Published for SISSA by Springer

Received: *January 4, 2018*
Accepted: *March 29, 2018*
Published: *April 4, 2018*

Energy flow polynomials: a complete linear basis for jet substructure

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

Center for Theoretical Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.

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

Abstract: We introduce the energy flow polynomials: a complete set of jet substructure observables which form a discrete linear basis for all infrared- and collinear-safe observables. Energy flow polynomials are multiparticle energy correlators with specific angular structures that are a direct consequence of infrared and collinear safety. We establish a powerful graph-theoretic representation of the energy flow polynomials which allows us to design efficient algorithms for their computation. Many common jet observables are exact linear combinations of energy flow polynomials, and we demonstrate the linear spanning nature of the energy flow basis by performing regression for several common jet observables. Using linear classification with energy flow polynomials, we achieve excellent performance on three representative jet tagging problems: quark/gluon discrimination, boosted W tagging, and boosted top tagging. The energy flow basis provides a systematic framework for complete investigations of jet substructure using linear methods.

Keywords: Jets, QCD Phenomenology

ArXiv ePrint: 1712.07124

Open Access, © The Authors.
Article funded by SCOAP[3].

https://doi.org/10.1007/JHEP04(2018)013

JHEP04(2018)013

Contents

JHEP04(2018)013

1 Introduction

Jet substructure is the analysis of radiation patterns and particle distributions within the collimated sprays of particles (jets) emerging from high-energy collisions [1–5]. Jet substructure is central to many analyses at the Large Hadron Collider (LHC), finding applications in both Standard Model measurements [6–19] and in searches for physics beyond the Standard Model [20–43]. An enormous catalog of jet substructure observables has been developed to tackle specific collider physics tasks [44–48], such as the identification of boosted heavy particles or the discrimination of quark- from gluon-initiated jets.

The space of possible jet substructure observables is formidable, with few known complete and systematic organizations. Previous efforts to define classes of observables around organizing principles include: the jet energy moments and related Zernike polynomials to classify energy flow observables [49]; a pixelated jet image [50] to represent energy deposits in a calorimeter; the energy correlation functions (ECFs) [51] to highlight the N-prong substructure of jets; the generalized energy correlation functions (ECFGs) [52] based around soft-collinear power counting [53]; and a set of N-subjettiness observables [54–56] to capture N-body phase space information [57]. With any of these representations, there is no simple method to combine individual observables, so one typically uses sophisticated multivariate techniques such as neural networks to fully access the information contained in several observables [57–73]. Furthermore, the sense in which these sets "span" the space of jet substructure is often unclear, sometimes relying on the existence of complicated nonlinear functions to map observables to kinematic phase space.

In this paper, we introduce a powerful set of jet substructure observables organized directly around the principle of infrared and collinear (IRC) safety. These observables are multiparticle energy correlators with specific angular structures which directly result from IRC safety. Since they trace their lineage to the hadronic energy flow analysis of ref. [74], we call these observables the *energy flow polynomials* (EFPs) and we refer to the set of EFPs as the *energy flow basis*. In the language of ref. [74], the EFPs can be viewed as a discrete set of C-correlators, though our analysis is independent from the original C-correlator logic. Crucially, the EFPs form a *linear* basis of all IRC-safe observables, making them suitable for a wide variety of jet substructure contexts where linear methods are applicable.

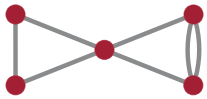
There is a one-to-one correspondence between EFPs and loopless multigraphs, which helps to visualize and calculate the EFPs. A multigraph is a graph where any two vertices can be connected by multiple edges; in this context, a *loop* is an edge from a vertex to itself, while a closed chain of edges is instead referred to as a *cycle*. For a multigraph G with N vertices and edges $(k,\ell) \in G$, the corresponding EFP takes the form:

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}, \tag{1.1}$$

where the jet consists of M particles, $z_i \equiv E_i / \sum_{j=1}^{M} E_j$ is the energy fraction carried by particle i, and θ_{ij} is the angular distance between particles i and j. The precise definitions of E_i and θ_{ij} will depend on the collider context, with energy and spherical (θ, ϕ) coordinates

JHEP04(2018)013

typically used for e^+e^- collisions, and transverse momentum p_T and rapidity-azimuth (y, ϕ) coordinates for hadronic collisions. For brevity, we often use the multigraph G to represent the formula for EFP_G in eq. (1.1), e.g.:



$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} \sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_1 i_3} \theta_{i_1 i_4} \theta_{i_1 i_5} \theta_{i_4 i_5}^2. \tag{1.2}$$

This paper is a self-contained introduction to the energy flow basis, with the following organization. Section 2 contains a general overview of the EFPs, with more detailed descriptions of eq. (1.1) and the correspondence to multigraphs. We also discuss a few different choices of measure for z_i and θ_{ij}. As already mentioned, EFPs are a special case of C-correlators [74], so not surprisingly, we find a close relationship between EFPs and other classes of observables that are themselves C-correlators, including jet mass, ECFs [51], certain generalized angularities [75], and energy distribution moments [49]. We also highlight features of the EFPs which are less well-known in the C-correlator-based literature.

In section 3, we give a detailed derivation of the EFPs as an (over)complete linear basis of all IRC-safe observables in the case of massless particles. Because this section is rather technical, it can be omitted on a first reading, though the logic just amounts to systematically imposing the constraints of IRC safety. In section 3.1, we use an independent (and arguably more transparent) logic from ref. [74] to show that any IRC-safe observable can be written as a linear combination of C-correlators:

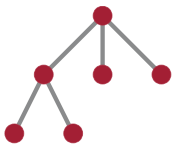
$$\mathcal{C}_N^{f_N} = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} \cdots E_{i_N} f_N(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu), \tag{1.3}$$

where f_N is an angular weighting function that is only a function of the particle directions $\hat{p}_i^\mu = p_i^\mu / E_i$ (and not their energies E_i). To derive eq. (1.3), we use the Stone-Weierstrass theorem [76] to expand an arbitrary IRC-safe observable in polynomials of particle energies, and then directly impose IRC safety and particle relabeling invariance. In section 3.2, we determine the angular structures of the EFPs by expanding f_N in terms of a discrete set of polynomials in pairwise angular distances. Remarkably, the discrete set of polynomials appearing in this expansion is in one-to-one correspondence with the set of non-isomorphic multigraphs, which facilitates indexing the EFPs by multigraphs to encode the geometric structure in eq. (1.1).

In section 4, we investigate the complexity of computing EFPs. Naively, eq. (1.1) has complexity $\mathcal{O}(M^N)$ due to the N nested sums over M particles. However, the rich analytic structure of eq. (1.1) and the graph representations of EFPs allow for numerous algorithmic speedups. Any EFP with a disconnected graph can be computed as the product of the EFPs corresponding to its connected components. Furthermore, we find that the Variable Elimination (VE) algorithm [77] can be used to vastly speed up the computation of many EFPs compared to the naive $\mathcal{O}(M^N)$ algorithm. VE uses the factorability of the summand to systematically determine a more efficient order for performing nested sums. For instance, all tree graphs can be computed in $\mathcal{O}(M^2)$ using VE. As an explicit example, consider an

JHEP04(2018)013

EFP with naive $\mathcal{O}(M^6)$ scaling:



$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} \theta_{i_1 i_2} \left(\sum_{i_3=1}^{M} z_{i_3} \theta_{i_1 i_3} \right)^2 \left(\sum_{i_4=1}^{M} z_{i_4} \theta_{i_2 i_4} \right)^2 . \tag{1.4}$$

The quantities in parentheses are computable in $\mathcal{O}(M^2)$, since they are length M lists with each element a sum over M objects, making the overall expression in eq. (1.4) computable in $\mathcal{O}(M^2)$. The efficient computation of the EFPs overcomes one of the main previous challenges in using higher-N multiparticle correlators in collider physics applications.[1]

In section 5, we perform numerical linear regression with EFPs for various jet observables. The linear spanning nature of the energy flow basis means that any IRC-safe observable $\mathcal{S}$ can be linearly approximated by EFPs, which we write as:

$$\mathcal{S} \simeq \sum_{G \in \mathcal{G}} s_G \, \mathrm{EFP}_G, \tag{1.5}$$

for some finite set of multigraphs $\mathcal{G}$ and some real coefficients s_G. One might worry that the number of EFPs needed to achieve convergence could be intractably large. In practice, though, we find that the required set of $\mathcal{G}$ needed for convergence is rather reasonable in a variety of jet contexts. While we find excellent convergence for IRC-safe observables, regressing with IRC-unsafe observables does not work as well, demonstrating the importance of IRC safety for the energy flow basis.

In section 6, we perform another test of eq. (1.5) by using linear classification with EFPs to distinguish signal from background jets. We consider three representative jet tagging problems: quark/gluon discrimination, boosted W tagging, and boosted top tagging. In this study, the observable appearing on the left-hand side of eq. (1.5) is the optimal IRC-safe discriminant for the two classes of jets. Remarkably, linear classification with EFPs performs comparably to multivariate machine learning techniques, such as jet images with convolutional neural networks (CNNs) [50, 63–66] or dense neural networks (DNNs) with a complete set of N-subjettiness observables [57]. Both the linear regression and classification models have few or no hyperparameters, illustrating the power and simplicity of linear learning methods combined with our fully general linear basis for IRC-safe jet substructure.

Our conclusions are presented in section 7, where we highlight the relevance of the energy flow basis to machine learning and discuss potential future applications and developments. A review of C-correlators and additional tagging plots are left to the appendices.

2 Energy flow polynomials

IRC-safe observables have long been of theoretical and experimental interest because observables which lack IRC safety are not well defined [78–81], or require additional care to calculate [82–86], in perturbative quantum chromodynamics (pQCD). More broadly,

[1] Sadly, fully-connected graphs, which correspond to the original ECFs [51], cannot be simplified using VE.

JHEP04(2018)013

though, IRC safety is a simple and natural organizing principle for high-energy physics observables, since IRC-safe observables probe the high-energy structure of an event while being insensitive to low-energy and collinear modifications. IRC safety is also an important property experimentally as IRC-safe observables are more robust to noise and finite detector granularity.

As argued in refs. [74, 87–89], the C-correlators in eq. (1.3) are a generic way to capture the IRC-safe structure of a jet, as long as one chooses an appropriate angular weighting function f_N. Later in section 3, we give an alternative proof that C-correlators span the space of IRC-safe observables and go on to give a systematic expansion for f_N. This expansion results in the EFPs, which yield an (over)complete linear basis for IRC-safe observables. In this section, we highlight the basic features of the EFPs and their relationship to previous jet substructure observables.

2.1 The energy flow basis

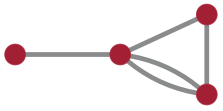
One can think of the EFPs as C-correlators that make specific, discrete choices for the angular weighting function f_N in eq. (1.3). True to their name, EFPs have angular weighting functions that are polynomial in pairwise angular distances θ_{ij}. The energy flow basis is therefore all C-correlators with angular structures that are unique monomials in θ_{ij}, meaning monomials that give algebraically different expressions once the sums in eq. (1.3) are performed. Since we intend to apply the energy flow basis for jet substructure, we remove the dependence on the overall jet kinematics by normalizing the particle energies by the total jet energy, $E_J \equiv \sum_{i=1}^{M} E_i$, leading to the EFPs written in terms of the energy fractions $z_i \equiv E_i/E_J$ as in eq. (1.1).

The uniqueness requirement on angular monomials can be better understood by developing a correspondence between monomials in θ_{ij} and multigraphs:

Multigraph/EFP Correspondence. *The set of loopless multigraphs on N vertices corresponds exactly to the set of angular monomials in $\{\theta_{i_k i_\ell}\}_{k<\ell\in\{1,\cdots,N\}}$. Each edge (k,ℓ) in a multigraph is in one-to-one correspondence with a term $\theta_{i_k i_\ell}$ in an angular monomial; each vertex j in the multigraph corresponds to a factor of z_{i_j} and summation over i_j in the EFP:*

(2.1)

Using eq. (2.1), the EFPs can be directly encoded by their corresponding multigraphs. For instance:



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_2 i_4}^2 \theta_{i_3 i_4}. \tag{2.2}$$

Since any labeling of the vertices gives an equivalent algebraic expression, we represent the graphs as unlabeled. The specification that the EFPs are unique monomials translates into the requirement that the corresponding multigraphs are non-isomorphic. Versions

JHEP04(2018)013

	Multigraph		Energy Flow Polynomial
N :	Number of vertices	$\Longleftrightarrow$	N-particle correlator
d :	Number of edges	$\Longleftrightarrow$	Degree of angular monomial
χ :	Treewidth $+1$	$\Longleftrightarrow$	Optimal VE complexity $\mathcal{O}(M^\chi)$
	Chromatic number	$\Longleftrightarrow$	Minimum number of prongs to not vanish
	Connected	$\Longleftrightarrow$	Prime
	Disconnected	$\Longleftrightarrow$	Composite

Table 1. Corresponding properties of multigraphs and EFPs.

of these multigraphs have previously appeared in the physics literature in the context of many-body configurations [90, 91], encoding all local scalar operators of a free theory [92], and in graphically depicting ECFs for jets [52, 93].

