# Cutting multiparticle correlators down to size

Patrick T. Komiske,[*] Eric M. Metodiev,[†] and Jesse Thaler[‡]
*Center for Theoretical Physics, Massachusetts Institute of Technology,*
*Cambridge, Massachusetts 02139, USA*
*and Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 17 November 2019; accepted 4 February 2020; published 24 February 2020)

Multiparticle correlators are mathematical objects frequently encountered in quantum field theory and collider physics. By translating multiparticle correlators into the language of graph theory, we can gain new insights into their structure as well as identify efficient ways to manipulate them. We highlight the power of this graph-theoretic approach by "cutting open" the vertices and edges of the graphs, allowing us to systematically classify linear relations among multiparticle correlators and develop faster methods for their computation. The naive computational complexity of an $N$-point correlator among $M$ particles is $\mathcal{O}(M^N)$, but when the pairwise distances between particles can be cast as an inner product, we show that all such correlators can be computed in linear $\mathcal{O}(M)$ run-time. With the help of new tensorial objects called energy flow moments, we achieve a fast implementation of jet substructure observables like $C_2$ and $D_2$, which are widely used at the Large Hadron Collider to identify boosted hadronic resonances. As another application, we compute the number of leafless multigraphs with $d$ edges up to $d = 16$ $(15, 641, 159)$, conjecturing that this is the same as the number of independent kinematic polynomials of degree $d$, previously known only to $d = 8$ (279).

DOI: 10.1103/PhysRevD.101.036019

## I. INTRODUCTION

Multiparticle correlators are ubiquitous mathematical structures that appear in a variety of physics domains, including kinematic polynomials for scattering amplitudes [1,2], operator bases for effective field theories [3,4], and many-body expansions for molecular analyses [5,6]. Broadly speaking, multiparticle correlators appear whenever the fundamental entities of a system are arranged in sets: unordered, variable-length collections of objects. These set elements may be nuclei, atoms, fields, particles, or other objects. Here, we refer to them as "particles," since we will eventually focus on multiparticle correlators in collider physics [7–9].

The basic structure of an $N$-point multiparticle correlator of $M$ particles is

$$\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\cdots\sum_{i_N=1}^{M} z_{i_1}z_{i_2}\cdots z_{i_N} f(p_{i_1}, p_{i_2}, \ldots, p_{i_N}), \qquad (1)$$

where $z_i$ are particle weights, $p_i$ are particle properties, and $f$ is a function of these properties. In the particle physics literature, Eq. (1) appears as the $C$-correlators of Ref. [10]. The function $f$ is often decomposed in terms of monomials of pairwise distances or invariants $\theta_{ij}$ between the particles:

$$f(p_{i_1}, \ldots, p_{i_N}) = \theta_{i_1 i_2}^{\alpha_{12}}\theta_{i_1 i_3}^{\alpha_{13}}\cdots\theta_{i_{N-1} i_N}^{\alpha_{N-1N}}, \qquad (2)$$

where $\alpha_{ij}$ are integer exponents. An argument for the generality of this restriction (up to isometries) in the context of collider observables is given in Ref. [9]. This simplification allows multiparticle correlators to be decomposed into smaller pieces, which is essential for deriving the results below.

In this paper, we develop the theory of multiparticle correlators by representing them as multigraphs, allowing us to obtain many useful results by manipulating their vertices and edges. Specifically, by "cutting" the multigraphs along either their vertices or edges, we derive otherwise opaque linear relations among sets of correlators. Furthermore, in special cases where the pairwise distances $\theta_{ij}$ take the form of an inner product, we show that the complexity of computing a multiparticle correlator can be reduced from $\mathcal{O}(M^N)$ down to the minimally required $\mathcal{O}(M)$. We highlight various connections of these mathematical results to high-energy physics, including counting kinematic polynomials relevant for superstring amplitudes [1,2] and speeding up the computation of jet substructure observables [7–9].

---

[*] pkomiske@mit.edu
[†] metodiev@mit.edu
[‡] jthaler@mit.edu

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.*

A detailed outline of this paper is as follows. In Sec. II, we summarize the basics of multiparticle correlators and review the multigraph-multiparticle correlator correspondence of Ref. [9]. Several well-known examples of multiparticle correlators are discussed, including kinematic polynomials and energy flow polynomials (EFPs) [9]. We also review the computational complexity of multiparticle correlators and show how the simplification in Eq. (2) allows us to improve upon the naive $\mathcal{O}(M^N)$ scaling using the variable elimination algorithm [9,11].

In Sec. III, we introduce a method to slice the vertices of multigraphs. By cutting vertices in half, the multiparticle correlator in Eq. (1) can be decomposed into contractions of "particle tensors" whose indices run from 1 to $M$ (the number of particles). We then derive linear relations among multiparticle correlators when $M$ is small compared to $N$. The key for understanding these linear relations is a well-known tensor antisymmetrization identity. In $m$ dimensions, antisymmetrizing any tensor over $\ell > m$ indices yields zero:

$$T^{a_1\cdots a_j}_{b_1\cdots b_k[c_1\cdots c_\ell]} = 0. \tag{3}$$

This follows immediately from the fact that any assignment of the $m$ possible values to the $\ell > m$ indices must have a repetition and therefore vanish. These particle tensors also enable further computational speedups through methods such as fast matrix multiplication and dynamic programming.

In Sec. IV, we explore the consequences of making an additional simplifying assumption, namely that the pairwise distances $\theta_{ij}$ are given by an inner product,

$$\theta_{ij} = \eta_{\mu\nu} u_i^\mu u_j^\nu, \tag{4}$$

for some vectors $u_i^\mu, u_j^\nu$ and metric $\eta_{\mu\nu}$ with the Einstein summation convention. This assumption applies, for instance, to kinematic polynomials of Mandelstam invariants and EFPs with angular exponent $\beta = 2$. Having made this assumption, we can slice the multigraph edges and the multiparticle correlators can be expressed as contractions of "moment tensors." This allows us to derive new linear relations and computational speedups by cutting the edges of the multigraphs. The antisymmetrization identity in Eq. (3), now applied to the moment tensors, explains redundancies that appear when the dimension of the inner product space is small compared to the number of edges. Multiparticle correlators can now be computed in $\mathcal{O}(M)$, since they are sewn together from moment tensors, each of which is $\mathcal{O}(M)$ to compute. This result generalizes the well-known fact that the invariant mass of a set of massless particles does not require $\mathcal{O}(M^2)$ time to compute, since one can simply sum over the four-vectors in $\mathcal{O}(M)$:

$$\sum_{i=1}^{M}\sum_{j=1}^{M} p_i \cdot p_j = \left(\sum_{i=1}^{M} p_i^\mu\right)^2. \tag{5}$$

This reduced complexity is particularly relevant for quantum algorithms for collider physics, where preprocessing and loading classical data into a quantum computer can be a key computational bottleneck [12].

In Sec. V, we show how momentum conservation can be used to trim away valency-1 leaves from the multigraphs. As an interesting physical application, we use these trimmed moment tensors to count independent kinematic polynomials up to degree $d$, which is relevant for enumerating superstring amplitudes [1]. By translating this problem to our graphical language, we are able to significantly extend existing integer sequences by simply enumerating leafless multigraphs, advancing previous results from $d = 8$ [1,2] to $d = 16$ [13,14].

In Sec. VI, we highlight the relevance of our results for collider physics. As shown in Ref. [9], the EFPs fully capture the infrared-and-collinear-safe (IRC-safe) information in the radiation pattern of an event, for any choice of angular exponent $\beta$. For the special choice of $\beta = 2$, the EFPs can be written as Lorentz contractions of energy flow moments (EFMs), which are IRC safe by construction. An EFM with $v$ indices takes the form

$$\mathcal{I}^{\mu_1\cdots\mu_v} = 2^{v/2} \sum_{i=1}^{M} E_i n_i^{\mu_1} \cdots n_i^{\mu_v}, \tag{6}$$

where $E_i$ are particle energies, $p_i^\mu$ are massless particle momenta, and $n_i^\mu \equiv p_i^\mu / E_i$. The EFMs are introduced for the first time in this paper, though we show that they are closely related to other moment tensors appearing previously in the literature. Using the EFMs, we demonstrate that the computationally expensive jet substructure observables $C_2$ [7] and $D_2$ [8], widely applied for new physics searches at the Large Hadron Collider (LHC), can significantly benefit from the computational speedups developed here. We make the fast implementations of these observables available in the ENERGYFLOW PYTHON package [15], which also provides tools to compute any multiparticle correlator.

Our conclusions are presented in Sec. VII. We derive additional linear relations for multiparticle correlators in the context of $e^+e^-$ collisions in the Appendix.

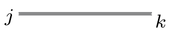
## II. MULTIPARTICLE CORRELATORS

This section summarizes the basic properties of multiparticle correlators. We review the correspondence between multiparticle correlators and multigraphs, provide two key examples of multiparticle correlators relevant for high-energy physics, and discuss previously known techniques to decrease their computational complexity.

### A. Correlators as graphs

As shown in Ref. [9], multigraphs are an efficient and intuitive way to represent multiparticle correlators. Each of the $N$ sums and weight factors in Eq. (1) corresponds to a vertex, and each pairwise distance factor in Eq. (2) corresponds to an edge. These rules can be graphically summarized as

$$_i = \sum_{i=1}^{M} z_i,$$



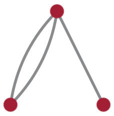
$$_j \quad\quad _k = \theta_{jk}. \tag{7}$$

The "multi" in multigraph refers to the fact that two vertices can be connected by more than one edge.

In this language, a multiparticle correlator with graph $G$ is written as

$$\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\cdots\sum_{i_N=1}^{M} z_{i_1}\cdots z_{i_N}\prod_{(k,\ell)\in G}\theta_{i_k i_\ell}, \tag{8}$$

where $(k,\ell)$ are the pairs of vertices connected by the edges in $G$. As an example, consider the following multigraph and its corresponding correlator:



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1}z_{i_2}z_{i_3}\theta_{i_1 i_2}^2\theta_{i_1 i_3}. \tag{9}$$

This correlator has three sums corresponding to the three vertices of the multigraph. Each of the edges contributes a factor of $\theta_{ij}$ to the argument, with $\theta_{i_2 i_3}$ absent due to no edge connecting two of the vertices. Any permutation $\sigma$ of $i_1, \ldots, i_N$ to $i_{\sigma(1)}, \ldots, i_{\sigma(N)}$ yields an identical correlator by the symmetry of the sum structure. Graphically, this translates to the fact that isomorphic multigraphs correspond to the same correlator, which allows us to write the graphs without labels on the vertices.

