979].

[182] M. Farina, Y. Nakai and D. Shih, *Searching for New Physics with Deep Autoencoders*, 1808.08992.

[183] T. S. Roy and A. H. Vijay, *A robust anomaly finder based on autoencoder*, 1903.02032.