Table 1 contains a summary of the correspondence between the properties of EFPs and multigraphs. The number of graph vertices N corresponds to the number of particle sums in the EFP, and the number of graph edges d corresponds to the *degree* of the EFP (i.e. the degree of the underlying angular monomial). The number of separated prongs for which an individual EFP is first non-vanishing is the *chromatic number* of the graph: the smallest number of colors needed to color the vertices of the graph with no two adjacent vertices sharing a color. For computational reasons discussed further in section 4, we also care about the treewidth of the graph, which is related to the computational complexity χ of an EFP. Also for computational reasons, we make a distinction between connected or *prime* multigraphs and disconnected or *composite* multigraphs; the value of a composite EFP is simply the product of the prime EFPs corresponding to its connected components.

Because the EFP basis is infinite, a suitable organization and truncation scheme is necessary to use the basis in practice. In this paper, we usually truncate by restricting to the set of all multigraphs with at most d edges. This is a natural choice because it corresponds to truncating the approximation of the angular function f_N at degree d polynomials. Furthermore, this truncation results in a *finite* number of EFPs at each order of truncation, which is not true for truncation by the number of vertices. The number of multigraphs with exactly d edges is Sequence A050535 in the On-Line Encyclopedia of Integer Sequences (OEIS) [94, 95]; the number of connected multigraphs with exactly d edges is Sequence A076864 in the OEIS [94]. The numbers of EFPs in our truncation of the energy flow basis are the partial sums of these sequences, which are listed in table 2a up to $d = 10$. Table 2b tabulates the number of prime EFPs of degree d binned by N up to $d = 10$. Table 3 illustrates all connected multigraphs with $d \leq 5$ edges.

2.2 Energy and angular measures

There are many possible choices for the energy fraction z_i and angular measure θ_{ij} used to define the EFPs. In the analysis of section 3, this choice arises because there are many systematic expansions of IRC-safe observables in terms of energy-like and angular-like

JHEP04(2018)013

Maximum degree d		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prime EFPs	A076864	1	1	2	5	12	33	103	333	1 183	4 442	17 576
	Cumul.	1	2	4	9	21	54	157	490	1 673	6 115	23 691
All EFPs	A050535	1	1	3	8	23	66	212	686	2 389	8 682	33 160
	Cumul.	1	2	5	13	36	102	314	1 000	3 389	12 071	45 231

(a)

	d	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
N	**2**	1	1	1	1	1	1	1	1	1	1
	3		1	2	3	4	6	7	9	11	13
	4			2	5	11	22	37	61	95	141
	5				3	11	34	85	193	396	771
	6					6	29	110	348	969	2 445
	7						11	70	339	1 318	4 457
	8							23	185	1 067	4 940
	9								47	479	3 294
	10									106	1 279
	11										235

(b)

Table 2. (a) The number of EFPs (prime and all) organized by degree d, for d up to 10. The cumulative rows tally the number of EFPs with degree at most d, i.e. the number of basis elements truncated at that d. While these sequences grow quickly, the total number of all basis elements is at most 1000 for $d \leq 7$, which is computationally tractable. (b) The number of prime EFPs broken down by number of vertices N and number of edges d in the multigraph. All connected graphs (prime EFPs) for d up to 5 are shown explicitly in table 3.

JHEP04(2018)013

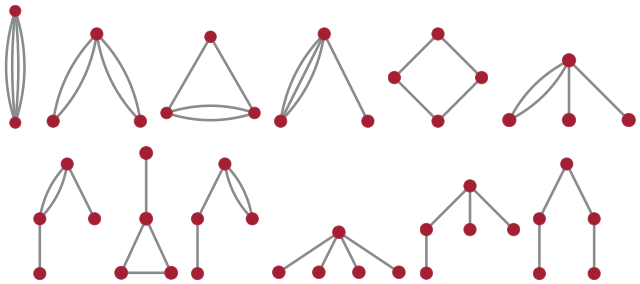
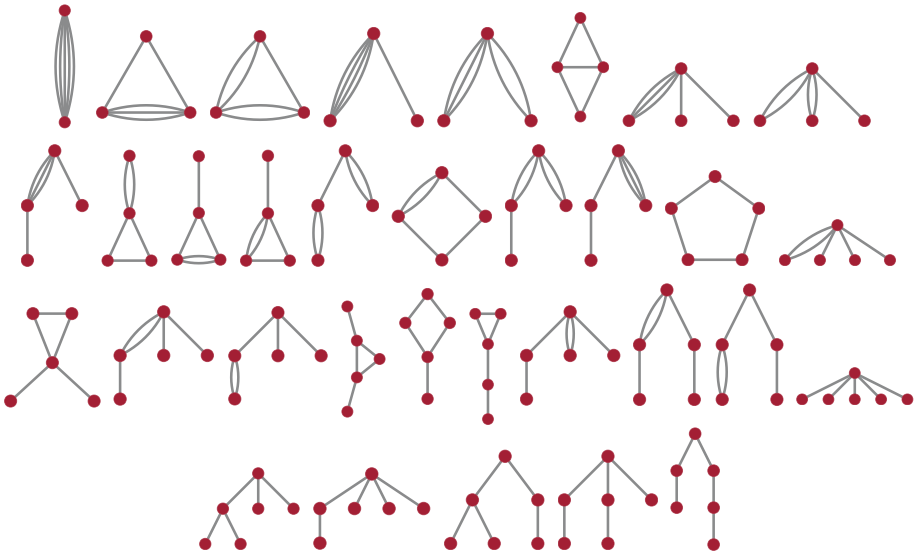
Degree	Connected Multigraphs
$d = 0$	
$d = 1$	
$d = 2$	
$d = 3$	
$d = 4$	
$d = 5$	

Table 3. All non-isomorphic, loopless, connected multigraphs organized by the total number of edges d, up to $d = 5$, sorted by their number of vertices N. Note that for a fixed number of edges d, the total number of multigraphs (connected or not) is finite. These graphs correspond to the $d \leq 5$ prime EFPs counted in table 2a. Image files for all of the prime EFP multigraphs up to $d = 7$ are available here.

JHEP04(2018)013

quantities. Typically, one wants to work with observables that respect the appropriate Lorentz subgroup for the collision type of interest. For e^+e^- colliders, the symmetries are the group of rotations about the interaction point, and for hadron colliders they are rotations about and boosts along the beam axis (sometimes with a reflection in the plane perpendicular to the beam). Therefore, the energy fractions z_i usually use particle energies E_i at an e^+e^- collider and particle transverse momenta $p_{T,i}$ at a hadron collider.

For the angular weighting function f_N, though, there are many different angular structures one can build out of the particle directions $\hat{p}_i^\mu$. The EFPs use the simplest and arguably most natural choice to expand the angular behavior: *pairwise* angular distances θ_{ij}, determined using spherical coordinates (θ, ϕ) at an e^+e^- collider and rapidity-azimuth coordinates (y, ϕ) at a hadron collider. Other classes of observables, such as ECFs [51] and ECFGs [52], also use pairwise angles since they manifestly respect the underlying Lorentz subgroup. For building the EFPs, is important that the θ_{ij}, or any other choice of geometric object, be sufficient to reconstruct the value of the original function f_N in terms of the $\hat{p}_i^\mu$. For pairwise angles, this property can be shown by triangulation, under the assumption that the observable in question does not depend on the overall jet direction nor on rotations or reflections about the jet axis. Since jets are collimated sprays of particles, the θ_{ij} are typically small and are good expansion parameters.

At various points in this paper, we explore three different energy/angular measures. For e^+e^- collisions, our default is:

$$\text{e^+e^- Default} \qquad \begin{aligned} z_i &= \frac{E_i}{E_J}, \qquad E_J \equiv \sum_{i=1}^{M} E_i, \\ \theta_{ij} &= \left(\frac{2\, p_i^\mu p_{j\mu}}{E_i E_j}\right)^{\beta/2}, \end{aligned} \tag{2.3}$$

where $\beta > 0$ is an angular weighting factor. For the hadron collider studies in sections 5 and 6, we use:

$$\text{Hadronic Default} \qquad \begin{aligned} z_i &= \frac{p_{T,i}}{p_{T,J}}, \qquad p_{T,J} \equiv \sum_{i=1}^{M} p_{T,i}, \\ \theta_{ij} &= \left(\Delta y_{ij}^2 + \Delta\phi_{ij}^2\right)^{\beta/2}, \end{aligned} \tag{2.4}$$

where $\Delta y_{ij} \equiv y_i - y_j$, $\Delta\phi_{ij} \equiv \phi_i - \phi_j$ are determined by the rapidity y_i and azimuth ϕ_i of particle i. This measure is rotationally-symmetric in the (y, ϕ) plane, which is the most commonly used case in jet substructure. For situations where this rotational symmetry is not desirable (such as for jet pull [96]), we can instead use a two-dimensional measure that treats the rapidity and azimuthal directions separately:

$$\text{Hadronic Two-Dimensional} \qquad \begin{aligned} z_i &= \frac{p_{T,i}}{p_{T,J}}, \qquad p_{T,J} \equiv \sum_{i=1}^{M} p_{T,i}, \\ \theta_{ij} &= \Delta y_{ij} \text{ or } \Delta\phi_{ij}, \end{aligned} \tag{2.5}$$

where each line on the multigraph now has an additional decoration to indicate whether it corresponds to Δy or $\Delta\phi$.

JHEP04(2018)013

We emphasize that the choice of measure is not unique, though it is constrained by the IRC-safety arguments in section 3. For example, IRC safety requires that the energy-like quantities appear linearly in z_i. For the default measures, the angular exponent β can take on any positive value and still be consistent with IRC safety. Depending on the context, different choices of β can lead to faster or slower convergence of the EFP expansion, with $\beta < 1$ emphasizing smaller values of θ_{ij} and $\beta > 1$ emphasizing larger values of θ_{ij}. For special choices of z_i and θ_{ij}, some EFPs may be linearly related, a point we return to briefly in section 4.1.

2.3 Relation to existing substructure observables

Many familiar jet observables can be nicely interpreted in the energy flow basis. When an observable can be written as a simple expression in terms of particle four-momenta or in terms of energies and angles, the energy flow decomposition can often be performed exactly. Some of the most well-known observables, such as jet mass and energy correlation functions, are exactly finite linear combinations of EFPs (with appropriate choice of measure), which one might expect since they also correspond to natural C-correlators. Unless otherwise specified, the analysis below uses the default hadronic measure in eq. (2.4) with $\beta = 1$ and treats all particles as massless.[2]

2.3.1 Jet mass

Jet mass is most basic jet substructure observable, and not surprisingly, it has a nice expansion in the energy flow basis. In particular, the squared jet mass divided by the jet energy squared is an exact $N = 2$ EFP using the e^+e^- measure in eq. (2.3) with $\beta = 1$:

$$e^+e^- : \qquad \frac{m_J^2}{E_J^2} = \frac{1}{2}\sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2} \left(\frac{2\, p_{i_1}^{\mu} p_{i_2\mu}}{E_{i_1} E_{i_2}}\right) = \frac{1}{2} \times$$

. (2.6)

Note that mass is exactly an EFP for any $\beta = 2/N$ measure choice.

For the hadronic measure in eq. (2.4) with $\beta = 1$, there is an approximate equivalence with the squared jet mass divided by the jet (scalar) transverse momentum:

$$\text{Hadronic} : \qquad \frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\cosh(\Delta y_{i_1 i_2}) - \cos(\Delta\phi_{i_1 i_2})) = \frac{1}{2} \times$$

$+ \cdots$. (2.7)

Since the jet mass is not exactly rotationally symmetric in the rapidity-azimuth plane, the subleading terms in eq. (2.7) are not fully encompassed by the simplified set of hadronic

[2] A proper treatment of non-zero particle masses would require an additional expansion in the *velocities* of the particles (see related discussion in refs. [97, 98]). To avoid these complications, one can interpret all particles as being massless in the E-scheme [97], i.e. $p^{\mu}_{\text{rescaled}} = E\,(1, \hat{p})$ with $\hat{p} = \vec{p}/|\vec{p}|$.

JHEP04(2018)013

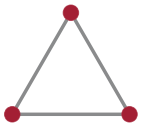
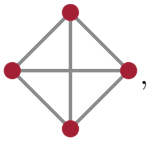
observables depending only on $\{\Delta y_{ij}^2 + \Delta\phi_{ij}^2\}$, but could be fully encompassed by using an expansion in $\{\Delta y_{ij}, \Delta\phi_{ij}\}$ as in eq. (2.5). For narrow jets, these higher-order terms in the expansion become less relevant since Δy_{ij}, $\Delta\phi_{ij} \ll 1$.[3]

2.3.2 Energy correlation functions

The ECFs are designed to be sensitive to N-prong jet substructure [51]. They can be written as a C-correlator, eq. (1.3), with a particular choice of angular weighting function:

$$f_N^{(\beta)}(\{\theta_{ij}\}) = \prod_{i<j} \theta_{ij}^{\beta}, \tag{2.8}$$

where $\theta_{ij} = (\Delta y_{ij}^2 + \Delta\phi_{ij}^2)^{1/2}$. In terms of multigraphs, the ECFs correspond to complete graphs on N vertices:

$e_2^{(\beta)} =$  , $e_3^{(\beta)} =$  , $e_4^{(\beta)} =$  , (2.9)

which are EFPs using the measure in eq. (2.4) with exponent β.