To quantify the number of unique analytic structures in the correlators, it is helpful to organize them by the degree $d$ of the monomial, or equivalently the number of edges in the associated multigraph. The number of unique multiparticle correlators of each $d$ up to $d = 7$ is tabulated in Table I based on entries in the On-Line Encyclopedia of Integer Sequences (OEIS) [16]. As an explicit demonstration, Table II shows the expressions and multigraphs for all 13 multiparticle correlators with degree $d \leq 3$. Correlators corresponding to connected multigraphs are also separately tabulated, as disconnected ones are simply the product of their connected components.

TABLE I. The number of multiparticle correlators of degree $d$. These are equivalent to the number of nonisomorphic multigraphs with $d$ edges, tabulated for connected and all multigraphs.

| | Degree $d$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| Connected | A076864 [17] | 1 | 1 | 2 | 5 | 12 | 33 | 103 | 333 |
| | Cumulative | 1 | 2 | 4 | 9 | 21 | 54 | 157 | 490 |
| All | A050535 [18] | 1 | 1 | 3 | 8 | 23 | 66 | 212 | 686 |
| | Cumulative | 1 | 2 | 5 | 13 | 36 | 102 | 314 | 1 000 |

### B. Kinematic polynomials

As one application in high-energy physics, multiparticle correlators arise in the study of kinematic polynomials. These are symmetric polynomials built from Mandelstam invariants:

$$s_{ij} = (p_i^\mu + p_j^\mu)^2 = 2p_i^\mu p_{j\mu}, \tag{10}$$

where the last equality assumes massless particles ($p^2 = 0$). Kinematic polynomials have been used to build operator bases [3,4] for quantum field theories and understand the structure of scattering amplitudes [1].

In our notation, kinematic polynomials are multiparticle correlators of the form

$$z_i = 1, \tag{11}$$

$$\theta_{ij} = s_{ij}, \tag{12}$$

and so the insights we develop here are directly applicable to this context.

It is worth noting that an analog of the graphical notation for multiparticle correlators was also developed in Ref. [3] and mentioned in Ref. [4]. Beyond this, Hogervorst *et al.* [3] remark that there are linear relations among the kinematic polynomials which can complicate certain analyses, and that a classification of those relations would be required to pursue particular strategies. We will explore these relations among multiparticle correlators at length, seeking to understand and classify these relations.

### C. Energy flow polynomials

A family of multiparticle correlators relevant for collider physics is the set of EFPs [9]. Explicitly, EFPs are defined as multiparticle correlators of the form

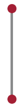
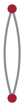
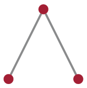
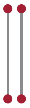
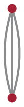
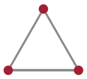
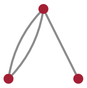
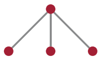
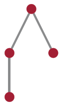
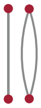
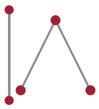
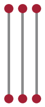
$$z_i = E_i, \tag{13}$$

$$\theta_{ij} = (2n_i^\mu n_{j\mu})^{\beta/2}, \qquad n_i^\mu \equiv p_i^\mu / E_i, \tag{14}$$

where $E_i$ is the energy of particle $i$, $n_i^\mu = (1, \hat{n}_i)^\mu$ for massless particles, and $\beta$ is an angular weighting factor. While we use particle energies $E_i$ here to simplify the notation, in a hadron collider context these would be replaced with particle transverse momenta $p_{Ti}$.

The EFPs are IRC-safe observables, which are guaranteed to be finite and computable in perturbative quantum field theory [19–23]. An observable is IRC safe if it is unchanged by the addition of a soft particle with $E \to 0$ or

TABLE II. All distinct multiparticle correlators with degree $d \leq 3$, shown as multigraphs and written explicitly.

| Degree | Multigraph | | Multiparticle correlator |
|---|---|---|---|
| $d = 0$ | | $=$ | $\sum_{i=1}^{M} z_i$ |
| $d = 1$ |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} \theta_{i_1 i_2}$ |
| $d = 2$ |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} \theta_{i_1 i_2}^2$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2} \theta_{i_1 i_3}$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_3 i_4}$ |
| $d = 3$ |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} \theta_{i_1 i_2}^3$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_1 i_3}$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_1 i_3}$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_1 i_3} \theta_{i_1 i_4}$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_3 i_4}$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_3 i_4}^2$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} \sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2} \theta_{i_3 i_4} \theta_{i_3 i_5}$ |
| |  | $=$ | $\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} \sum_{i_5=1}^{M} \sum_{i_6=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} z_{i_6} \theta_{i_1 i_2} \theta_{i_3 i_4} \theta_{i_5 i_6}$ |

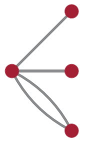
by the collinear splitting of one particle into two with $p^\mu \to \{\lambda p^\mu, (1-\lambda) p^\mu\}$ for any $\lambda \in [0,1]$. While Ref. [9] shows via direct computation that the EFPs are IRC safe, we will later develop a simple notation which makes this fact manifest in the $\beta = 2$ case (see Sec. VI A).

For any choice of $\beta > 0$, the EFPs form a linear basis of all IRC-safe observables [9], meaning that any IRC-safe observable can be approximated by a finite linear combination of EFPs. Many common collider observables are encompassed by the EFPs [7,24] or can be cast as exact linear combinations of EFPs [25–29]. While it has been shown that the EFP basis can be useful for jet classification [9,30], the basis is overcomplete even for fixed $\beta$, in the sense that there are linear relations among the elements, which complicates the use of linear fitting methods. Understanding these relations is a key goal of this paper, which we will accomplish in Secs. III C and IV C.

### D. Computational complexity

An $N$-particle correlator in Eq. (8) is naively computable in $\mathcal{O}(M^N)$ by evaluating the $N$ nested sums over $M$ particles. This computational complexity can typically be significantly improved upon, however, by exploiting the algebraic structure of the sums, as we have detailed in Ref. [9] and summarize here. Efficiently computing multiparticle correlators is especially important for collider physics applications, where hundreds to thousands of final-state particles are produced in each collision.

The key insight is to iteratively perform the sums in a carefully chosen order. For example, we can write the following correlator in a suggestive way:



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2}\theta_{i_1 i_3}\theta^2_{i_1 i_4}$$

$$= \sum_{i_1=1}^{M} z_{i_1}\left(\sum_{i_2=1}^{M} z_{i_2}\theta_{i_1 i_2}\right)^2\left(\sum_{i_4=1}^{M} z_{i_4}\theta^2_{i_1 i_4}\right), \tag{15}$$

which can be evaluated in $\mathcal{O}(M^2)$ by computing the parenthetical quantities in $\mathcal{O}(M)$ for all $M$ terms in the outer sum. This strategy gives a significant improvement over the naive $\mathcal{O}(M^4)$ complexity of the correlator.

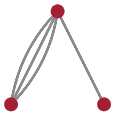
The general procedure to determine the order in which to compute the nested sums is known as the variable elimination algorithm [11]. Using variable elimination yields a computational complexity of $\mathcal{O}(M^{\mathrm{tw}(G)+1})$, where $\mathrm{tw}(G)$ is the treewidth of $G$, neglecting multiple edges. The treewidth measures how "treelike" a graph is, with $\mathrm{tw}(G) = 1$ for trees and higher values for other graphs. For all noncomplete graphs, $\mathrm{tw}(G) + 1 < N$ and so variable elimination provides a significant improvement in computational speed in most cases. Here, we see the graphical notation is not only useful for enumerating all possible correlators, but it also provides a natural strategy for efficiently evaluating the correlators. In the next two sections, we exploit this graphical language for further computational gains.

## III. CUTTING OPEN VERTICES

In this section, we begin to surgically disassemble the multigraphs by slicing them through their vertices. This yields particle tensors whose indices index particles and allows us to view the correlators as contractions over these indices. Using these particle tensors, we can understand linear relations that emerge when the number of particles is small as well as speed up certain computations beyond what is possible through variable elimination alone. These results hold for any multiparticle correlator of the form in Eq. (8), regardless of the choice of pairwise distances $\theta_{ij}$.

### A. Particle tensors

One can view the sums over particle indices $i_1, \ldots, i_N$ in Eq. (8) as contractions of tensorial objects. We can explicitly find these tensors by slicing open the multigraph vertices. For example, the following correlator can be written suggestively as



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3}\theta^3_{i_1 i_2}\theta_{i_2 i_3}$$

$$= \mathcal{T}_{i_1}\mathcal{T}^{(3)}_{i_1 i_2}\mathcal{T}^{(1)}_{i_2 i_3}\mathcal{T}_{i_3}, \tag{16}$$

where the indices $i_1$, $i_2$, $i_3$ are contracted via Einstein summation notation, and we have defined the particle tensors as

$$\mathcal{T}_i = \sqrt{z_i}, \qquad \mathcal{T}^{(\alpha_{12})}_{i_1 i_2} = \sqrt{z_{i_1} z_{i_2}}\,\theta^{\alpha_{12}}_{i_1 i_2}. \tag{17}$$

Note that this decomposition is not unique, and we could have represented Eq. (16) alternatively as

$$\mathcal{T}_{i_1}\mathcal{T}_{i_2}\mathcal{T}_{i_3}\mathcal{T}^{(3,0,1)}_{i_1 i_2 i_3}, \tag{18}$$

where the three-index particle tensor is

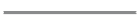
$$\mathcal{T}^{(\alpha_{12},\alpha_{13},\alpha_{23})}_{i_1 i_2 i_3} = \sqrt{z_{i_1} z_{i_2} z_{i_3}}\,\theta^{\alpha_{12}}_{i_1 i_2}\theta^{\alpha_{13}}_{i_1 i_3}\theta^{\alpha_{23}}_{i_2 i_3}. \tag{19}$$

This is an example of an $N$-index particle tensor:

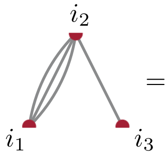
$$\mathcal{T}^{(\{\alpha_{ab}\})}_{i_1\cdots i_N} = \sqrt{\prod_{j=1}^{N} z_{i_j}\prod_{k=1}^{N}\prod_{\ell=k+1}^{N}\theta^{\alpha_{k\ell}}_{i_k i_\ell}}, \tag{20}$$

where $\{\alpha_{ab}\}$ are integers ordered by ascending $b$ then $a$.