The ECFs have since been expanded to a more flexible set of observables referred to as the ECFGs [52]. Letting $\min^{(m)}$ indicate the m-th smallest element in a set, the ECFGs are also C-correlators with angular weighting function:

$${}_v f_N^{(\beta)}(\{\theta_{ij}\}) = \prod_{m=1}^{v} \min_{i<j}^{(m)} \{\theta_{ij}^{\beta}\}. \tag{2.10}$$

The ECFGs do not have an exact multigraph correspondence due to the presence of the min function, but are evidently closely related to the EFPs since they share a common energy structure. The min function itself can be approximated by polynomials in its arguments, which induces an approximating series for the ECFGs in terms of EFPs when plugged into the common energy structure.

Both the EFPs and the ECFGs represent natural extensions of the ECFs but in different directions. From our graph-theoretic perspective, the EFPs extend the ECFs to non-fully-connected graphs. The ECFGs extend the scaling properties of the ECFs into observables with independent energy and angular scalings. As discussed in section 2.4, there are angular structures possible in the EFPs that are not possible in the ECFGs. As with any jet substructure analysis, the choice of which set of observables to use depends on the physics of interest, with the EFPs designed for linear completeness and the ECFGs designed for nice power-counting properties.

[3] Alternatively, we could use a measure with $\theta_{ij} = \left(\frac{2\, p_i^\mu p_{j\mu}}{p_{T,i} p_{T,j}}\right)^{\beta/2}$, similar in spirit to the Conical Geometric measure of ref. [99], to exactly recover the jet mass.

JHEP04(2018)013

2.3.3 Angularities

Next, we consider the IRC-safe jet angularities [75] (see also refs. [100–103]) defined by:

$$\lambda^{(\alpha)} = \sum_{i=1}^{M} z_i \, \theta_i^{\alpha}, \tag{2.11}$$

where $\alpha > 0$ is an angular exponent and θ_i denotes the distance of particle i to the jet axis. For concreteness and analytic tractability, we take the jet axis to be the p_T-weighted centroid in (y, ϕ)-space, such that the jet axis is located at:

$$y_J = \sum_{j=1}^{M} z_j y_j, \qquad \phi_J = \sum_{j=1}^{M} z_j \phi_j. \tag{2.12}$$

With this, the angularities can be expressed as:

$$\begin{aligned}
\lambda^{(\alpha)} &= \sum_{i_1=1}^{M} z_{i_1} \left((y_{i_1} - y_J)^2 + (\phi_{i_1} - \phi_J)^2 \right)^{\alpha/2} \\
&= \sum_{i_1=1}^{M} z_{i_1} \left(\left(\sum_{i_2=1}^{M} z_{i_2} \Delta y_{i_1 i_2} \right)^2 + \left(\sum_{i_2=1}^{M} z_{i_2} \Delta \phi_{i_1 i_2} \right)^2 \right)^{\alpha/2} \\
&= \sum_{i_1=1}^{M} z_{i_1} \left(\sum_{i_2=1}^{M} z_{i_2} \theta_{i_1 i_2}^2 - \frac{1}{2} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} z_{i_2} z_{i_3} \theta_{i_2 i_3}^2 \right)^{\alpha/2}.
\end{aligned} \tag{2.13}$$

For even α, the parenthetical in eq. (2.13) can be expanded and identified to be a linear combination of EFPs with $N = \alpha$ and $d = \alpha$ (see ref. [49] for a related discussion). For $\alpha = 2$, eq. (2.13) implies:

$$\lambda^{(2)} = \frac{1}{2} \sum_{i \in J} \sum_{j \in J} z_i z_j \theta_{ij}^2 = \frac{1}{2} \times$$

. (2.14)

For $\alpha = 4$ and $\alpha = 6$, eq. (2.13) implies:

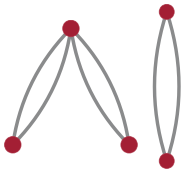
$$\lambda^{(4)} =$$

$$- \frac{3}{4} \times$$

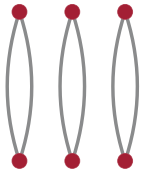
, (2.15)

$$\lambda^{(6)} =$$

$$- \frac{3}{2} \times$$



$$+ \frac{5}{8} \times$$



. (2.16)

JHEP04(2018)013

This can be continued for arbitrarily high, even α. Thus, the even α angularities are exact, non-trivial linear combinations of EFPs, illustrating the close connections between the two classes of observables. While angularities with odd or non-integer α do not have the same analytic tractability, the specific case of $\alpha = 1/2$ is shown to be numerically well approximated by EFPs in section 5.3.

2.3.4 Geometric moment tensors

Next, we consider observables based on the two-dimensional geometric moment tensor of the energy distribution in the (y, ϕ)-plane [49, 60]:

$$\mathbf{C} = \sum_{i \in J} z_i \begin{pmatrix} \Delta y_i^2 & \Delta y_i \Delta \phi_i \\ \Delta \phi_i \Delta y_i & \Delta \phi_i^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \sum_{i,j} z_i z_j \Delta y_{ij}^2 & \frac{1}{2} \sum_{i,j} z_i z_j \Delta y_{ij} \Delta \phi_{ij} \\ \frac{1}{2} \sum_{i,j} z_i z_j \Delta \phi_{ij} \Delta y_{ij} & \frac{1}{2} \sum_{i,j} z_i z_j \Delta \phi_{ij}^2 \end{pmatrix}, \tag{2.17}$$

where the distances are measured with respect to the p_T-weighted centroid axis (y_J, ϕ_J) from eq. (2.12). Useful observables can be constructed from the trace and determinant of $\mathbf{C}$, such as planar flow $\text{Pf} = 4 \det \mathbf{C} / (\text{tr}\, \mathbf{C})^2$ [101, 104], which is a ratio of two IRC-safe observables.

We see that eq. (2.17) is exactly a matrix of EFPs with $N = 2$ and the two-dimensional hadronic measure from eq. (2.5). The trace $\text{tr}\, \mathbf{C}$ and determinant $\det \mathbf{C}$ have the rotational symmetry in the (y, ϕ)-plane of the default hadronic measure from eq. (2.4), allowing them to be written as linear combinations of EFPs with that measure:

$$\text{tr}\, \mathbf{C} = \frac{1}{2} \times$$  ,

$$4 \det \mathbf{C} =$$  $$- \frac{1}{2} \times$$  . (2.18)

In ref. [49], a general class of energy flow moments was explored and categorized, with the goal of classifying observables according to their energy flow distributions. These energy flow moments are defined with respect to a specified jet axis:

$$I_{k_1 \cdots k_N} \equiv \sum_{i=1}^{M} z_i \, x_{k_1}^{(i)} \cdots x_{k_N}^{(i)}, \tag{2.19}$$

where $k_i \in \{1, 2\}$, $x_1^{(i)} = \Delta y_i = y_i - y_J$ and $x_2^{(i)} = \Delta \phi_i = \phi_i - \phi_J$. Using the p_T-weighted centroid axis, this is the natural generalization of eq. (2.17), with the special case of $I_{k_1 k_2} = (\mathbf{C})_{k_1 k_2}$. By performing a similar analysis to the one used to arrive at eq. (2.18), one can show that any scalar constructed by contracting the indices of a product of objects in eq. (2.19) can be decomposed into an exact linear combination of EFPs.

2.4 Going beyond existing substructure observables

Because the EFPs are C-correlators that span the space of IRC-safe observables, their angular structures should encompass all possible behaviors of C-correlators. By contrast,

JHEP04(2018)013

the ECFs and ECFGs mentioned in section 2.3.2 have more restricted behaviors, and it is illuminating to understand the new kinds of structures present in the EFPs.

Without loss of generality, the angular weighting function f_N in eq. (1.3) can be taken to be a symmetric function of the particle directions $\hat{p}_i^\mu$ due to the symmetrization provided by the sum structure (see eq. (3.14) below). The ECFs and ECFGs exhibit a stronger symmetry, though, since the angular functions in eqs. (2.8) and (2.10) are invariant under the swapping any two pairwise angles θ_{ij}. This symmetry is manifested in the ECFs multigraphs in eq. (2.9) by the fact that all pairs of indices are connected by the same number of edges.

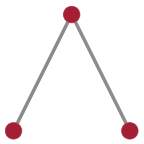
We can easily see that the pairwise swap symmetry of the ECFs is stronger than the full permutation symmetry of the EFPs: the group of permutations of the angular distances θ_{ij} has $\binom{N}{2}!$ elements, whereas the group of permutations of the indices $\{i_a\}$ has $N!$ elements. An example of an EFP that does not satisfy the stronger symmetry is the following $N = 4$ graph:

$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_1 i_3} \theta_{i_1 i_4}. \tag{2.20}$$

The angular weighting function of the EFP in eq. (2.20) is symmetric under the 4! permutations in the indices (vertices) $i_a \to i_{\sigma(a)}$ but not under the exchange of pairwise angles (edges) $\theta_{i_1 i_3} \to \theta_{i_2 i_3}$ which would result in a different EFP, namely:

$$\neq \quad . \tag{2.21}$$

Another feature of the ECFs and ECFGs is that their angular weighting function f_N vanishes whenever two of its arguments become collinear. Indeed, one of the present authors made the erroneous claim in ref. [52] that this vanishing behavior was required by collinear safety.[4] Instead, the argument in section 3.1.3 shows this not to be the case, and observables defined by eq. (1.3) are IRC safe for any sufficiently smooth and non-singular f_N. An example of an EFP that does not necessarily vanish when two of its arguments become collinear is the following $N = 3$ graph:



$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2} \theta_{i_1 i_3}, \tag{2.22}$$

which does not vanish when $\hat{p}_{i_2}^\mu \to \hat{p}_{i_3}^\mu$. More generally, any non-fully-connected graph will not vanish in every collinear limit, but the corresponding EFP will still be collinear safe.

By relaxing the restrictions on the angular weighting function f_N to those minimally required by IRC safety, the energy flow basis captures all topological structures which can possibly appear in a C-correlator, beyond just the ones described by ECFs and ECFGs.

[4] If the sums are taken over distinct N-tuples as in ref. [52], then the angular function does have to vanish on collinearity for C safety. In general, non-collinearly-vanishing angular functions are C safe if the sum is taken over all N-tuples of particles, including sets with repeated indices.

JHEP04(2018)013

3 Constructing a linear basis of IRC-safe observables

Having introduced the EFPs, we now give a detailed argument that they linearly span the space of IRC-safe observables. Due to its more technical nature, this section can be omitted on a first reading, and the reader may skip to section 4. Refs. [74, 87–89] argue that, from the point of view of quantum field theory, all IRC-safe information about the jet structure should be contained in the C-correlators. In section 3.1, we independently arrive at the same conclusion by a direct application of IRC safety. We then go on in section 3.2 to expand the angular structure of the C-correlators to find a correspondence between multigraphs and EFPs.

An IRC-safe observable $\mathcal{S}$ depends only on the unordered set of particle four-momenta $\{p_i^\mu\}_{i=1}^M$, and not any non-kinematic quantum numbers. An observable defined on $\{p_i^\mu\}_{i=1}^M$ can alternatively be thought of as a collection of functions, one for each number of particles M. IRC safety then imposes constraints on this collection and thereby induces relations between the functions. The requirement of IR safety imposes the constraint [81]:

$$\mathcal{S}(\{p_1^\mu, \ldots, p_M^\mu\}) = \lim_{\varepsilon \to 0} \mathcal{S}(\{p_1^\mu, \ldots, p_M^\mu, \varepsilon\, p_{M+1}^\mu\}), \qquad \forall p_{M+1}^\mu, \tag{3.1}$$

while the requirement of C safety imposes the constraint:

$$\mathcal{S}(\{p_1^\mu, \ldots, p_M^\mu\}) = \mathcal{S}(\{p_1^\mu, \ldots, (1-\lambda)p_M^\mu, \lambda p_M^\mu\}), \qquad \forall \lambda \in [0,1]. \tag{3.2}$$

Eq. (3.1) says that the observable is unchanged by the addition of infinitesimally soft particles, while eq. (3.2) guarantees that the observable is insensitive to a collinear splitting of particles.

As written, only particle M is affected in eq. (3.2). The indexing used to identify particles, however, is arbitrary and these properties continue to hold when the particles are reindexed. This *particle relabeling symmetry* is not an additional constraint that is imposed but rather a consequence of assigning labels to an unordered set of particles. These three restrictions — IR safety, C safety, and particle relabeling symmetry — are necessary and sufficient conditions for obtaining the energy flow basis.

Throughout this analysis, particles are treated as massless, $p_i^\mu = E_i\, \hat{p}_i^\mu$, where $\hat{p}_i^\mu$ is purely geometric. Note that we could replace E_i with any quantity linearly dependent on energy, such as the transverse momentum $p_{T,i}$, which corresponds to making a different choice of measure in section 2.2.

3.1 Expansion in energy

Consider an arbitrary IRC-safe observable $\mathcal{S}$, expanded in terms of the particle energies. If the observable has a simple analytic dependence on the energies, then the usual Taylor expansion can be used:

$$\mathcal{S} = \mathcal{S}_M|_{\{E\}=0} + \sum_{i_1=1}^M E_i \left.\frac{\partial \mathcal{S}_M}{\partial E_{i_1}}\right|_{\{E\}=0} + \frac{1}{2}\sum_{i_1=1}^M \sum_{i_2=1}^M E_{i_1} E_{i_2} \left.\frac{\partial^2 \mathcal{S}_M}{\partial E_{i_1} \partial E_{i_2}}\right|_{\{E\}=0} + \cdots, \tag{3.3}$$

JHEP04(2018)013

where M is the particle multiplicity and the derivatives are evaluated at vanishing energies. An example of this is the jet mass from eq. (2.6):

$$m_J^2 = \sum_{i=1}^{M}\sum_{j=1}^{M} \eta_{\mu\nu} p_i^\mu p_j^\nu = \sum_{i=1}^{M}\sum_{j=1}^{M} E_i E_j \eta_{\mu\nu}\, \hat{p}_i^\mu \hat{p}_j^\nu, \tag{3.4}$$

where $\eta_{\mu\nu}$ is the Minkowski metric. This expression is already in the form of eq. (3.3) with:

$$\frac{\partial^2 m_J^2}{\partial E_{i_1} \partial E_{i_2}} = 2\eta_{\mu\nu} \hat{p}_{i_1}^\mu \hat{p}_{i_2}^\nu, \tag{3.5}$$

and all other Taylor coefficients zero. See section 2.3 for additional examples of observables with explicit formulas for which eq. (3.3) can be applied.

For some observables, though, a Taylor expansion may be difficult or impossible to obtain. The simplest example is a non-differentiable observable. This is the case for m_J (rather than m_J^2); the presence of the square root spoils the existence of a Taylor expansion, but the square root can be nonetheless approximated by polynomials arbitrarily well in a bounded interval. A more complicated case is if the observable is defined in terms of an algorithm, such as a groomed jet mass [5, 105–109], and an explicit formula in terms of particle four-momenta would not be practical to differentiate or write down. Similarly, the observable could be a non-obvious function of the particles, i.e. the optimal observable to accomplish some task.

In cases without a Taylor expansion, the Stone-Weierstrass theorem [76] still guarantees that the observable can be approximated over some bounded energy range by polynomials in the energies.[5] We write down such an expansion by considering all possible polynomials in the energies and multiplying each one by a different geometric function. Combining all terms of degree N into $\mathcal{C}_N$, the expansion is:

$$\mathcal{S} \simeq \sum_{N=0}^{N_{\max}} \mathcal{C}_N, \qquad \mathcal{C}_N \equiv \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} C^{(M)}_{i_1\cdots i_N}(\hat{p}_1^\mu, \ldots, \hat{p}_M^\mu) \prod_{j=1}^{N} E_{i_j}, \tag{3.6}$$

where $C^{(M)}_{i_1\cdots i_n}(\hat{p}_1^\mu, \ldots, \hat{p}_M^\mu)$ are geometric angular functions, which depend on the indices of the energy factors $i_1 \cdots i_n$ and could in general be different for different multiplicities M. The Stone-Weierstrass theorem guarantees that there is a maximum degree $N_{\max}$ in this energy expansion for any given desired accuracy, but places no further restrictions on the $\mathcal{C}_N$.

To derive constraints on these angular functions $C^{(M)}_{i_1\cdots i_N}$, we impose the three key properties of IR safety in section 3.1.1, particle relabeling invariance in section 3.1.2, and C safety in section 3.1.3, which we summarize in section 3.1.4. In applying these properties, we will often use the fact that when setting two expressions for the observable $\mathcal{S}$ equal to each other, we can read off term-by-term equality by treating the particle energies as

[5] A version of this theorem that suffices for our purposes can be phrased as follows: for any continuous, real-valued function f defined on a compact subset $X \subset \mathbb{R}^n$, for all $\epsilon > 0$ there exists a polynomial p of finite degree at most $N_{\max}$ such that $|p(\mathbf{x}) - f(\mathbf{x})| < \epsilon$ for all $\mathbf{x} \in X$. Conceptually, this theorem is used to approximate any continuous function on a bounded region by a polynomial.

JHEP04(2018)013

independent quantities:

$$\mathcal{S} = \mathcal{S}' \quad \Longrightarrow \quad \mathcal{C}_N = \mathcal{C}'_N, \quad \forall N \leq N_{\max}. \tag{3.7}$$

Note that the sum structure in eq. (3.6) implies that, without loss of generality, the angular functions can be taken to depend only on the labels $i_1, \ldots, i_N$ as an unordered set.

3.1.1 Infrared safety

IR safety constrains the angular functions appearing in the expansion of eq. (3.6) in two ways: by restricting which particle directions contribute to a particular term in the sum and by relating angular functions of different multiplicities.

First, consider a particular angular function, $C^{(M)}_{i_1\cdots i_N}$ in eq. (3.6), and some particle $j \notin \{i_1, \ldots, i_N\}$. Consider particle j in the soft limit: if $C^{(M)}_{i_1\cdots i_N}$ depends on $\hat{p}^\mu_j$ in any way, then IR safety is violated because E_j does not appear in the product of energies but the value of the observable changes as the direction of j is changed. Hence, IR safety imposes the requirement that

$$C^{(M)}_{i_1\cdots i_N}(\hat{p}^\mu_1, \ldots, \hat{p}^\mu_M) = C^{(M)}_{i_1\cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}), \tag{3.8}$$

namely the indices of the arguments must match those of the angular function. Note that we must always write $C^{(M)}_{i_1\cdots i_N}$ with N arguments, even if some are equal due to indices coinciding.

Next, consider two polynomial approximations of the same observable: one as a function of M particles and the other as a function of $M+1$ particles. In the soft limit of particle $M+1$, $E_{M+1} \to 0$, the IR safety of $\mathcal{S}$, written formally in eq. (3.1), guarantees that the function of $M+1$ particles approaches the function of M particles. In terms of the corresponding polynomial approximations, we have that:

$$\sum_{i_1=1}^{M+1}\cdots\sum_{i_N=1}^{M+1} C^{(M+1)}_{i_1\cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) \prod_{j=1}^{N} E_{i_j} = \sum_{i_1=1}^{M}\cdots\sum_{i_N=1}^{M} C^{(M)}_{i_1\cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) \prod_{j=1}^{N} E_{i_j} + \mathcal{O}(E_{M+1}). \tag{3.9}$$

We see from eq. (3.9) that the same angular coefficients from the polynomial approximation of the function of $M+1$ particles can be validly chosen for the approximation of the function of M particles, with the following equality of angular functions:

$$C^{(M+1)}_{i_1\cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) = C^{(M)}_{i_1\cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) \equiv C_{i_1\cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}), \tag{3.10}$$

which says that the multiplicity label on the angular functions can be dropped.

As a result of enforcing IR safety, the dependence of the angular functions on multiplicity has been eliminated, as well as the dependence of a given angular function on any particles with indices not appearing in its subscripts.

3.1.2 Particle relabeling symmetry

Now, using particle relabeling symmetry, for all $\sigma \in S_M$, where S_M is the group of permutations of M objects, we have that $\mathcal{C}_N$ is unchanged by the replacement $E_{i_j} \to E_{\sigma(i_j)}$ and

JHEP04(2018)013

$\hat{p}^\mu_{i_j} \to \hat{p}^\mu_{\sigma(i_j)}$. With the angular functions as constrained by IR safety, the particle relabeling invariance of $\mathcal{C}_N$ can be written as:

$$\mathcal{C}_N = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} C_{i_1 \cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) \prod_{j=1}^{N} E_{i_j} \tag{3.11}$$

$$= \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} C_{i_1 \cdots i_N}(\hat{p}^\mu_{\sigma(i_1)}, \ldots, \hat{p}^\mu_{\sigma(i_N)}) \prod_{j=1}^{N} E_{\sigma(i_j)}$$

$$= \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} C_{\sigma^{-1}(i_1) \cdots \sigma^{-1}(i_N)}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) \prod_{j=1}^{N} E_{i_j}, \tag{3.12}$$

where the sums were reindexed according to σ^{-1}. In particular, from eq. (3.12), we have for any $\sigma \in S_M$ that:

$$C_{i_1 \cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}) = C_{\sigma(i_1) \cdots \sigma(i_N)}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N}). \tag{3.13}$$

Eq. (3.13) allows us to permute the indices of $C_{i_1 \cdots i_N}$ within S_M, equating previously unrelated angular functions.

As written, $C_{i_1 \cdots i_N}$ is not necessarily symmetric in its arguments. Without loss of generality, though, we can symmetrize $C_{i_1 \cdots i_N}$ without changing the value of $\mathcal{C}_N$ as follows:

$$\mathcal{C}_N = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} \underbrace{\frac{1}{N!} \sum_{\sigma \in S_N} C_{i_1 \cdots i_N}(\hat{p}^\mu_{\sigma(i_1)}, \ldots, \hat{p}^\mu_{\sigma(i_N)})}_{C'_{i_1 \cdots i_N}(\hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_N})} \prod_{j=1}^{N} E_{i_j}, \tag{3.14}$$

where $C'_{i_1 \cdots i_N}$ is now symmetric in its arguments. We assume in the next step of the derivation that the angular weighting functions are symmetric in their arguments.

3.1.3 Collinear safety

The key requirement for restricting the form of $C_{i_1 \cdots i_N}$ is C safety. If the angular weighting function(s) were required to vanish whenever two of the inputs were collinear, then the observable would be manifestly C safe (see e.g. [52]); this is a sufficient condition for C safety but not a necessary one. More generally, one can have non-zero angular functions of N arguments even when subsets of the arguments are collinear.

Using the IR safety argument of eq. (3.10) and the particle relabeling symmetry of eq. (3.13), we can relate any angular function $C_{i_1 \cdots i_N}$ to one of the following:

$$\begin{aligned}
&C_{123\cdots N}(\hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_2}, \hat{p}^\mu_{i_3}, \ldots, \hat{p}^\mu_{i_N}),\\
&C_{1123\cdots(N-1)}(\hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_2}, \hat{p}^\mu_{i_3}, \ldots, \hat{p}^\mu_{i_{N-1}}),\\
&C_{112234\cdots(N-2)}(\hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_2}, \hat{p}^\mu_{i_2}, \hat{p}^\mu_{i_3}, \hat{p}^\mu_{i_4}, \ldots, \hat{p}^\mu_{i_{N-2}}),\\
&C_{1112234\cdots(N-3)}(\hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_2}, \hat{p}^\mu_{i_2}, \hat{p}^\mu_{i_3}, \hat{p}^\mu_{i_4}, \ldots, \hat{p}^\mu_{i_{N-3}}),\\
&\vdots\\
&C_{11\cdots1}(\hat{p}^\mu_{i_1}, \hat{p}^\mu_{i_1}, \ldots, \hat{p}^\mu_{i_1}),
\end{aligned} \tag{3.15}$$

JHEP04(2018)013

where there is one of these "standard" angular functions for each integer partition of N. In particular, the length of the integer partition is how many unique indices appear in the subscript and the values of the partition indicate how many times each index is repeated.

The role of C safety is to impose relationships between these standard angular functions, eventually showing that the only required function is $C_{123\cdots N}$. Intuitively, this means that as any set of particles become collinear, the angular dependence is that of collinear limit of N arbitrary directions. The proof that this follows from C safety, however, is the most technically involved step of this derivation.