The general decomposition, demonstrated by Eq. (16), has a simple graphical interpretation as graphs with cut vertices:

$_i = \sqrt{z_i},$ $\quad j$  $_k = \theta_{jk}.$ (21)

Here, we have sliced open a vertex into two free indices, which later will be contracted to obtain the sums. For example, the alternative decomposition in Eq. (18) involves a graph with three cut vertices:



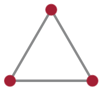
$$= \mathcal{T}^{(3,0,1)}_{i_1 i_2 i_3} = \sqrt{z_{i_1} z_{i_2} z_{i_3}}\,\theta^3_{i_1 i_2}\theta_{i_2 i_3}. \tag{22}$$

The generic $N$-index particle tensor in Eq. (20) can be represented by a multigraph $G$:

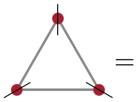
$$\sqrt{z_{i_1}}\cdots\sqrt{z_{i_N}}\prod_{(k,\ell)\in G}\theta_{i_k i_\ell}, \qquad (23)$$

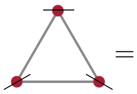
which is similar to the multiparticle correlator in Eq. (8), albeit without the sums over particles and with square roots on the weights $z_i$. We also note that the particle tensors, unlike the multiparticle correlators, depend on the particular vertex labeling of $G$.

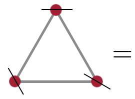
For any multigraph, its vertices can be sliced in half in many different ways, each yielding a valid tensor contraction expression for that correlator. All the different ways of cutting the vertices of the triangle graph yield

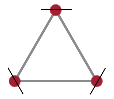


$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1}z_{i_2}z_{i_3}\theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_1 i_3}$$

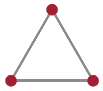
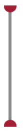
$$=$$  $$= \mathcal{T}^{(1)}_{i_1 i_2}\mathcal{T}^{(1)}_{i_2 i_3}\mathcal{T}^{(1)}_{i_1 i_3}$$

$$=$$  $$= \mathcal{T}_{i_2}\mathcal{T}^{(1,0,1)}_{i_1 i_2 i_3}\mathcal{T}^{(1)}_{i_1 i_3}$$

$$=$$  $$= \mathcal{T}_{i_1}\mathcal{T}_{i_2}\mathcal{T}^{(1,0,1,1,0,0)}_{i_1 i_2 i_3 i_3}$$

$$=$$  $$= \mathcal{T}_{i_1}\mathcal{T}_{i_2}\mathcal{T}_{i_3}\mathcal{T}^{(1,1,1)}_{i_1 i_2 i_3}. \qquad (24)$$

The first contraction in Eq. (24) is the trace of the cube of a two-index tensor treated as a matrix, which we can graphically represent as

 $$= \mathrm{tr}\Bigg($$  $$\Bigg)^3. \qquad (25)$$

## B. Computational complexity

Casting the particle sums in multiparticle correlators as tensor contractions can be used to improve their computation in several ways. First, improvements in the computational complexity of individual correlators can be achieved using algorithms for fast matrix operations. While the triangle graph in Eq. (24) is $\mathcal{O}(M^3)$ to evaluate both naively and with variable elimination, casting it as a matrix product in Eq. (25) allows us to go beyond this limit. Using Strassen's algorithm [31] yields an $\mathcal{O}(M^{2.81})$ evaluation of the correlator, and methods as fast as $\mathcal{O}(M^{2.38})$ exist [32–34]. For instance, Strassen's algorithm uses the structure of the matrix product to reduce the number of operations required to multiply $2\times 2$ block matrices to 7 rather than 8, which gives rise to a power of $\log_2 7 \simeq 2.81$ in its complexity when recursively applied. We will showcase this speedup for computing collider observables in Sec. VI C.

Beyond improving individual correlator evaluation, the constituent particle tensors can be used to improve the computation of collections of correlators. Many correlators can be built out of the same particle tensors contracted in different ways. This allows for a dynamic programming approach where various particle tensors are computed and then reused many times to evaluate additional correlators. While determining the optimal set of subgraphs to compute for a target collection of correlators is beyond the scope of this work, we note that it is in an interesting avenue for further exploration.

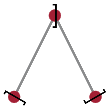
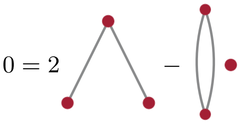
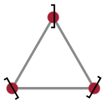
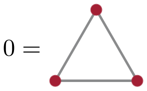
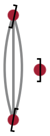
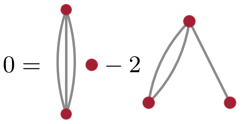
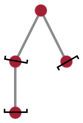
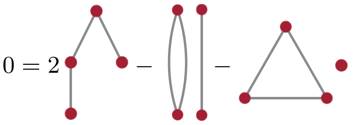
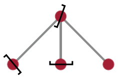
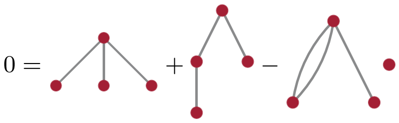
## C. Finite-particle linear relations

When there are small numbers of particles, algebraic relations among the multiparticle correlators emerge due to the simplified nature of the configuration. Relations of this type were also explored in Ref. [1] for the purpose of expanding superstring amplitudes. In the collider context, understanding these relations is also important for calculating EFPs in fixed-order perturbative quantum field theory with small numbers of particles. In general, understanding these finite-particle relations can improve the computation of collections of correlators by avoiding the computation of redundant elements. Further, these relations can involve multigraphs of different computational complexities, allowing for individual correlators to be evaluated more efficiently via these relations.

The tensor identity in Eq. (3) concisely encodes many well-known results, including several Riemann tensor identities as well as the Cayley-Hamilton theorem [35]. The second fundamental theorem of invariant theory indicates that any identity among a set of tensors in $m$ dimensions, aside from ones governed by their existing symmetries, can be obtained as a consequence of Eq. (3) [36].

For the case of particle tensors with $M$ particles, antisymmetrizing over any choice of $L > M$ indices must vanish. We can translate this fact into the language of cutting graphs. Graphically, we denote the antisymmetrization over an index by a bracket through a vertex whose direction indicates which of the two particle tensors to use, and the vertex itself represents contracting the indices. The sum over all permutations of those indices, weighted by the sign of the permutation, must vanish by Eq. (3). After contracting the indices, each permutation gives rise to an associated multigraph, so this procedure gives rise to an alternating sum of multigraphs which must vanish. Some choices of antisymmetrizations will vanish trivially by the symmetries of the tensors, but many do not trivially vanish and give rise to nontrivial algebraic relations.

TABLE III. All linear relations with $M = 2$ particles for $d \leq 3$ multiparticle correlators, derived by antisymmetrizing the particle tensor indices.

| Antisymmetrization | Identity for $M = 2$ |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

To showcase this approach, in Table III we enumerate all of the finite particle linear relations for the case of $M = 2$ and $d \leq 3$. For example, the first row of Table III corresponds to the relation

$$3!\mathcal{T}_{[i_1}\mathcal{T}^{i_1}_{i_2}\mathcal{T}^{i_2}_{i_3]}\mathcal{T}^{i_3} = 2\mathcal{T}_{i_1}\mathcal{T}^{i_1}_{i_2}\mathcal{T}^{i_2}_{i_3}\mathcal{T}^{i_3} - \mathcal{T}^{i_1}_{i_2}\mathcal{T}^{i_2}_{i_1}\mathcal{T}^{i_3}\mathcal{T}_{i_3}, \tag{26}$$

where we have raised and lowered indices to make the summation convention more clear. This procedure can be extended to enumerate the relations for larger numbers of particles.

## IV. CUTTING OPEN EDGES

In this section, we show that the structure of multiparticle correlators can be dramatically simplified if the pairwise distances $\theta_{ij}$ can be written as an inner product. In particular, this allows us to slice open the multigraph edges to obtain moment tensors, resulting in further computational speedups and a greater understanding of various linear relations.

### A. Moments from an inner product

The following results assume that $\theta_{ij}$ can be written as an inner product. Repeating Eq. (4) for convenience,

$$\theta_{ij} = \eta_{\mu\nu} u_i^\mu u_j^\nu, \tag{27}$$

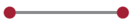
for some vectors $u_i^\mu, u_j^\nu$ and (symmetric) metric $\eta_{\mu\nu}$. This assumption holds for the kinematic polynomials in Eq. (12) and for the EFP angular measure in Eq. (14) for $\beta = 2$. While we use the notation of Lorentz indices and the Minkowski metric here, the conclusions hold for any inner product among vectors using any choice of metric.

With this assumption, any multiparticle correlator can be written via tensor contractions, where in this case the indices are Lorentz indices rather than particle indices. To that end, we introduce the following moment tensors:

$$\mathcal{M}^{\mu_1\cdots\mu_v} = \sum_{i=1}^{M} z_i u_i^{\mu_1}\cdots u_i^{\mu_v}, \tag{28}$$

where $v$ is the rank of the tensor. We apply this moment logic to kinematic polynomials in Sec. V A and to $\beta = 2$ EFPs in Sec. VI A.

A simple example demonstrates the connection between multiparticle correlators and these moment tensors. Consider the following equivalent ways of writing the simplest 2-correlator:



$$\begin{aligned} &= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\eta_{\mu\nu} u_{i_1}^\mu u_{i_2}^\nu) \\ &= \left(\sum_{i_1=1}^{M} z_{i_1} u_{i_1}^\mu\right)\left(\sum_{i_2=1}^{M} z_{i_2} u_{i_2}^\nu\right)\eta_{\mu\nu} \\ &= \mathcal{M}^\mu \mathcal{M}_\mu, \end{aligned} \tag{29}$$

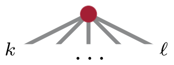
where the first line uses the definition of a multiparticle correlator and Eq. (27), and the last line is explicitly a contraction of moment tensors from Eq. (28) with Einstein summation notation. This is the generalization of the familiar mass identity in Eq. (5), where the inner product allows us to separate the factors into tensors with only a single Lorentz index.