The requirement of C safety in eq. (3.2) implies that $\mathcal{S}$ is unchanged whether one considers $\{E_i, \hat{p}_i^\mu\}_{i=1}^M$ or the same particles with a collinear splitting of the first particle, $\{\tilde{E}_i, \hat{p}_i^\mu\}_{i=0}^M$, where:

$$\tilde{E}_0 = (1-\lambda)E_1, \qquad \tilde{E}_1 = \lambda E_1, \qquad \hat{p}_0^\mu = \hat{p}_1^\mu, \qquad \tilde{E}_i = E_i, \tag{3.16}$$

for all $\lambda \in [0,1]$ and $i > 1$. Rewriting eq. (3.11), we can explicitly separate out the terms of the sums involving k collinearly split indices $\{0,1\}$:

$$\mathcal{C}_N = \sum_{i_1=0}^{M} \cdots \sum_{i_N=0}^{M} C_{i_1\cdots i_N}(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu) \prod_{j=1}^{N} \tilde{E}_{i_j} \tag{3.17}$$

$$= \sum_{k=0}^{N} \binom{N}{k} \sum_{i_1=0}^{1} \cdots \sum_{i_k=0}^{1} \sum_{i_{k+1}=2}^{M} \cdots \sum_{i_N=2}^{M} C_{i_1\cdots i_N}(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu) \prod_{j=1}^{N} \tilde{E}_{i_j}, \tag{3.18}$$

where in going to this last expression, we have used the symmetry of $C_{i_1\cdots i_N}$ in its arguments and accounted for the degeneracy of such terms using the binomial factor $\binom{N}{k}$. We then insert the collinear splitting kinematics of eq. (3.16) into eq. (3.18),

$$\mathcal{C}_N = \sum_{k=0}^{N} \binom{N}{k} \sum_{i_1=0}^{1} \cdots \sum_{i_k=0}^{1} \lambda^{\sum_{a=1}^k i_a} (1-\lambda)^{k-\sum_{a=1}^k i_a} E_1^k \tag{3.19}$$

$$\times \sum_{i_{k+1}=2}^{M} \cdots \sum_{i_N=2}^{M} C_{i_1\cdots i_N}(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu) \prod_{j=k+1}^{N} E_{i_j}$$

$$= \sum_{k=0}^{N} \binom{N}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} \lambda^\ell (1-\lambda)^{k-\ell} E_1^k \tag{3.20}$$

$$\times \sum_{i_{k+1}=2}^{M} \cdots \sum_{i_N=2}^{M} C_{\underbrace{0\cdots 0}_{\ell}\underbrace{1\cdots 1}_{k-\ell} i_{k+1}\cdots i_n}(\hat{p}_1^\mu, \ldots, \hat{p}_1^\mu, \hat{p}_{i_{k+1}}^\mu, \ldots, \hat{p}_{i_N}^\mu) \prod_{j=k+1}^{N} E_{i_j},$$

where in going to this last expression, we have used the particle relabeling symmetry of eq. (3.13) to sort the $\{0,1\}$ subscript indices of the angular functions.

The constraint of C safety says that eq. (3.20) is equal to eq. (3.11) on the non-collinearly split event. To make this constraint more useful, we use the binomial theorem to write 1 in a suggestive way:

$$1 = (\lambda + 1 - \lambda)^k = \sum_{\ell=0}^{k} \binom{k}{\ell} \lambda^\ell (1-\lambda)^{k-\ell}, \tag{3.21}$$

JHEP04(2018)013

and insert this expression into eq. (3.11), separating out factors where k of the indices are equal to 1:

$$\mathcal{C}_N = \sum_{k=0}^{N} \binom{N}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} \lambda^\ell (1-\lambda)^{k-\ell} E_1^k \times \sum_{i_{k+1}=2}^{M} \cdots \sum_{i_N=2}^{M} C_{\underbrace{1\cdots1}_{k} i_{k+1}\cdots i_N}(\hat{p}_1^\mu, \ldots, \hat{p}_1^\mu, \hat{p}_{i_{k+1}}^\mu, \ldots, \hat{p}_{i_N}^\mu) \prod_{j=k+1}^{N} E_{i_j}. \tag{3.22}$$

Subtracting eq. (3.22) from eq. (3.20) and treating the energies as independent quantities, the following constraint can be read off:

$$\sum_{\ell=0}^{k} \binom{k}{\ell} \lambda^\ell (1-\lambda)^{k-\ell} \left(C_{\underbrace{0\cdots0}_{\ell}\underbrace{1\cdots1}_{k-\ell} i_{k+1}\cdots i_N} - C_{\underbrace{1\cdots1}_{k} i_{k+1}\cdots i_N} \right) = 0, \tag{3.23}$$

where the identical arguments of the angular functions are suppressed for compactness.

We would like to obtain that the quantity in parentheses in eq. (3.23) vanishes since the equation holds for all λ. To see this, suppose that the quantity in parentheses does not vanish, and let $\hat{\ell}$ be the smallest such ℓ where this happens. Consider the regime $0 < \lambda \ll 1$: by the definition of $\hat{\ell}$, there are no $\mathcal{O}(\lambda^\ell)$ terms for $\ell < \hat{\ell}$ and thus the left-hand side of eq. (3.23) is $\mathcal{O}(\lambda^{\hat{\ell}}) \neq 0$, contradicting eq. (3.23). We thus obtain:

$$C_{\underbrace{0\cdots0}_{\ell}\underbrace{1\cdots1}_{k-\ell} i_{k+1}\cdots i_N}(\hat{p}_1^\mu, \ldots, \hat{p}_1^\mu, \hat{p}_{i_{k+1}}^\mu, \ldots, \hat{p}_{i_n}^\mu) = C_{\underbrace{1\cdots1}_{k} i_{k+1}\cdots i_N}(\hat{p}_1^\mu, \ldots, \hat{p}_1^\mu, \hat{p}_{i_{k+1}}^\mu, \ldots, \hat{p}_{i_N}^\mu), \tag{3.24}$$

for $0 \leq \ell \leq k$. Note that in this expression, the first k arguments of the functions are identical.

The constraint in eq. (3.24) is very powerful, especially when combined with the relabeling symmetry of eq. (3.13). While we obtained eq. (3.24) using the collinear limit, the particle direction $\hat{p}_0$ appears nowhere in this expression, so the 0 subscript is simply an index on the angular function. Therefore, when any k arguments of one of the angular functions become collinear, any $\ell \leq k$ of the corresponding subscript labels may be swapped out for values not appearing anywhere else in the indices. A concrete example of this is

$$C_{1123\cdots N-1}(\hat{p}_{i_1}^\mu, \hat{p}_{i_1}^\mu, \hat{p}_{i_2}^\mu, \ldots, \hat{p}_{i_{N-1}}^\mu) = C_{1234\cdots N}(\hat{p}_{i_1}^\mu, \hat{p}_{i_1}^\mu, \hat{p}_{i_2}^\mu, \ldots, \hat{p}_{i_{N-1}}^\mu), \tag{3.25}$$

where the N index here plays the role of the 0 index in eq. (3.24). This then implies that all of the angular functions in eq. (3.15) can related to a single function:

$$C_{i_1\cdots i_N}(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu) = C_{123\cdots N}(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu) \equiv f_N(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu), \tag{3.26}$$

yielding the intuitive result that the angular dependence when some number of particles become collinear should follow from the collinear limit of N arbitrary directions.

JHEP04(2018)013

3.1.4 A new derivation of C-correlators

Finally, substituting eq. (3.26) into eq. (3.11) implies that

$$\mathcal{S} \simeq \sum_{N=0}^{N_{\max}} \mathcal{C}_N^{f_N}, \qquad \mathcal{C}_N^{f_N} = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} \cdots E_{i_N} f_N(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu), \tag{3.27}$$

where we recognize $\mathcal{C}_N^{f_N}$ as the C-correlators of eq. (1.3). This expression says that an arbitrary IRC-safe observable can be approximated arbitrarily well by a linear combination of C-correlators. In this way, we have given a new derivation that C-correlators linearly span the space of IRC-safe observables by directly imposing the constraints of IRC safety and particle relabeling symmetry on an arbitrary observable.

The argument presented here suffices to show the IRC-safety of the C-correlators with any continuous angular weighting function, even if it is not symmetric. Though we used the symmetrization in eq. (3.14) to aid the C-safety derivation in section 3.1.3, it is now perfectly valid to relax this constraint on f_N. In particular, we can simply consider eq. (3.14) applied in reverse and select a single term in the symmetrization sum to represent f_N. Thus we are not constrained merely to symmetric f_N, which will be helpful in obtaining the EFPs.

3.2 Expansion in geometry

Having now established that the C-correlators linearly span the space of IRC-safe observables, we now expand the angular weighting function f_N in eq. (3.27) in terms of a discrete linear angular basis.[6] By virtue of the sum structure of the C-correlators, this angular basis directly translates into a basis of IRC-safe observables, i.e. the energy flow basis.

Following the discussion in section 2.2, we take the angular function f_N to depend only on the pairwise angular distances θ_{ij}. Note that the results of section 3.1 continue to hold with pairwise angular distances in place of particle directions, as long as θ_{ij} is a dimensionless function of $\hat{p}_i^\mu$ and $\hat{p}_j^\mu$ with no residual dependence on energy. Of course, this choice would not be valid for expanding IRC-safe observables that do not respect the symmetries implied by θ_{ij}, such as trying to use the default hadronic measure in eq. (2.4) for observables that depend on the overall jet rapidity. In such cases, one can perform an expansion directly in the $\hat{p}_i^\mu$, though we will not pursue that here.

Expanding the angular function f_N in terms of polynomials up to order $d_{\max}$ in the pairwise angular distances yields:

$$f_N(\hat{p}_{i_1}^\mu, \ldots, \hat{p}_{i_N}^\mu) \simeq \sum_{d=0}^{d_{\max}} \sum_{\mathcal{M} \in \Theta_d} b_{\mathcal{M}} \mathcal{M}, \tag{3.28}$$

where Θ_d is the set of monomials in $\{\theta_{ij} \,|\, i < j \in \{i_1, \ldots, i_N\}\}$ of degree d, $\mathcal{M}$ is one of these monomials, and the $b_{\mathcal{M}}$ are numerical coefficients. While this is a perfectly valid expansion, it represents a vast overcounting of the number of potential angular structures.

[6] Our approach here turns out to be similar to the construction of kinematic polynomial rings for operator bases in ref. [110].

JHEP04(2018)013

Our goal is to substitute eq. (3.28) into the definition of a C-correlator in eq. (3.27) and identify the *unique* analytic structures that emerge. Note that two monomials $\mathcal{M}_1, \mathcal{M}_2 \in \Theta_d$ that are related by a permutation $\sigma \in S_N$ with action $\theta_{i_a i_b} \to \theta_{i_{\sigma(a)} i_{\sigma(b)}}$ give rise to identical C-correlators, $\mathcal{C}^{\mathcal{M}_1} = \mathcal{C}^{\mathcal{M}_2}$, as a result of the relabeling symmetry in section 3.1.2. Thus, we can greatly simplify the angular expansion by summing only over equivalence classes of monomials not related by permutations, which we write as Θ_d/S_N. Writing this out in terms of $\mathcal{E} \in \Theta_d/S_N$:

$$\mathcal{C}_N^{f_N} \simeq \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} \cdots E_{i_N} \sum_{d=0}^{d_{\max}} \sum_{\mathcal{E} \in \Theta_d/S_N} \sum_{\mathcal{M} \in \mathcal{E}} b_{\mathcal{M}} \mathcal{M} \tag{3.29}$$

$$= \sum_{d=0}^{d_{\max}} \sum_{\mathcal{E} \in \Theta_d/S_N} b_{\mathcal{E}} \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} \cdots E_{i_N} \mathcal{M}_{\mathcal{E}}, \tag{3.30}$$

where, by the relabeling symmetry, $\mathcal{M}_{\mathcal{E}}$ can be any representative monomial in the equivalence class $\mathcal{E}$, and the coefficient $b_{\mathcal{E}} = |\mathcal{E}|\, b_{\mathcal{M}}$ absorbs the size $|\mathcal{E}|$ of the equivalence class.

As described in section 2.1, the set of monomials Θ_d is in bijection with the set of multigraphs with d edges and N vertices, and the set of equivalence classes Θ_d/S_N is in bijection with the set of non-isomorphic multigraphs with d edges and N vertices. In particular, each edge (k, ℓ) in a multigraph G corresponds to a factor of $\theta_{i_k i_\ell}$ in the monomial $\mathcal{M}_{\mathcal{E}}$:

$$\mathcal{M}_{\mathcal{E}} = \prod_{(k,\ell) \in G} \theta_{i_k i_\ell}, \tag{3.31}$$

where G corresponds to the equivalence class $\mathcal{E}$. By substituting eq. (3.31) into eq. (3.30) and relabeling the coefficient $b_{\mathcal{E}}$ to b_G, we can identify the resulting analytic structures that linearly span the space of C-correlators as the (unnormalized) EFPs:

$$\mathcal{C}_N^{f_N} \simeq \sum_{d=0}^{d_{\max}} \sum_{G \in \mathcal{G}_{N,d}} b_G \, \mathrm{EFP}_G, \qquad \mathrm{EFP}_G \equiv \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} \cdots E_{i_N} \prod_{(k,\ell) \in G} \theta_{i_k i_\ell}, \tag{3.32}$$

where $\mathcal{G}_{N,d}$ is the set of non-isomorphic multigraphs with d edges on N vertices.

In section 3.1.4, it was shown that the set of IRC-safe observables is linearly spanned by the set of C-correlators, summarized in eq. (3.27). In this section, we have shown in eq. (3.32) that the C-correlators themselves are linearly spanned by the EFPs, whose angular structures are efficiently encoded by multigraphs. By linearity, the EFPs therefore form a complete linear basis for all IRC-safe observables, completing our argument.

4 Computational complexity of the energy flow basis

Since we would like to apply the energy flow basis in the context of jet substructure, the efficient computation of EFPs is of great practical interest. Naively, calculating an EFP whose graph has a large number of vertices requires a prohibitively large amount of computation time, especially as the number of particles in the jet grows large. In practice, though, we can dramatically speed up the implementation of the EFPs by making use of the correspondence with multigraphs. Beta code to calculate the EFPs using these methods is available through our `EnergyFlow` module.