We now generalize this procedure to demonstrate that all multiparticle correlators can be obtained from these moment tensors when the inner product condition holds. The general strategy is to regroup the summand of a general correlator into a product of factors, one for each sum index. The resulting expression manifestly factors the dependence on the constituent particles into tensorial objects contracted according to the multigraph. Substituting Eq. (27) into Eq. (8), we have

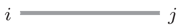
$$\begin{aligned} &\sum_{i_1=1}^{M}\cdots\sum_{i_N=1}^{M} z_{i_1}\cdots z_{i_N}\prod_{(k,\ell)\in G}\eta_{\mu\nu}u_{i_k}^\mu u_{i_\ell}^\nu \\ &\quad = \left(\prod_{j=1}^{N}\sum_{i_j=1}^{M} z_{i_j} u_{i_j}^{\mu_1^j} u_{i_j}^{\mu_2^j}\cdots u_{i_j}^{\mu_{v_j}^j}\right)\prod_{(k,\ell)\in G}\eta_{\mu^k_{A_{k\ell}}\mu^\ell_{A_{\ell k}}} \\ &\quad = \left(\prod_{j=1}^{N}\mathcal{M}^{\mu_1^j\mu_2^j\cdots\mu_{v_j}^j}\right)\prod_{(k,\ell)\in G}\eta_{\mu^k_{A_{k\ell}}\mu^\ell_{A_{\ell k}}}, \end{aligned} \tag{30}$$

where $v_j$ is the valency of vertex $j$, $\mu_b^a$ is an index corresponding to the $b$th edge associated with vertex $a$, and $A$ is a matrix such that the $A_{k\ell}$th instance of $u_{i_k}^\mu$ is contracted with the $A_{\ell k}$th instance of $u_{i_\ell}^\mu$.

Demonstrating their continued usefulness, multigraphs provide a simpler recipe for translating between multiparticle correlators and moment tensors than the unavoidably opaque notation of Eq. (30). Each vertex in the graph is associated with a moment having $v$ indices, where $v$ is the valency of the vertex. The edges in the graph specify which vertices are connected, and hence they encode the contractions between indices of different moments. These rules can be summarized as



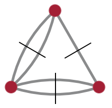
$$= \mathcal{M}^{\mu_k\cdots\mu_\ell} \tag{31}$$



$$= \eta_{\mu_i\mu_j}. \tag{32}$$

Graphically, we have cut the edges of the multigraphs in half to obtain these tensors.

As an example, consider the following multigraph expressed both as a multiparticle correlator and in terms of contractions of moments:



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1}z_{i_2}z_{i_3}\theta_{i_1i_2}\theta_{i_1i_3}^2\theta_{i_2i_3}^2 \tag{33}$$

$$= \mathcal{M}^{\mu\nu\rho\sigma}\mathcal{M}^{\tau}{}_{\mu\nu}\mathcal{M}_{\rho\sigma\tau}, \tag{34}$$

where pairs of indices are contracted with the metric. Interestingly, these graphical rules are closely related to Penrose graphical notation for tensor contractions [37].

### B. Computational complexity

The moments in Eq. (28) are symmetric tensor structures summed over $M$ particles, and hence are $\mathcal{O}(M)$ to compute. Since the tensor structures are fully symmetric, they only have $(n+v-1)!/v!/(n-1)!$ independent components, where $n$ is the dimensionality of the inner product vectors $u_i^\mu$ and $v$ is the rank of the tensor. The number of independent components is therefore polynomial in the rank $v$ as opposed to exponential. In $3+1$ spacetime dimensions, $n=4$ and there are $(v+3)(v+2)(v+1)/6$ independent components. Since this is independent of $M$ and $v^3$ is smaller than the number of particles for many events and multiparticle correlators of interest, computing the moment tensors is typically computationally efficient.

Since any multiparticle correlator with an inner product structure can be written as a contraction of moment tensors via Eq. (30), all such correlators become $\mathcal{O}(M)$ to compute. The tensorial rearrangement into moments provides a way to circumvent the previous computational limits and achieve linear complexity. This can be seen explicitly in Eq. (29), where an $\mathcal{O}(M^2)$ tree graph is written as a contraction of two tensors that are each $\mathcal{O}(M)$ to compute. Since the computation of any correlator must probe all particles at least once, the linear complexity in $M$ of computing the correlators via the moments is optimal. Thus, by computing all of the particle moments up to a desired order $v$, all correlators can be obtained via tensor contractions which are independent of $M$ and depend only on the dimensionality of the space. We provide an explicit demonstration of this computational improvement for collider observables in Sec. VI C.

### C. Finite-dimension linear relations

When the dimensionality of the inner product space is small, there are additional linear relations among the multiparticle correlators. These relations were pointed out by Ref. [3], and we now seek to understand them systematically. Similar to the finite-particle relations in Sec. III C, the key is to figure out the appropriate way to apply Eq. (3) to antisymmetrize over more indices than dimensions.

In this case, we antisymmetrize over the inner product (Lorentz) indices instead of the particle indices. We specialize to the cases of $n=4$ and $n=3$ dimensions for the following discussion, since these are the dimensionalities relevant for Sec. VI and the Appendix, respectively. It is worth noting that $n$-dimensional identities continue to hold in fewer than $n$ dimensions.

The Cayley-Hamilton theorem is a special case of Eq. (3), so we will first explore this as a method to understand the finite-dimension identities. The Cayley-Hamilton theorem, as typically stated, says that a matrix $A$ satisfies its own characteristic polynomial. Using the Newton identities, this translates to a matrix relation between powers of $A$, the trace of powers of $A$, and the determinant of $A$. Written out explicitly for small matrices, we have for 2 × 2 matrices

$$A^2 - (\mathrm{tr}\,A)A + (\det A)I = 0, \tag{35}$$

where $I$ is the identity matrix, and for 3 × 3 matrices

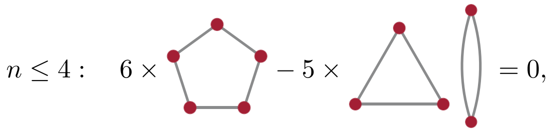
$$A^3 - (\mathrm{tr}\,A)A^2 + \frac{1}{2}((\mathrm{tr}\,A)^2 - \mathrm{tr}\,A^2)A - (\det A)I = 0. \tag{36}$$

For our first $n=4$ identity, we apply the 4 × 4 Cayley-Hamilton theorem to the matrix

$$\sqrt{\eta}^{\mu_1\nu}\mathcal{M}_{\nu\rho}\sqrt{\eta}^{\rho\mu_2}, \tag{37}$$

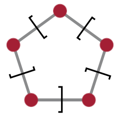
where $\sqrt{\eta}$ is defined because $\eta$ is symmetric.

Multiplying the resulting Cayley-Hamilton relation by the matrix and taking a trace yields



(38)

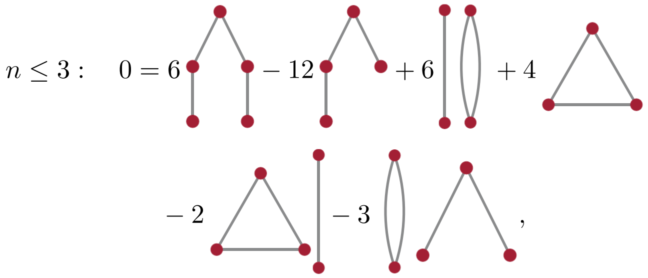
where we assume that $\mathcal{M}^{\mu}_{\mu}=0$, as in our cases of interest due to the masslessness of the particles, with the full expression being easily obtainable but unwieldy. Note that Eq. (38) could also have been obtained directly from the antisymmetrization identity



$$= \mathcal{M}^{\mu_2}_{[\mu_1}\mathcal{M}^{\mu_3}_{\mu_2}\mathcal{M}^{\mu_4}_{\mu_3}\mathcal{M}^{\mu_5}_{\mu_4}\mathcal{M}^{\mu_1}_{\mu_5]} = 0. \tag{39}$$

Higher-dimensional versions of the Cayley-Hamilton theorem can be used to derive relations for larger 2-regular graphs (hexagon, heptagon, etc.).

In addition to identities involving 2-regular graphs, we can also multiply the expression of the Cayley-Hamilton theorem on both sizes by $\sqrt{\eta}^{\mu\nu}\mathcal{M}_{\nu}$ in order to produce relations among chainlike graphs, such as



(40)

which holds in three or fewer dimensions. In general, the Cayley-Hamilton theorem provides a matrix expression valid in some number of dimensions which can be combined with any graph fragment to give an identity. While this approach only yields a subset of the finite-dimension identities, Eq. (3) captures all of them in full generality.

## V. TRIMMING THE LEAVES

When there are simple global constraints on the particles, their effects on multiparticle correlators can often be made clear graphically. In this section, we consider cases where center-of-momentum relations allows us to "trim the leaves" of a multigraph, which gives us a way to count kinematic polynomials from Ref. [1]. More exotic graphical rules are relevant for multiparticle correlators in $e^+e^-$ collisions, which we study in detail in the Appendix.

### A. Revisiting kinematic polynomials

As anticipated in Sec. IV A, kinematic polynomials can be written as contractions of moment tensors. With massless particles, the relevant moment tensors take the following form:

$$\mathcal{P}^{\mu_1\cdots\mu_v} = 2^{v/2}\sum_{i=1}^{M} p_i^{\mu_1}\cdots p_i^{\mu_v}, \tag{41}$$

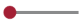
with the prefactor accounting for the factor of two in the Mandelstam invariants $s_{ij}=2p_i^{\mu}p_{j\mu}$. Following the logic of Sec. IV A, all Lorentz-invariant kinematic polynomials can be built from complete Lorentz contractions of these moments.

### B. Center-of-momentum relations

Kinematic polynomials can be further simplified when imposing energy-momentum conservation $\sum_i p_i^{\mu}=0$. Written in the generic multiparticle correlator language, this corresponds to the following constraint:

$$\mathcal{M}^{\mu} = \sum_{i=1}^{M} z_i u_i^{\mu} = 0. \tag{42}$$

Since $\mathcal{M}^{\mu}$ corresponds to a vertex connected to a single edge, we can write this constraint graphically as



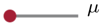
$$= 0. \tag{43}$$

Therefore, momentum conservation allows us to restrict our attention to "leafless" graphs, whose minimum vertex valency is greater than 1.

Leafless graphs are also relevant for studying EFPs on a collection of final-state particles. It is sometimes convenient to work in the center-of-momentum frame with the constraint

$$\mathcal{M}^{\mu} = (E,0,0,0)^{\mu}, \tag{44}$$

where $E$ is a fixed energy in this frame, corresponding to the graphical rule



$$= \quad E\,\delta^{\mu}_{0}. \tag{45}$$

In the $e^+e^-$ case where $u_i^0=1$, valency-1 vertices simply contribute factors of $E$ and we have

$$\mathcal{M}_{\mu_1}\mathcal{M}^{\mu_1\mu_2\cdots\mu_v} = E\mathcal{M}^{\mu_2\cdots\mu_v}. \tag{46}$$

TABLE IV. The number of multigraphs with $d$ edges that have no vertices of valency one (i.e., are leafless), tabulated for both connected and all graphs. These have been added to the OEIS as new sequences [13,14]. We conjecture that the sequence for all graphs is the same as A226919 [2], which was discovered in a string theory context [1]. The values for $d > 8$ (bold font) are new results.

| | Leafless multigraphs | |
|---|---:|---:|
| | Connected | All |
| Edges $d$ | A307317 [13] | A307316 [14] |
| 1 | 0 | 0 |
| 2 | 1 | 1 |
| 3 | 2 | 2 |
| 4 | 4 | 5 |
| 5 | 9 | 11 |
| 6 | 26 | 34 |
| 7 | 68 | 87 |
| 8 | 217 | 279 |
| 9 | **718** | **897** |
| 10 | **2553** | **3129** |
| 11 | **9574** | **11 458** |
| 12 | **38 005** | **44 576** |
| 13 | **157 306** | **181 071** |
| 14 | **679 682** | **770 237** |
| 15 | **3 047 699** | **3 407 332** |
| 16 | **14 150 278** | **15 641 159** |

Hence only leafless multigraphs need to be considered in the center-of-momentum frame for EFPs in the case of $e^+e^-$ collisions.

### C. Counting superstring amplitudes

There is an interesting dividend of our graphical understanding of the center-of-momentum relation in Eq. (43). Reference [1] sought to count the number of independent, symmetric polynomials of kinetic variables of degree $d$, which is relevant for determining five-point superstring amplitudes. They worked in the limit of many particles with many spacetime dimensions (i.e., without finite spacetime dimension identities). Using an interesting technique based on Molien series, they conjectured the counts for $d \leq 8$. In our language, counting these independent polynomials of degree $d$ in the center-of-momentum frame corresponds to simply counting leafless multigraphs with $d$ edges.

We can efficiently enumerate and count the number of leafless multigraphs using NAUTY [38], with our results summarized in Table IV. Our sequence agrees with the results of Ref. [1] where they overlap [2], but we are able to overcome previous computational limitations and double the number of known terms. These values have been added as new sequences in the OEIS [13,14]. Enumerating multigraphs not only allows us to efficiently count these polynomials, but it also provides explicit constructions of them, which may be useful in further exploring the space of kinematic polynomials.

## VI. APPLICATIONS TO COLLIDER PHYSICS

In this section, we apply the general lessons that we have developed to obtain concrete computational and conceptual improvements for collider observables.[1] We first introduce EFMs as novel tensorial structures to efficiently encode and compute $\beta = 2$ EFPs, and we show their relation to previous moment tensors in the literature. We then exploit the results of Sec. IV B to provide a maximally efficient $\mathcal{O}(M)$ implementation of key jet substructure observables.

### A. Introducing energy flow moments

With $\beta = 2$, the pairwise distance measure of the EFPs in Eq. (14) has an inner product structure. Following Sec. IV A, the $\beta = 2$ EFPs can therefore be written as contractions of EFMs defined in Eq. (6), repeated for convenience:

$$\mathcal{I}^{\mu_1 \cdots \mu_v} = 2^{v/2} \sum_{i=1}^{M} E_i n_i^{\mu_1} \cdots n_i^{\mu_v}, \tag{47}$$

where $n_i^\mu \equiv p_i^\mu / E_i$, and the prefactor accounts for the factor of 2 in $\theta_{ij}$. The above definition is suitable for $e^+e^-$ collisions, whereas in a hadron collider context, one would typically replace energy $E_i$ with transverse momentum $p_{Ti}$ in the definitions of both $\mathcal{I}^{\mu_1 \cdots \mu_v}$ and $n_i^\mu$.

The fact that the $\beta = 2$ EFPs are IRC safe is immediately clear via inspection of Eq. (47). Due to the linear energy weighting, an EFM is manifestly invariant to the addition of a zero-energy particle or the collinear splitting of one particle into two. To form the $\beta = 2$ EFPs, we need only Lorentz contract the EFMs, which respects this IRC-safe structure.

As an aside, parity violation is interesting to explore at colliders [40–48]. By invoking the $\epsilon$ symbol, we can construct parity-violating contractions of the EFMs that go beyond the parity-invariant EFPs. Two simple parity-violating observables of potential interest are

$$\epsilon_{\alpha\beta\gamma\delta} \mathcal{I}^{\alpha} \mathcal{I}^{\beta\rho} \mathcal{I}^{\gamma\sigma\tau} \mathcal{I}^{\delta}_{\rho\sigma\tau}, \qquad \epsilon_{\alpha\beta\gamma\delta} \mathcal{I}^{\alpha\rho} \mathcal{I}^{\beta\sigma}_{\rho} \mathcal{I}^{\gamma\tau}_{\sigma} \mathcal{I}^{\delta}_{\tau}, \tag{48}$$

which are manifestly permutation symmetric, rotationally invariant, and IRC safe. Further investigations in this direction are left to future work and we restrict ourselves here to contractions with the Minkowski metric.

### B. Relation to existing moments

The EFMs are closely related to existing moment-based approaches for collider physics, dating back several decades. Here, we show that EFMs directly encompass or suitably generalize these approaches.

---

[1] This section also explains two cryptic comments we made in previous papers: footnote 8 of Ref. [9] and footnote 4 of Ref. [39].

A number of tensorial quantities similar in spirit to the EFMs have been previously defined and explored for both $e^+e^-$ and hadronic collisions. In Ref. [49], the following objects built out of particle three-momenta were defined for $e^+e^-$ collisions:

$$\Theta^{j_1 j_2 \cdots j_v} = \sum_{i=1}^{M} E_i \hat{n}_i^{j_1} \hat{n}_i^{j_2} \cdots \hat{n}_i^{j_v}, \tag{49}$$

where $\hat{n}_i = \vec{p}_i / E_i$ is a unit three-vector and $j_k \in \{1,2,3\}$. We refer to the tensors defined by Eq. (49) as “generalized sphericity tensors,” since the second-rank tensor $\Theta_2 = \Theta^{j_1 j_2}$ is often called the sphericity tensor. In an $e^+e^-$ context, the spatial components of the EFMs are exactly these tensors:

$$\mathcal{I}^{j_1 j_2 \cdots j_v} = 2^{v/2} \Theta^{j_1 j_2 \cdots j_v}, \tag{50}$$

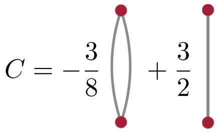
and, like the EFMs, the generalized sphericity tensors are IRC safe.

The second-rank sphericity tensor $\Theta_2 = \Theta^{j_1 j_2}$ has seen significant application in the form of defining observables from its eigenvalues. The $C$ and $D$ parameter event shapes are defined in terms of the eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$ of the sphericity tensor $\Theta_2$ as

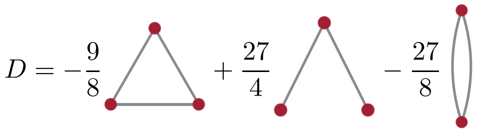
$$C = 3(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3), \qquad D = 27\lambda_1\lambda_2\lambda_3. \tag{51}$$

These IRC-safe observables have been widely analyzed for $e^+e^-$ collisions both theoretically [49–53] and experimentally [54–58].

To highlight the close relationship with EFMs, we note that the $C$ and $D$ parameters are themselves both linear combinations of EFPs and contractions of EFMs:



$$= -\frac{3}{8}\mathcal{I}^{\mu\nu}\mathcal{I}_{\mu\nu} + \frac{3}{2}\mathcal{I}^{\mu}\mathcal{I}_{\mu}, \tag{52}$$



$$= -\frac{9}{8}\mathcal{I}^{\mu\nu}\mathcal{I}_{\nu\rho}\mathcal{I}^{\rho}_{\mu} + \frac{27}{4}\mathcal{I}^{\mu}\mathcal{I}_{\mu\nu}\mathcal{I}^{\nu} - \frac{27}{8}\mathcal{I}^{\mu\nu}\mathcal{I}_{\mu\nu}\mathcal{I}. \tag{53}$$

We derive these expressions in the Appendix. In this graphical notation, we leave factors of total energy implicit:

$$\bullet = \mathcal{I} = \sum_i E_i, \tag{54}$$

which must be included whenever graphs with different numbers of vertices are added together.

The third-rank tensor $\Theta^{j_1 j_2 j_3}$ has seen additional study [59], but the general formulation of Eq. (49) has seen very limited application at the LHC. One reason for this may be that is not clear how to use the tensors in Eq. (49) in a hadronic context: the factor of energy can be promoted to transverse momentum but the explicit appearance of the three-momenta prevents covariance under boosts and rotations about the collision axis. The EFMs are precisely the generalization of the generalized sphericity tensors that applies in both the $e^+e^-$ and hadronic contexts. Including the energy component of the particle four-momenta avails a hadronic interpretation simply by replacing all instances of energy with transverse momentum. Relations among these $\Theta^{j_1 j_2 \cdots j_v}$ tensions are studied further in the Appendix.

Other constructions of moment tensors have been proposed solely in the hadronic context. In Ref. [60], a set of transverse-momentum-weighted moments in the rapidity-azimuth $(y, \phi)$ plane were defined for narrow jets, and a number of substructure observables were identified in terms of their products and contractions. Letting $x_i = (y_i, \phi_i)$ indicate the rapidity-azimuth position of particle $i$, and letting $k_j \in \{1,2\}$ be rapidity-azimuth indices, these hadronic moment tensors are

$$I^{k_1 k_2 \cdots k_v} = \sum_{i=1}^{M} p_{T,i} x_i^{k_1} x_i^{k_2} \cdots x_i^{k_v}. \tag{55}$$

The two-index moment $I^{k_1 k_2}$ has been used to define the planar flow observable [61,62], which quantifies the linear versus planar nature of the radiation pattern.