JHEP04(2018)013

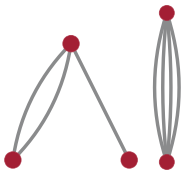
4.1 Algebraic structure

The set of EFPs has a rich algebraic structure which will allow in some cases for faster computation. Firstly, they form a monoid (a group without inverses) under multiplication. In analogy with the natural numbers, the composite EFPs, those with disconnected multigraphs, can be expressed as a product of the prime EFPs corresponding to the connected components of a disconnected graph:

$$\mathrm{EFP}_G = \prod_{g \in C(G)} \mathrm{EFP}_g, \tag{4.1}$$

where $C(G)$ is the set of connected components of the multigraph G.

As a concrete example of eq. (4.1), consider:



$$= \left(\sum_{i_1=1}^{M} \sum_{i_1=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta^2_{i_1 i_2} \theta_{i_2 i_3} \right) \left(\sum_{i_4=1}^{M} \sum_{i_5=1}^{M} z_{i_4} z_{i_5} \theta^4_{i_4 i_5} \right). \tag{4.2}$$

Thus, we only need to perform summations for the computation of prime EFPs, with the composite ones given by eq. (4.1). Note that if one were combining EFPs with a nonlinear method, such as a neural network, the composite EFPs would not be needed as separate inputs since the model could in principle learn to compute them on its own. The composite EFPs are, however, required to have a linear basis and should be included when linear methods are employed, such as those in sections 5 and 6.

The relationship between prime and composite EFPs is just the simplest example of the algebraic structure of the energy flow basis. The EFPs depend on M energies and $\binom{M}{2}$ pairwise angles, but there are only $3M - 4$ degrees of freedom for the phase space of M massless particles, leading generically to additional (linear) relations among the EFPs. Hence, the EFPs are an *over*complete linear basis. We leave further analysis and exploration of these relations to future work, and simply remark here that linear methods continue to work even if there are redundancies in the basis elements.

4.2 Dispelling the $\mathcal{O}(M^N)$ myth for N-particle correlators

It is useful to analyze the complexity of computing an EFP.[7] A naive implementation of eq. (1.1) runs in $\mathcal{O}(M^N)$ due to the N nested sums over M particles. There is a computational simplification, however, that can be used to tremendously speed up calculations of certain EFPs by making use of the graph structure of G. As an example, consider the following EFP:

$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_1 i_3} \theta_{i_1 i_4} = \sum_{i_1=1}^{M} z_{i_1} \left(\sum_{i_2=1}^{M} z_{i_2} \theta_{i_1 i_2} \right)^3, \tag{4.3}$$

[7] The title of this section is inspired by ref. [111].

JHEP04(2018)013

d			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prime	χ	2	1	2	4	9	21	55	146	415	1 212	3 653
		3			1	3	12	47	185	757	3 181	13 691
		4						1	2	11	49	231
		5										1
	Total		1	2	5	12	33	103	333	1 183	4 442	17 576
All	χ	2	1	3	7	19	48	135	371	1 077	3 161	9 539
		3			1	4	18	76	312	1 296	5 447	23 268
		4						1	3	16	74	352
		5										1
	Total		1	3	8	23	66	212	686	2 389	8 682	33 160

(a)

		N														
d	χ	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	2	1														
2	2	1	1	1												
3	2	1	1	3	1	1										
	3		1													
4	2	1	2	5	5	4	1	1								
	3		1	2	1											
5	2	1	2	8	10	14	7	4	1	1						
	3		2	5	7	3	1									
6	2	1	3	12	21	33	30	21	8	4	1	1				
	3		3	12	23	23	11	3	1							
	4			1												
7	2	1	3	16	35	71	82	81	45	23	8	4	1	1		
	3		4	23	65	92	76	36	12	3	1					
	4			1	1	1										
8	2	1	4	21	58	134	205	245	197	122	52	24	8	4	1	1
	3		5	41	153	311	355	257	118	40	12	3	1			
	4			3	5	5	2	1								

(b)

Table 4. (a) The number of prime/all EFPs binned by degree d and complexity χ up to $d = 10$. The complexity is that of our `EnergyFlow` implementation, running in time $\mathcal{O}(M^\chi)$. The partial sums of the "Total" rows are the entries of table 2a. (b) The number of EFPs binned by degree d, complexity χ, and N up to $d = 8$. Note that the majority of EFPs shown here have $N > 4$, which would be computationally intractable without algorithmic speedups such as VE.

JHEP04(2018)013

which can be computed in $\mathcal{O}(M^2)$ rather than $\mathcal{O}(M^4)$ by first computing the M objects in parentheses in eq. (4.3) and then performing the overall sum.

In general, since the summand is a product of factors, the distributive property allows one to put parentheses around combinations of sum operators and factors. A clever choice of such parentheses, known as an *elimination ordering*, can often be used to perform the N sums of eq. (1.1) in a way which greatly reduces the number of operations needed to obtain the value of the EFP for a given set of particles. This technique is known as the Variable Elimination (VE) algorithm [77] (see also ref. [112] for a review).

When run optimally, the VE algorithm reduces the complexity of computing EFP_G to $\mathcal{O}(M^{\text{tw}(G)+1})$ where $\text{tw}(G)$ is the *treewidth* of the graph G, neglecting multiple edges in the case of multigraphs. The treewidth is a measure which captures how tangled a graph is, with trees (graphs with no cycles) being the least tangled (with treewidth 1) and complete graphs the most tangled (with treewidth $N-1$). Additionally, we have that for graphs with a single cycle the treewidth is 2 and for complete graphs minus one edge the treewidth is $N-2$. Thus the EFPs corresponding to tree multigraphs can be computed with VE in $\mathcal{O}(M^2)$ whereas complete graphs do require the naive $\mathcal{O}(M^N)$ to compute with VE. Since the ECFs correspond to complete graphs (see eq. (2.9)), they do not benefit from VE. Similarly, VE cannot speed up the computation of ECFGs, since the ECFGs do not have a factorable summand.

Finding the optimal elimination ordering and computing the treewidth for a graph G are both NP-hard. In practice, heuristics are used to decide on a pretty-good elimination ordering (which for the small graphs we consider here is often optimal) and to approximate the treewidth. In principle, these orderings need only be computed once for a fixed set of graphs of interest. Similarly, many algebraic structures reappear when computing a set of EFPs for the same set of particles, making dynamic programming a viable technique for further improving the computational complexity of the method.

Table 4a shows the number of EFPs listed by degree d and VE complexity χ (with respect to the heuristics used in our implementation), and table 4b further breaks up the EFP counts by N. Figure 1 shows the time to compute the average $d \leq 7$ EFP as a function of multiplicity M for different VE complexity χ. Finally, we note that though VE often provides a significant speedup over the naive algorithm, there may be even faster ways of computing the EFPs.[8]

5 Linear regression with jet observables

Regression, classification, and generation are three dominant machine learning paradigms. Machine learning applications in collider physics have been largely focused on classification (e.g. jet tagging) [65–73] with recent developments in regression [113] and generation [114, 115]. For a more complete review of modern machine learning techniques in jet substructure, see ref. [48]. The lack of established regression problems in jet physics is due

[8] At the risk of burying the lede in a footnote, we have found that with certain choices of the angular measure, it is possible to compute all EFPs in $\mathcal{O}(M)$. We leave a further exploration of these interesting special cases to future work.

JHEP04(2018)013

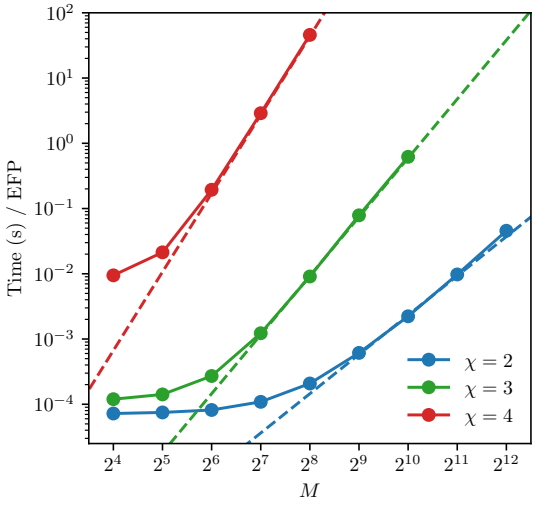


Figure 1. Compute time (in seconds) per EFP for different VE complexities χ as a function of the number of inputs M. The quoted value is based on all EFPs with $d \leq 7$, and each data point is the average of 10 computations. The dashed lines show the expected $\mathcal{O}(M^\chi)$ scaling behavior. As χ increases, the relative amount of overhead decreases and the asymptotic behavior is achieved more rapidly than for smaller χ. Computations were run with Python 3.5.2 and `NumPy` 1.13.3 on a 2.3 GHz Intel Xeon E5-2673 v4 (Broadwell) processor on Microsoft Azure using our `EnergyFlow` module.

in part to the difficulty of theoretically probing multivariate combinations as well as the challenges associated with extracting physics information from trained regressions models.

In this section, we show that the linearity of the energy flow basis mitigates many of these problems, providing a natural regression framework using simple linear models, probing the learned observable combinations, and gaining insight into the physics of the target observables. Since regression requires training samples, we observe how the regression performance compares on jets with three characteristic phase-space configurations: one-prong QCD jets, two-prong boosted W jets, and three-prong boosted top jets. We use linear regression to demonstrate convergence of the energy flow basis on IRC-safe observables, while illustrating their less-performant behavior for non-IRC-safe observables.

5.1 Linear models with the energy flow basis

Linear models assume a linear relationship between the input and target variables, making them the natural choice for (machine) learning with the energy flow basis for both regression and classification. A linear model $\mathscr{M}$ with EFPs as the inputs is defined by a finite set $\mathcal{G}$ of multigraphs and numerical coefficients $\mathbf{w} = \{w_G\}_{G\in\mathcal{G}}$:

$$\mathscr{M} = \sum_{G\in\mathcal{G}} w_G \,\mathrm{EFP}_G. \tag{5.1}$$

The fundamental relationship between EFPs, linear models, and IRC-safe observables is highlighted by comparing eq. (5.1) to eq. (1.5), where the linear model $\mathscr{M}$ in eq. (5.1) takes the place of the IRC-safe observable $\mathcal{S}$ in eq. (1.5). Because the EFPs are a complete linear basis, $\mathscr{M}$ is capable of approximating any $\mathcal{S}$ for a sufficiently large set of EFPs.

JHEP04(2018)013

The linear structure of eq. (5.1) allows for an avenue to "open the box" and interpret the learned coefficients as defining a unique multiparticle correlator for each N. To see this, partition the set $\mathcal{G}$ into subsets $\mathcal{G}_N$ of graphs with N vertices. The sum in eq. (5.1) can be broken into two sums, one over N and the other over all graphs in $\mathcal{G}_N$. The linear energy structure of the EFPs in eq. (1.1) allows for the second sum to be pushed inside the product of energies onto the angular weighting function:

$$\mathscr{M} = \sum_{N=0}^{N_{\max}} \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \left(\sum_{G \in \mathcal{G}_N} w_G \prod_{(k,\ell) \in G} \theta_{i_k i_\ell} \right), \tag{5.2}$$

where $N_{\max}$ is the maximum number of vertices of any graph in $\mathcal{G}$. The quantity in parentheses in eq. (5.2) may be though of as a *single* angular weighting function. The linear model written in this way reveals itself to be a sum of C-correlators (similar to eq. (3.27)), one for each N, where the linear coefficients within each $\mathcal{G}_N$ parameterize the angular weighting function f_N of that C-correlator. This arrangement of the learned parameters of the linear model into $N_{\max}$ C-correlators contrasts sharply with the lack of a physical organization of parameters in nonlinear methods such as neural networks or boosted decision trees.

5.2 Event generation and EFP computation

For the studies in this section and in section 6, we generate events using Pythia 8.226 [116–118] with the default tunings and shower parameters at $\sqrt{s} = 13\,\text{TeV}$. Hadronization and multiple parton interactions (i.e. underlying event) are included, and a 400 GeV parton-level p_T cut is applied. For quark/gluon distribution, quark (signal) jets are generated through $pp \to qZ(\to \nu\bar{\nu})$, and gluon (background) jets through $pp \to gZ(\to \nu\bar{\nu})$, where only light-quarks (uds) appear in the quark sample. For W and top tagging, signal jets are generated through $pp \to W^+W^-(\to \text{hadrons})$ and $pp \to t\bar{t}(\to \text{hadrons})$, respectively. For both W and top events, the background consists of QCD dijets.

Final state, non-neutrino particles were made massless, keeping y, ϕ, and p_T fixed,[9] and then were clustered with FastJet 3.3.0 [119] using the anti-k_T algorithm [120] with a jet radius of $R = 0.4$ for quark/gluon samples and $R = 0.8$ for W and top samples (and the relevant dijet background). The hardest jet with rapidity $|y| < 1.7$ and $500\,\text{GeV} \leq p_T \leq 550\,\text{GeV}$ was kept. For each type of sample, 200k jets were generated. For the regression models, 75% were used for training and 25% for testing.

For these events, all EFPs up to degree $d \leq 7$ were computed in Python using our `EnergyFlow` module making use of `NumPy`'s einsum function. See tables 2 and 4 for counts of EFPs tabulated by various properties such as N, d, and χ. Note that all but 4 of the 1000 $d \leq 7$ EFPs can be computed in $\mathcal{O}(M^2)$ or $\mathcal{O}(M^3)$ in the VE paradigm, making the set of EFPs with $d \leq 7$ efficient to compute.

5.3 Spanning substructure observables with linear regression

We now consider the specific case of training linear models to approximate substructure observables with linear combinations of EFPs. For an arbitrary observable O, we use

[9] Using massless inputs is not a requirement for using the EFPs, but for these initial EFP studies, we wanted to avoid the caveats associated with massive inputs for the validity of section 3.

JHEP04(2018)013

	Observable	Properties
$m_J/p_{T,J}$	Scaled jet mass	No Taylor expansion about zero energy limit
$\lambda^{(\alpha=1/2)}$	Les Houches Angularity	No analytic relationship beyond even integer α
$\tau_2^{(\beta=1)}$	2-subjettiness	Algorithmically defined IRC-safe observable
$\tau_{21}^{(\beta=1)}$	N-subjettiness ratio	Sudakov safe, safe for two-prong kinematics
$\tau_{32}^{(\beta=1)}$	N-subjettiness ratio	Sudakov safe, safe for three-prong kinematics
M	Particle multiplicity	IRC unsafe

Table 5. The six substructure observables used as targets for linear regression, listed with relevant properties. The first three are IRC safe, the next two are Sudakov safe in general (and IRC safe in the noted regions of phase space), and particle multiplicity is IRC unsafe. The Les Houches Angularity [124, 125] is calculated with respect to the p_T-weighted centroid axis in eq. (2.12), and the N-subjettiness observables [54, 55] are calculated using k_T axes.

least-squares regression to find a suitable set of coefficients $\mathbf{w}^*$:

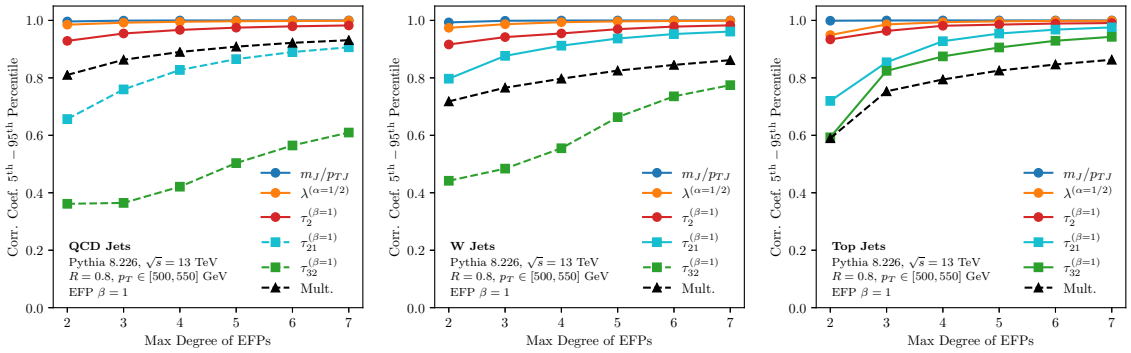
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left\{ \sum_{J \in \text{jets}} \left(O(J) - \sum_{G \in \mathcal{G}} w_G \, \text{EFP}_G(J) \right)^2 \right\}, \tag{5.3}$$

where $O(J)$ is the value of the observable and $\text{EFP}_G(J)$ the value of the EFP given by multigraph G on jet J. There are possible modifications to eq. (5.3) which introduce penalties proportional to $\|\mathbf{w}\|_1$ or $\|\mathbf{w}\|_2^2$ where $\|\cdot\|_1$ is the 1-norm and $\|\cdot\|_2$ is the 2-norm. The first of these choices, referred to as lasso regression [121], may be particularly interesting because of the variable selection behavior of this model, which would aid in selecting the most important EFPs to approximate a particular observable. We leave such investigation to future work. See ref. [122] for a review of linear models for regression.

We use the `LinearRegression` class of the `scikit-learn` python module [123] to implement eq. (5.3) with no regularization on the samples described in section 5.2. In general, the smallest possible regularization which prevents overfitting (if any) should be used. Because of the linear nature of linear regression and the analytic tractability of eq. (5.3), the $\mathbf{w}^*$ corresponding to the global minimum of the squared loss function can be found efficiently using convex optimization techniques. Such techniques include closed-form solutions or convergent iterative methods.

As targets for the regression, we consider the six jet observables in table 5 to highlight some interesting test cases. As our measure of the success of the regression, we use a variant of the correlation coefficient between the true and predicted observables that is less sensitive to outliers than the unadulterated correlation coefficient. When evaluating the trained linear model on the test set, only test samples with predicted values within the 5^{th} and 95^{th} percentiles of the predictions are included. In the contexts considered in this paper, narrowing this percentile range lowers the correlation coefficient and widening the range out toward all of the test set increases the correlation coefficient. The qualitative nature of the results are insensitive to the specific choice of percentile cutoffs. We perform

JHEP04(2018)013



(a) **(b)** **(c)**

Figure 2. Correlation coefficients between true and predicted values for the jet observables in table 5, plotted as a function of maximum EFP degree. Shown are the (a) QCD dijet, (b) W jet, and (c) top jet samples, and as explained in the text, we restrict to predictions in the 5^{th}–95^{th} percentiles. Observables in IRC-safe regions of phase space are shown with solid lines and those in IRC-unsafe regions (including Sudakov-safe regions) are shown with dashed lines. The IRC-safe observables are all learned with correlation coefficient above 0.98 in all three cases by $d = 7$. Multiplicity (black triangles) sets the scale for the regression performance on IRC-unsafe observables. Note that τ_{21} has performance similar to the IRC-safe observables only when jets are characteristically two-pronged or higher (W and top jets), and similarly for τ_{32} when the jets are characteristically three-pronged (top jets).

this regression using EFPs of degree up to d for d from 2 to 7 on all three jet samples, with the results shown in figure 2. Histograms of the true and predicted distributions for a subset of these observables are shown in figure 3 for the three types of jets considered here.

Since the learned coefficients depend on the training set, in principle different linear combinations may be learned to approximate the substructure observables in different jet contexts. This stands in contrast to the analysis in section 2.3, where many jet substructure observables were identified as exact linear combinations of EFPs, independent of the choice of inputs. The IRC-safe observables — mass, Les Houches angularity, and 2-subjettiness — are all learned with a correlation coefficient above 0.98 in all three cases by $d = 7$.

The IRC-unsafe multiplicity sets the scale of performance for observables that are not IRC safe. For the N-subjettiness ratios, the regression performance depends on whether the observable is IRC safe or only Sudakov safe [82, 83]. The ratio τ_{21} is only IRC safe for regions of phase space with two prongs or more (i.e. the W and top samples), and τ_{32} is only IRC safe for three prongs or more (i.e. just the top sample). In cases where the N-subjettiness ratio is IRC safe, the regression performs similarly to the other IRC-safe observables, whereas for the cases where the N-subjettiness ratio is only Sudakov safe, the regression performance is poor (even worse than for multiplicity). It is satisfying to see the expected behavior between the safety of the observable and the quality of the regression with EFPs.

As a final cross check of the regression, we can use the linear model in eq. (5.1) to confirm some of the analytic results of section 2.3. Specifically, we perform a linear

JHEP04(2018)013

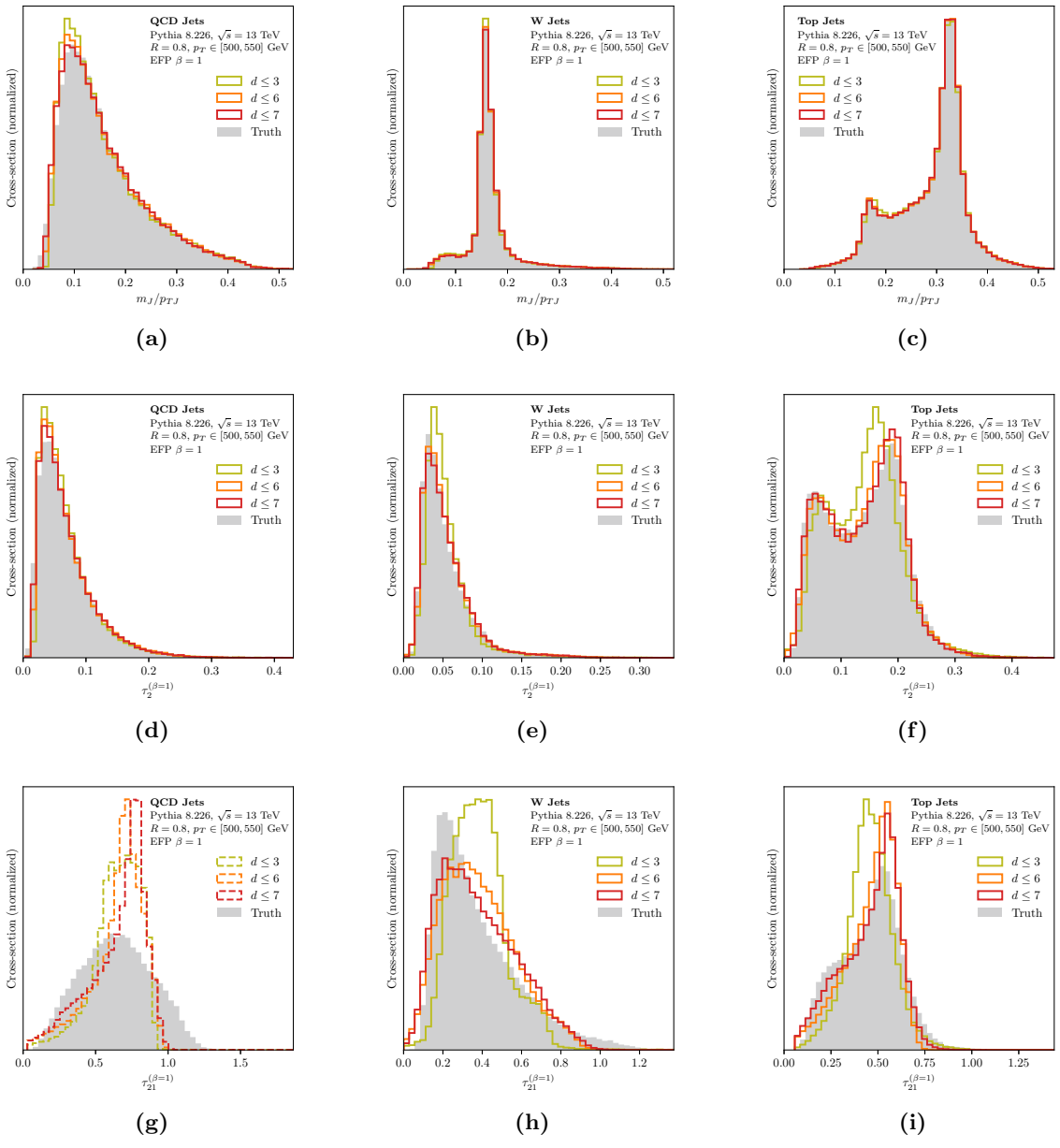
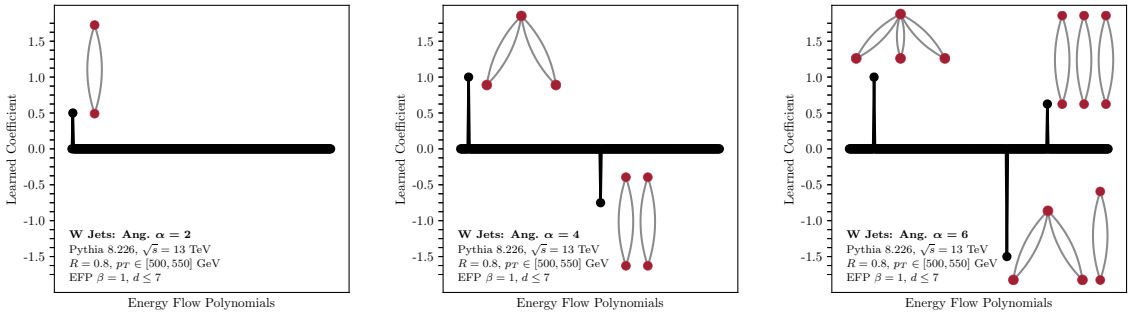


Figure 3. The distributions of true and predicted scaled jet mass (top), $\tau_2^{(\beta=1)}$ (middle), and $\tau_{21}^{(\beta=1)}$ (bottom) using linear regression with EFPs up to different maximum degrees d on QCD jets (left), W jets (center), and top jets (right). Note the excellent agreement for the IRC-safe observables in the first two rows. Observables in IRC-safe regions of phase space are shown with solid lines and those in IRC-unsafe regions are shown with dashed lines. The Sudakov-safe $\tau_{21}^{(\beta=1)}$ predicted distributions match the true distributions for jets typically with two or more prongs (W and top jets) better than for typically one-pronged (QCD) jets.