When boosting a jet to be central ($y = 0$) and taking the narrow limit $y_i, \phi_i \ll 1$, $x_i$ become precisely two spatial indices of the EFMs with the hadronic prescription of $z_i = p_{T,i}$. The tensors of Eq. (55) were defined in Ref. [60] relative to a jet axis, which we neglect here (by boosting to $y = 0$) to make the relation to EFMs immediate. Choosing a $p_T$-weighted centroid axis, a related analysis carries through with a jet axis included. In this way, the EFMs generalize $I^{k_1 \cdots k_n}$ beyond the narrow jet limit and do not require referencing a jet axis.

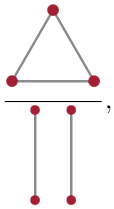
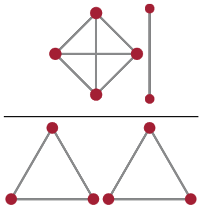
Thus, we see that the EFMs can be related to moments developed for both $e^+e^-$ and hadronic collisions. Unlike these previous moments, which were developed for a specific type of collision, EFMs can operate in both realms by simply exchanging energies for transverse momenta. In this way, EFMs are a general framework for moment-based approaches to collider physics analyses.

### C. Speeding up multiprong taggers

Using the results of Sec. IV B, EFPs with $\beta = 2$ are manifestly $\mathcal{O}(M)$ to compute through the EFMs. This provides a significant computational speedup over approaches based on variable elimination; see Sec. II D.

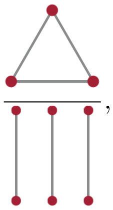
This computational speedup also applies to substructure observables derived from $\beta = 2$ EFPs. A number of different observables have been proposed to tag jets with multiprong hadronic substructure at the LHC, including two-pronged decays of boosted $W$ bosons and three-pronged decays of boosted top quarks, along with other more exotic scenarios [63,64]. These jet substructure observables have since found significant experimental application at the LHC for multiprong tagging and new physics searches [65]. While most current LHC analyses focus on $\beta = 1$ observables, comparable (or better) performance can be obtaining using $\beta = 2$ for standard tagging applications [7,8].

As concrete examples, Ref. [7] established a family of $N$-prong tagging observables $C_N^{(\beta)}$, which can be written as ratios of EFPs with complete graphs [9]. For two- and three-prong tagging, these dimensionless ratio observables are

$C_2^{(\beta)} =$ , $C_3^{(\beta)} =$ , (56)

where factors of $\mathcal{I} =$ ● can be restored by ensuring equal numbers of vertices in the numerator and denominator. Note that these observables are traditionally defined using the rapidity-azimuth distance rather than Eq. (14), though these are equivalent for narrow jets and our $\mathcal{O}(M)$ moment logic can be applied in both cases.

These $C_N$ observables are in general $\mathcal{O}(M^{N+1})$ to compute, which becomes quickly computationally intractable with increasing $N$. Using power counting arguments, Ref. [8] established a different combination of energy correlators as an improved two-pronged tagger called $D_2^{(\beta)}$, given by the expression

$D_2^{(\beta)} =$ , (57)

which naively requires $\mathcal{O}(M^3)$ to compute. These observables can be measured before or after grooming [66–70] is applied, as they operate on generic sets of particles. Note that the jet substructure observable $N_2$ cannot be expressed as an exact combination of EFPs, since the angular structure of the generalized correlators in Ref. [24] does not take the form of Eq. (2).

In the case of $\beta = 2$, these tagging observables can all be computed in $\mathcal{O}(M)$ using our moment-based approach. Further, $C_2$ and $D_2$ benefit (for any $\beta$) from the fast matrix multiplication speedups to $\mathcal{O}(M^{2.81})$ discussed in Sec. III B.

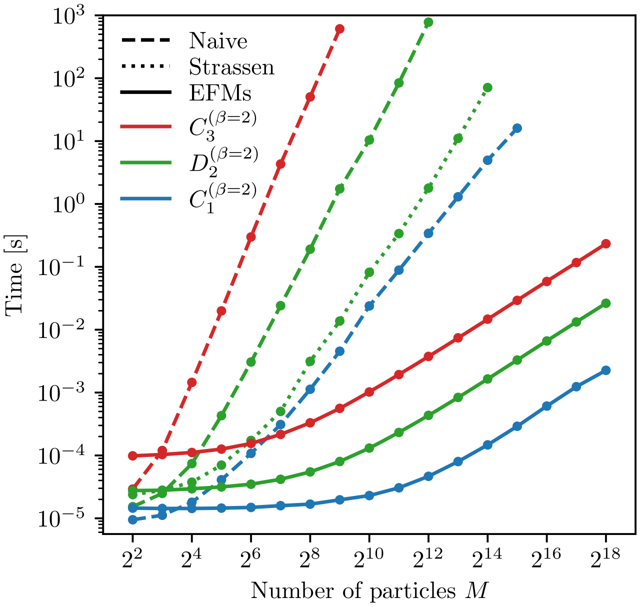
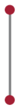


FIG. 1. Compute time (in seconds) for three jet substructure observables: $C_3^{(\beta=2)}$ [naively $\mathcal{O}(M^4)$], $D_2^{(\beta=2)}$ [naively $\mathcal{O}(M^3)$], and $C_1^{(\beta=2)}$ [naively $\mathcal{O}(M^2)$] on an Intel Xeon 2.0 GHz processor. The observables are computed for different numbers of particles by (dashed lines) evaluating the nested sums and (solid lines) using the moment-based approach developed here. The fast matrix multiplication approach using Strassen's algorithm (dotted line) is also used to compute $D_2$. The moment strategy shows the expected linear scaling across all observables.

In Fig. 1, we compare the overall computational time of $C_3^{(\beta=2)}$ and $D_2^{(\beta=2)}$ using the naive approach and the moment-based approach. We also show the fast matrix multiplication approach for computing $D_2$ using Basic Linear Algebra Subprograms [71] via NUMPY [72,73] on eight CPU cores. (The use of multiple cores merely shifts the curve vertically and does not alter the asymptotic scaling.) For comparison, we include the computational time of a single 2-correlator (relevant for one-prong quark/gluon discrimination [8]):

$C_1^{(\beta)} =$ , (58)

which has a naive scaling of $\mathcal{O}(M^2)$.

We can see the significant benefit of linear computational complexity, resulting in very large practical speedups for realistic numbers of particles at the LHC (i.e., 100–1000). Furthermore, this opens the door to the use of higher-$N$ correlators for collider physics, which until now have appeared computationally prohibitive. The $\mathcal{O}(M)$ implementations of all these observables are available in the ENERGYFLOW PYTHON package [15].

## VII. CONCLUSIONS

A greater understanding of multiparticle correlators has profound implications for fields ranging from physics to applied mathematics. In this paper, we have investigated multiparticle correlators, attempting to understand and categorize linear relations that appear among them as well as developing faster computational techniques to improve their practical application.

When the summand of a multiparticle correlator satisfies the monomial form in Eq. (2), the correlator has a multigraph structure that enables us to effectively represent and manipulate it. This was used already in Ref. [9] to apply the variable elimination algorithm to speed up the computation of many multiparticle correlators for collider physics beyond the naive expectation. Here, we have gone significantly further by developing techniques to "cut" the multigraph vertices and edges.

By slicing open the vertices of correlator graphs, we have arrived at particle tensors and demonstrated that each correlator can be written as a contraction of these tensors. The fundamental tensor antisymmetrization identity in Eq. (3) resulted in an infinite class of linear relations that hold when the number of particles $M$ is small compared to the number of vertices $N$. The existence of faster-than-expected tensor operations, such as fast matrix multiplication, has allowed us to reduce the computational complexity of correlators in general. These computational and conceptual strategies extend beyond correlators with monomial summands, applying also to noninteger exponents or indeed products of any pairwise interparticle quantities.

We have also sliced open the edges of the multigraphs when the pairwise distances had the inner product structure of Eq. (4). This assumption holds for kinematic polynomials as well as for EFPs with $\beta = 2$ in both the $e^+e^-$ and hadron collider contexts. Cutting the edges has resulted in moment tensors that carry the indices of the inner product space. Applying the tensor antisymmetrization identity has led to another infinite class of linear relations arising when the dimensionality of the inner product space (e.g., four spacetime dimensions) is small compared to the number of edges $d$. We have also made contact with efforts to analytically enumerate structures appearing in superstring amplitudes, deriving a method that has allowed us to significantly extend previous results.

In a collider context, the moment tensors yield the EFMs. The EFMs are novel tensorial structures that provide a natural and efficient way to compute collider correlators, unifying and encompassing existing moment-based approaches in the collider physics literature. We have demonstrated a significant speedup of widely used jet substructure observables using the graphical techniques that we have developed. We expect that the EFMs will be a useful development for experimental applications and theoretical calculations of multiparticle correlators in collider physics.

## ACKNOWLEDGMENTS

We are grateful to Brian Henning, Andrew Larkoski, Thomas Melia, and Ian Moult for helpful conversations. We thank the Harvard Center for the Fundamental Laws of Nature for its hospitality while this work was completed. This work was supported by the Office of Nuclear Physics of the U.S. Department of Energy (DOE) under Grant No. DE-SC-0011090 and by the DOE Office of High Energy Physics under Grants No. DE-SC0012567 and No. DE-SC0019128. J. T. is supported by the Simons Foundation through a Simons Fellowship in Theoretical Physics. Cloud computing resources were provided through a Google Cloud allotment from the MIT Quest for Intelligence.

## APPENDIX: EUCLIDEAN SLICING LINEAR RELATIONS

In this appendix, we derive linear relations for the generalized sphericity tensors in Eq. (49), and show how they can be used to constrain $\beta = 2$ EFPs in $e^+e^-$ collisions. The generalized sphericity tensors appear in the spatial parts of EFMs. For example, the rank-2 tensors are related by

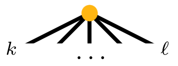
$$\mathcal{I}^{00} = 2, \quad \mathcal{I}^{0i} = \mathcal{I}^{i0} = \sqrt{2}\mathcal{I}^{i} = 2\Theta^{i}, \quad \mathcal{I}^{ij} = 2\Theta^{ij}. \tag{A1}$$

For massless particles in $e^+e^-$ collisions, the EFP measure in Eq. (14) has the property that all $n_i^\mu$ take the special form $n_i^\mu = (1, \hat{n})^\mu$. The appearance of 1 in the zeroth component is the reason why relations such as $\mathcal{I}^{i0} = \sqrt{2}\mathcal{I}^{i}$ exist. In general, we have "subslicing" relations of the form

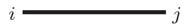
$$\mathcal{I}^{0\mu_1\cdots\mu_v} = \sqrt{2}\mathcal{I}^{\mu_1\cdots\mu_v}. \tag{A2}$$

Since the $\Theta$ tensors live in three dimensions, they satisfy finite-dimensional tensor identities involving antisymmetrization over four or more indices, i.e., one fewer than the four-dimensional EFMs. Since any EFM for $e^+e^-$ collisions can be written in terms of the $\Theta$ tensors, these redundancies carry over and manifest as new EFM relations. We showcase several specific examples of these identities in this appendix.

To begin, we introduce a graphical notation to keep track of the different contractions of the sphericity tensors, in analogy with the graphs for EFMs. Explicitly, the graphical rules are

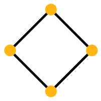


$$= \Theta^{a_k\cdots a_\ell}, \tag{A3}$$



$$= 2\,\delta_{a_i a_j}. \tag{A4}$$

For instance, the square graph has the following meaning in terms of the sphericity tensor:



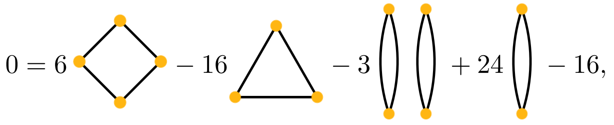
$$= 2^4\,\Theta^{ab}\Theta_{bc}\Theta^{cd}\Theta_{da} = 2^4\,\mathrm{Tr}[\Theta_2^4]. \qquad \text{(A5)}$$

Note that $\mathrm{Tr}[\Theta_2] = 1$, and the factors of 2 come from the spatial part of our $2\delta_{ab}$ convention. In general, the traces of powers of $\Theta_2$ give rise to 2-regular graphs.

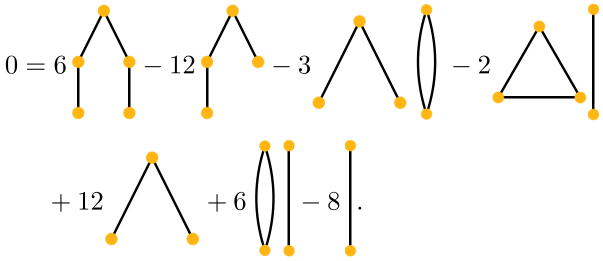
Next, we apply the $3 \times 3$ Cayley-Hamilton theorem from Eq. (36) to $\Theta_2$. Multiplying Eq. (36) through by $\Theta_2$, we have

$$\Theta_2^4 - \Theta_2^3 + \frac{1}{2}(1 - \mathrm{tr}\,\Theta_2^2)\Theta_2^2 - \Theta_2 \det \Theta_2 = 0. \qquad \text{(A6)}$$

Substituting in for the det $\Theta_2$ [from tracing over Eq. (36)] and taking the trace of this expression gives rise to one identity. Multiplying Eq. (36) on both sides by $\Theta_1$ gives rise to a separate identity. These two identities are



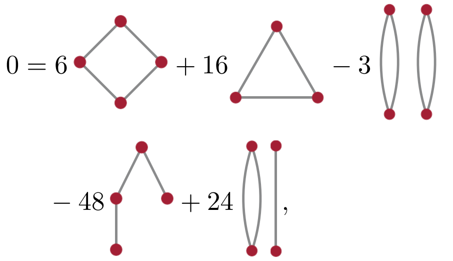
(A7)



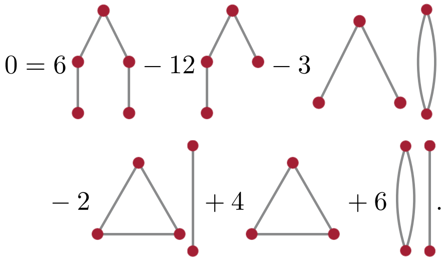
(A8)

In this analysis, we used a Cayley-Hamilton approach for simplicity. In general, these graphical identities can be systematically derived from the tensor antisymmetrization identity in Eq. (3) by enumerating the possible antisymmetrizations (here, over four or more indices) of each graph structure.

To translate these identities among sphericity tensors into identities among EFPs, we develop a look-up table which writes out each contraction of sphericity tensors as a specific linear combination of EFPs. This can be worked out in the moment picture by recursively writing a Lorentzian contraction as a Euclidean contraction plus an additional term of opposite sign (with the appropriate factors of 2). Our results are summarized in Table V for all graphs up through the connected graphs with four edges. With this table, we can quickly translate between the sphericity and energy flow pictures. As an example, the identities in Eqs. (A7) and (A8) give rise to the following identities among the EFPs:



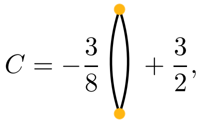
(A9)



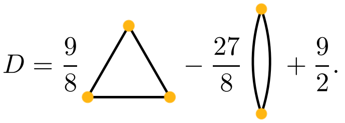
(A10)

Hence the finite dimension identities of the three-dimensional sphericity tensor induce related identities among the EFPs, with an interesting interplay between dimensionalities.

We can use similar reasoning to make contact with the $C$ and $D$ parameters. From Eq. (51), these event shapes are related to the trace and determinant of the sphericity tensor via $C = \frac{3}{2}((\mathrm{tr}\,\Theta)^2 - \mathrm{tr}\,\Theta^2)$ and $D = 27 \det \Theta$. Using Eq. (36) to express the determinant of $\Theta$, we can write these observables in terms of the sphericity graphs in the following way:



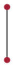
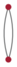
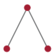
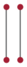
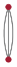
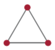
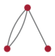
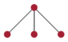
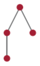
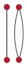
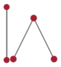
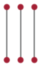
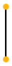
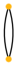
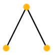
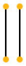
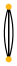
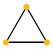
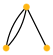
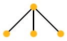
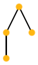
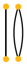
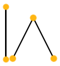
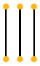
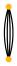
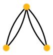
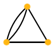
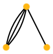
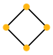
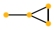
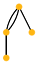
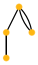
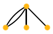
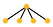
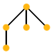
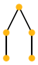
(A11)



(A12)

We can use Table V to translate Eqs. (A11) and (A12) into the energy flow picture. Doing so yields the EFP expressions for the $C$ and $D$ parameters already presented in Eqs. (52) and (53).

TABLE V. All contractions of generalized sphericity tensors (left column) as linear combinations of EFPs (top row). The entries of the table are the specific linear coefficients of each EFP for a particular sphericity contraction. This table can be used to translate between linear combinations of sphericity contractions and EFPs by matrix multiplication. Shown are all graphs up to the connected graphs with five edges, though in principle it can be extended arbitrarily. Note that a single diagonal of ones continues for the $d = 4$ graphs and is not shown explicitly.

| | $d = 0$ | $d = 1$ | $d = 2$ | | | $d = 3$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | |  |  |  |  |  |  |  |  |  |  |  |  | ... |
| | 1 | | | | | | | | | | | | | |
|  | 2 | −1 | | | | | | | | | | | | |
|  | 4 | −4 | 1 | | | | | | | | | | | |
|  | 4 | −4 | | 1 | | | | | | | | | | |
|  | 4 | −4 | | | 1 | | | | | | | | | |
|  | 8 | −12 | 6 | | | −1 | | | | | | | | |
|  | 8 | −12 | | 6 | | | −1 | | | | | | | |
|  | 8 | −12 | 2 | 4 | | | | −1 | | | | | | |
|  | 8 | −12 | | 6 | | | | | −1 | | | | | |
|  | 8 | −12 | | 4 | 2 | | | | | −1 | | | | |
|  | 8 | −12 | 2 | | 4 | | | | | | −1 | | | |
|  | 8 | −12 | | 2 | 4 | | | | | | | −1 | | |
|  | 8 | −12 | | | 6 | | | | | | | | −1 | |
|  | 16 | −32 | 24 | | | −8 | | | | | | | | 1 |
|  | 16 | −32 | 8 | 16 | | | | −8 | | | | | | |
|  | 16 | −32 | 4 | 20 | | | −4 | −4 | | | | | | |
|  | 16 | −32 | 12 | 12 | | −2 | | −6 | | | | | | |
|  | 16 | −32 | | 16 | 8 | | | | | −8 | | | | |
|  | 16 | −32 | | 20 | 4 | | −2 | | −2 | −4 | | | | |
|  | 16 | −32 | 4 | 16 | 4 | | | −4 | | −4 | | | | |
|  | 16 | −32 | 4 | 12 | 8 | | | −2 | | −4 | −2 | | | |
|  | 16 | −32 | 4 | 20 | | | | −4 | −4 | | | | | |
|  | 16 | −32 | | 24 | | | | | −8 | | | | | |
|  | 16 | −32 | | 16 | 8 | | | | −2 | −4 | | −2 | | |
|  | 16 | −32 | | 12 | 12 | | | | | −4 | | −4 | | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ |

[1] R. H. Boels, On the field theory expansion of superstring five point amplitudes, Nucl. Phys. **B876**, 215 (2013).

[2] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, https://oeis.org/A226919 (2013); asymptotic number of completely symmetric polynomials of degree $n$ up to momentum conservation in the limit as the number of particles increases.

[3] M. Hogervorst, S. Rychkov, and B. C. van Rees, Truncated conformal space approach in $d$ dimensions: A cheap alternative to lattice field theory?, Phys. Rev. D **91**, 025005 (2015).

[4] B. Henning, X. Lu, T. Melia, and H. Murayama, Operator bases, $S$-matrices, and their partition functions, J. High Energy Phys. 10 (2017) 199.

[5] P. Dallot, P. D. Bristowe, and M. Demazure, Reduced coordinates on the configuration space of three and four atoms, Phys. Rev. B **46**, 2133 (1992).

[6] Z. Xie and J. M. Bowman, Permutationally invariant polynomial basis for molecular energy surface fitting via monomial symmetrization, J. Chem. Theory Comput. **6**, 26 (2010).

[7] A. J. Larkoski, G. P. Salam, and J. Thaler, Energy correlation functions for jet substructure, J. High Energy Phys. 06 (2013) 108.

[8] A. J. Larkoski, I. Moult, and D. Neill, Power counting to better jet observables, J. High Energy Phys. 12 (2014) 009.

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

[10] F. V. Tkachov, Measuring multi-jet structure of hadronic energy flow, or What is a jet?, Int. J. Mod. Phys. A **12**, 5411 (1997).

[11] N. L. Zhang and D. Poole, Exploiting causal independence in Bayesian network inference, J. Artif. Intell. Res. **5**, 301 (1996).

[12] A. Y. Wei, P. Naik, A. W. Harrow, and J. Thaler, Quantum algorithms for jet clustering, arXiv:1908.08949.

[13] E. M. Metodiev, The On-Line Encyclopedia of Integer Sequences, https://oeis.org/A307317 (2019); number of connected leafless multigraphs with $n$ edges.

[14] P. T. Komiske, The On-Line Encyclopedia of Integer Sequences, https://oeis.org/A307316 (2019); number of leafless multigraphs with $n$ edges.

[15] P. T. Komiske and E. M. Metodiev, ENERGYFLOW, https://energyflow.network/ (2019).

[16] N. J. A. Sloane, The On-line encyclopedia of integer sequences, Not. Am. Math. Soc. **50**, 912 (2003).

[17] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, https://oeis.org/A076864 (2002); number of connected multigraphs with $n$ edges.

[18] V. Jovovic, The On-Line Encyclopedia of Integer Sequences, https://oeis.org/A050535 (1999); number of multigraphs on infinite set of nodes with $n$ edges.

[19] G. Sterman, J. Smith, J. C. Collins, J. Whitmore, R. Brock, J. Huston, J. Pumplin, W.-K. Tung, H. Weerts, C.-P. Yuan *et al.*, Handbook of perturbative QCD, Rev. Mod. Phys. **67**, 157 (1995).

[20] S. Weinberg, *The Quantum Theory of Fields. Volume I: Foundations* (Cambridge University Press, Cambridge, England, 2005).

[21] T. Kinoshita, Mass singularities of Feynman amplitudes, J. Math. Phys. (N.Y.) **3**, 650 (1962).

[22] T. D. Lee and M. Nauenberg, Degenerate systems and mass singularities, Phys. Rev. **133**, B1549 (1964).

[23] C. Frye, H. Hannesdottir, N. Paul, M. D. Schwartz, and K. Yan, Infrared finiteness and forward scattering, Phys. Rev. D **99**, 056015 (2019).

[24] I. Moult, L. Necib, and J. Thaler, New angles on energy correlation functions, J. High Energy Phys. 12 (2016) 153.

[25] C. F. Berger, T. Kucs, and G. F. Sterman, Event shape/energy flow correlations, Phys. Rev. D **68**, 014012 (2003).

[26] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, and J. Virzi, Substructure of high-$p_T$ Jets at the LHC, Phys. Rev. D **79**, 074017 (2009).

[27] S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig, and C. Lee, Jet shapes and jet algorithms in SCET, J. High Energy Phys. 11 (2010) 101.

[28] A. J. Larkoski, D. Neill, and J. Thaler, Jet shapes with the broadening axis, J. High Energy Phys. 04 (2014) 017.

[29] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, Gaining (Mutual) Information about Quark/Gluon Discrimination, J. High Energy Phys. 11 (**2014**) 129.

[30] A. Butter *et al.*, The machine learning landscape of top taggers, SciPost Phys. **7**, 014 (2019).

[31] V. Strassen, Gaussian elimination is not optimal, Numer. Math. **13**, 354 (1969).

[32] D. Coppersmith and S. Winograd, Matrix multiplication via arithmetic progressions, J. Symb. Comput. **9**, 251 (1990).

[33] A. M. Davie and A. J. Stothers, Improved bound for complexity of matrix multiplication, Proc. R. Soc. Edinburgh, Sect. A **143**, 351 (2013).

[34] F. L. Gall, Powers of tensors and fast matrix multiplication, in *Proceedings of the International Symposium on Symbolic and Algebraic Computation (ISSAC '14), Kobe, Japan, 2014*, edited by K. Nabeshima (ACM, New York, 2014), pp. 296–303.

[35] S. B. Edgar and A. Hoglund, Dimensionally dependent tensor identities by double antisymmetrization, J. Math. Phys. (N.Y.) **43**, 659 (2002).

[36] G. Sneddon, The identities of the algebraic invariants of the four-dimensional Riemann tensor. II, J. Math. Phys. (N.Y.) **39**, 1659 (1998).

[37] R. Penrose, Applications of negative dimensional tensors, Comb. Math. Appl. **1**, 221 (1971).

[38] B. D. McKay and A. Piperno, Practical graph isomorphism, II, J. Symb. Comput. **60**, 94 (2014).

[39] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow networks: Deep sets for particle jets, J. High Energy Phys. 01 (2019) 121.

[40] O. Nachtmann, A new tool for the study of fundamental interactions: Parity odd correlations in quark fragmentation, Nucl. Phys. **B127**, 314 (1977).

[41] R. H. Dalitz, G. R. Goldstein, and R. Marshall, On the helicity of charm jets, Z. Phys. C **42**, 441 (1989).

[42] A. V. Efremov, L. Mankiewicz, and N. A. Tornqvist, Jet handedness as a measure of quark and gluon polarization, Phys. Lett. B **284**, 394 (1992).

[43] A. Brandenburg, L. J. Dixon, and Y. Shadmi, Event handedness in $e^+e^-$ annihilation to three jets, Phys. Rev. D **53**, 1264 (1996).

[44] S. Chatrchyan *et al.* (CMS Collaboration), Forward-backward asymmetry of Drell-Yan lepton pairs in $pp$ collisions at $\sqrt{s} = 7$ TeV, Phys. Lett. B **718**, 752 (2013).

[45] G. Aad *et al.* (ATLAS Collaboration), Measurement of the forward-backward asymmetry of electron and muon pair-production in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, J. High Energy Phys. 09 (2015) 049.

[46] V. Khachatryan *et al.* (CMS Collaboration), Forward-backward asymmetry of Drell-Yan lepton pairs in $pp$ collisions at $\sqrt{s} = 8$ TeV, Eur. Phys. J. C **76**, 325 (2016).

[47] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the weak mixing angle using the forward-backward asymmetry of Drell-Yan events in $pp$ collisions at 8 TeV, Eur. Phys. J. C **78**, 701 (2018).

[48] C. G. Lester and M. Schott, Search for non-standard sources of parity violation in jets at $\sqrt{s} = 8$ TeV with CMS open data, J. High Energy Phys. 12 (2019) 120.

[49] J. F. Donoghue, F. E. Low, and S.-Y. Pi, Tensor analysis of hadronic jets in quantum chromodynamics, Phys. Rev. D **20**, 2759 (1979).

[50] G. Parisi, Super inclusive cross-sections, Phys. Lett. **74B**, 65 (1978).

[51] R. Keith Ellis, D. A. Ross, and A. E. Terrano, The perturbative calculation of jet structure in $e^+e^-$ annihilation, Nucl. Phys. **B178**, 421 (1981).

[52] S. Catani and B. R. Webber, Resummed C-parameter distribution in $e^+e^-$ annihilation, Phys. Lett. B **427**, 377 (1998).

[53] A. J. Larkoski and A. Procita, New insights on an old problem: Resummation of the $D$-parameter, J. High Energy Phys. 02 (2019) 104.

[54] K. Ackerstaff *et al.* (OPAL Collaboration), QCD studies with $e^+e^-$ annihilation data at 161 GeV, Z. Phys. C **75**, 193 (1997).

[55] P. Achard *et al.* (L3 Collaboration), Determination of $\alpha_s$ from hadronic event shapes in $e^+e^-$ annihilation at $192 \leq s \leq 208$ GeV, Phys. Lett. B **536**, 217 (2002).

[56] A. Heister *et al.* (ALEPH Collaboration), Studies of QCD at $\mathrm{e}^+\mathrm{e}^-$ centre-of-mass energies between 91 and 209 GeV, Eur. Phys. J. C **35**, 457 (2004).

[57] J. Abdallah *et al.* (DELPHI Collaboration), The measurement of $\alpha_s$ from event shapes with the DELPHI detector at the highest LEP energies, Eur. Phys. J. C **37**, 1 (2004).

[58] S. Bethke, S. Kluth, C. Pahl, and J. Schieck (JADE Collaboration), Determination of the strong coupling $\alpha_\mathrm{S}$ from hadronic event shapes with $\mathcal{O}(\alpha_\mathrm{S}{}^3)$ and resummed QCD predictions using JADE data, Eur. Phys. J. C **64**, 351 (2009).

[59] S. Dagan and A. Vainshtein, Momentum tensors of second and third rank, as tools for jets' analysis, arXiv:1102.4053.

[60] G. Gur-Ari, M. Papucci, and G. Perez, Classification of energy flow observables in narrow jets, arXiv:1101.2905.

[61] J. Thaler and L.-T. Wang, Strategies to identify boosted tops, J. High Energy Phys. 07 (2008) 092.

[62] L. G. Almeida, S. J. Lee, G. Perez, I. Sung, and J. Virzi, Top jets at the LHC, Phys. Rev. D **79**, 074012 (2009).

[63] 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, arXiv:1709.04464.

[64] S. Marzani, G. Soyez, and M. Spannowsky, Looking inside jets: An introduction to jet substructure and boosted-object phenomenology, Lect. Notes Phys. **958**, 1 (2019).

[65] R. Kogler *et al.*, Jet substructure at the Large Hadron Collider: Experimental review, Rev. Mod. Phys. **91**, 045003 (2019).

[66] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, J. High Energy Phys. 02 (2010) 084.

[67] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches, Phys. Rev. D **81**, 094023 (2010).

[68] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Techniques for improved heavy particle searches with jet substructure, Phys. Rev. D **80**, 051501 (2009).

[69] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, Towards an understanding of jet substructure, J. High Energy Phys. 09 (2013) 029.

[70] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, J. High Energy Phys. 05 (2014) 146.

[71] L. S. Blackford, A. Petitet, R. Pozo, K. Remington, R. C. Whaley, J. Demmel, J. Dongarra, I. Duff, S. Hammarling, G. Henry *et al.*, An updated set of Basic Linear Algebra Subprograms (BLAS), ACM Trans. Math. Softw. **28**, 135 (2002).

[72] T. Oliphant, Guide to NUMPY, https://web.mit.edu/dvp/Public/numpybook.pdf (2006).

[73] Computer code NUMPY, https://www.numpy.org/.