JHEP04(2018)013



(a) **(b)** **(c)**

Figure 4. The linear combinations of EFPs learned by linear regression for even-α angularities with the W jet samples. Shown are (a) $\alpha = 2$, (b) $\alpha = 4$, and (c) $\alpha = 6$. All but the highlighted EFP coefficients are learned to be near zero. The EFPs corresponding to those non-zero coefficients are illustrated directly on the figure. The learned linear coefficients are exactly those predicted analytically in eqs. (2.14), (2.15) and (2.16). The same behavior is found with the QCD and top jet samples.

regression with the target observable being the even-α angularities with respect to the p_T-weighted centroid axis. These were shown to be non-trivial linear combinations of EFPs in section 2.3.3. Regressing onto $\lambda^{(2)}$, $\lambda^{(4)}$, and $\lambda^{(6)}$, the linear model learned the observables with effectively 100% accuracy and the learned linear combination was exactly that predicted by eqs. (2.14), (2.15) and (2.16), up to a precision of 10^{-6}. Figure 4 shows the learned linear combinations of EFPs for the W jet sample.

6 Linear jet tagging

We now apply the energy flow basis to three representative jet tagging problems — light-quark/gluon classification, W tagging, and top tagging — providing a broad set of contexts in which to study the EFPs. Since the energy flow basis is linear, we can (in principle) access the optimal IRC-safe observable for jet tagging by training a linear classifier for this problem. As mentioned in section 5.3, one benefit of linear models, in addition to their inherent simplicity, is that they are typically convex problems which can be solved exactly or with gradient descent to a global minimum. See ref. [122] for a review of linear models for classification.

A (binary) linear classifier learns a vector $\mathbf{w}^*$ that defines a hyperplane orthogonal to the vector. A bias term, which can be related to the distance of this hyperplane from the origin, sets the location of the decision boundary, which is the hyperplane translated away from the origin. The decision function for a particular point in the input space is the normal distance to the decision boundary. In contrast with regression, where the target variable is usually continuous, classification predictions are classes, typically 0 or 1 for a binary classifier.

Different methods of determining the vector $\mathbf{w}^*$ — such as logistic regression, support vector machines, or linear discriminant analysis — may learn different linear classifiers since the methods optimize different loss functions. For our linear classifier, we use Fisher's linear discriminant [126] provided by the `LinearDiscriminantAnalysis` class of

JHEP04(2018)013

the `scikit-learn` python module [123]. The choice of logistic regression was also explored, and jet tagging performance was found to be insensitive to which type of linear classifier was used.

The details of the event generation and EFP computation are the same as in section 5.2. To avoid a proliferation of plots, we present only the case of W tagging in the text and refer to appendix B for the corresponding results for quark/gluon discrimination and top tagging. Qualitatively similar results are obtained on all three tagging problems, with the conclusion that linear classification with EFPs yields comparable classification performance to other powerful machine learning techniques. This is good evidence that the EFPs provides a suitable linear expansion of generic IRC-safe information relevant for practical jet substructure applications.

6.1 Alternative jet representations

In order to benchmark the EFPs, we compare them to two alternative jet tagging paradigms:

- The **jet images** approach [50] treats calorimeter deposits as pixels and the jet as an image, often using convolutional neural networks to determine a classifier. Jet images have been applied successfully to the same tagging problems considered here: quark/gluon discrimination [65], W tagging [63], and top tagging [66, 68].

- The ***N*-subjettiness basis** was introduced for W tagging in ref. [57] and later applied to tagging non-QCD jets [73]. We use the same choice of N-subjettiness basis elements as ref. [57], namely:

$$\{\tau_1^{(1/2)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(1/2)}, \tau_2^{(1)}, \tau_2^{(2)}, \cdots, \tau_{N-2}^{(1/2)}, \tau_{N-2}^{(1)}, \tau_{N-2}^{(2)}, \tau_{N-1}^{(1)}, \tau_{N-1}^{(2)}\}, \tag{6.1}$$

  with $3N-4$ elements needed to probe N-body phase space. These are then used as inputs to a DNN.

Both of these learning paradigms are expected to perform well, and we will see below that this is the case. As a strawman, we also consider linear classification with the N-subjettiness basis elements in eq. (6.1), which is not expected to yield good performance. For completeness, we also perform DNN classification with the energy flow basis.

We now summarize the technical details of these alternative jet tagging approaches. For jet images, we create 33×33 jet images spanning $2R \times 2R$ in the rapidity-azimuth plane. Motivated by ref. [65], both single-channel "grayscale" jet images of the p_T per pixel and two-channel "color" jet images consisting of the p_T channel and particle multiplicity per pixel were used. The p_T-channel of the jet image was normalized such that the sum of the pixels was one. Standardization was used to ensure that each pixel had zero mean and unit standard deviation by subtracting the training set mean and dividing by the training set standard deviation of each pixel in each channel. A jet image CNN architecture similar to that used in ref. [65] was employed: three 36-filter convolutional layers with filter sizes of 8×8, 4×4, and 4×4, respectively, followed by a 128-unit dense layer and a 2-unit softmaxed output. A rectified linear unit (ReLU) activation [127] was applied to the output

JHEP04(2018)013

of each internal layer. Maxpooling of size 2×2 was performed after each convolutional layer with a stride length of 2. The dropout rate was taken to be 0.1 for all layers. He-uniform initialization [128] was used to initialize the model weights.

For the DNN (both for the N-subjettiness basis and for the EFPs), we use an architecture consisting of three dense layers of 100 units each connected to a 2-unit softmax output layer, with ReLU activation functions applied to the output of each internal layer. For the training of all networks, 300k samples were used for training, 50k for validation, and 50k for testing. Networks were trained using the Adam algorithm [129] using categorical cross-entropy as a loss function with a learning rate of 10^{-3} and a batch size of 100 over a maximum of 50 epochs. Early stopping was employed, monitoring the validation loss, with a patience parameter of 5. The python deep learning library `Keras` [130] with the `Theano` backend [131] was used to instantiate and train all neural networks. Training of the CNNs was performed on Microsoft Azure using NVIDIA Tesla K80 GPUs and the NVIDIA CUDA framework. Neural network performance was checked to be mildly insensitive to these parameter choices, but these parameter choices were not tuned for optimality. As a general rule, the neural networks used here are employed to give a sense of scale for the performance attainable with jet images and the N-subjettiness basis using out-of-the-box techniques; improvements in classification accuracy may be possible for these methods with additional hyperparameter tuning.

6.2 W tagging results and comparisons

We present results for the W tagging study here, with the other two classification problems discussed in appendix B. The performance of a binary classifier is encapsulated by the background mistag rate ε_b at a given signal efficiency ε_s. For all of the figures below, we plot inverse receiver operator characteristic (ROC) curves, $1/\varepsilon_b$ as a function of ε_s, on a semi-log scale; a higher ROC curve indicates a better classifier. The corresponding standard ROC (ε_b vs. ε_s) and significance improvement ($\varepsilon_s/\sqrt{\varepsilon_b}$ vs. ε_s) curves are available in the source files of the `arXiv` preprint as additional pages in the figure.

We begin by studying the performance for different choices of angular exponent β in the default hadronic measure from eq. (2.4). Figure 5 shows ROC curves for the choices of $\beta = 0.2$, $\beta = 0.5$, and $\beta = 1$, using all EFPs with $d \leq 7$. The differences in performance are mild, but $\beta = 0.5$ slightly improves the ROC curves for W tagging, so we use $\beta = 0.5$ for the remainder of our studies. The choice of $\beta = 0.5$ was also found to be optimal for the cases of quark/gluon and top tagging discussed in appendix B.

Next, in figure 6a, we test the linear spanning nature of the EFPs by comparing the ROC curves of the linear and nonlinear models trained on EFPs up to different d. With linear regression, there is a large jump in performance in going from $d \leq 3$ (13 EFPs) to $d \leq 6$ (314 EFPs), and a slight increase in performance from $d \leq 6$ to $d \leq 7$ (1000 EFPs), indicating good convergence to the optimal IRC-safe observable for W jet discrimination. To avoid cluttering the plot, $d \leq 4$ and $d \leq 5$ are not shown in figure 6a, but their ROC curves fall between those of $d \leq 3$ and $d \leq 6$, highlighting that the higher d EFPs carry essential information for linear classification. By contrast, using nonlinear classification with a DNN, the EFPs performance with $d \leq 3$ is already very good, since functions of the

JHEP04(2018)013

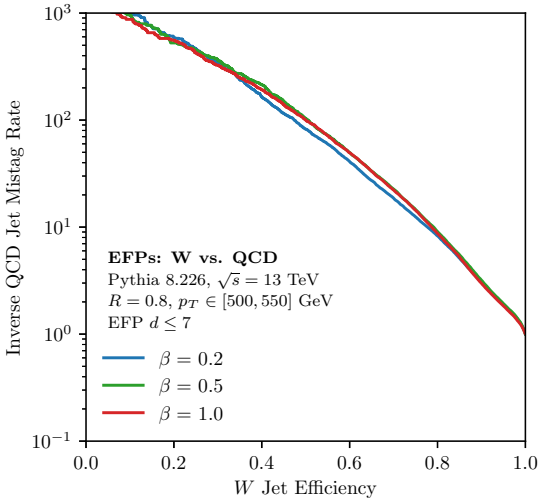
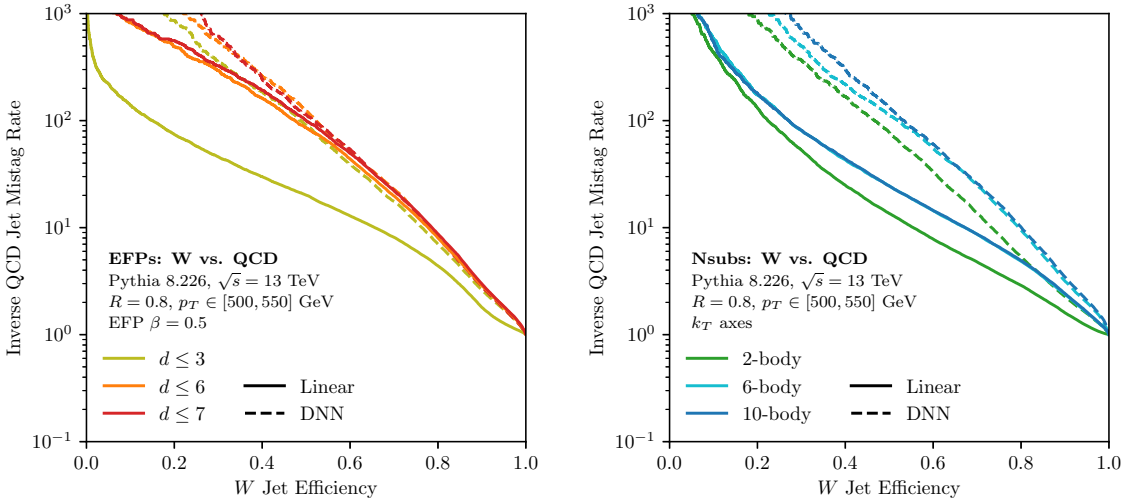


Figure 5. Inverse ROC curves for linear W tagging with the energy flow basis using different choices of angular exponent β in eq. (2.4). Though the improvement is mild, $\beta = 0.5$ shows the best overall performance. See figure 10 for the corresponding quark/gluon discrimination and top tagging results, where $\beta = 0.5$ is also the best choice by a slight margin.



(a) **(b)**

Figure 6. Inverse ROC curves for W tagging with (a) the energy flow basis including degrees up to $d = 7$ and (b) the N-subjettiness basis up to 10-body phase space information. In both cases, we show the observables combined linearly (solid) and with a DNN (dashed). The linear combinations of EFPs can be seen to approach the nonlinear combinations, particularly for higher signal efficiencies, while the linear combinations of the N-subjettiness basis can be seen to saturate well below the nonlinear combinations as the number of observables is increased. See figure 11 for the corresponding quark/gluon discrimination and top tagging results.

JHEP04(2018)013

low d EFPs can be combined in a nonlinear fashion to construct information contained in higher d composite EFPs. The linear and nonlinear performance is similar with the $d \leq 7$ EFPs for operating points of $\varepsilon_s \gtrsim 0.5$, though the nonlinear DNN outperforms the linear classifier in the low signal efficiency region. It should be noted that the linear classifier is not trained specifically for the low signal efficiency region and it may be possible that choosing a different hyperplane could boost performance there. We leave to future work a more detailed investigation of optimizing the choice of linear classifier.

The performance of the N-subjettiness basis with both linear and nonlinear classifiers is shown in figure 6b. For both linear classification and the DNN, performance appears to saturate with the 6-body (14 τ_Ns) phase space, with not much gained in going to 10-body (26 τ_Ns) phase space, except for a small increase in the low signal efficiency region for the DNN; we confirmed up to 30-body (86 τ_Ns) phase space that no change in ROC curves was observed compared to 10-body phase space. As expected, there is relativity poor performance with linear classification even as the dimension of phase space is increased. Classification with a DNN, though, shows an immense increase in performance over linear classification, as expected since the N-subjettiness basis is expected to nonlinearly capture all of the relevant IRC-safe kinematic information [57]. This illustrates that nonlinear combinations of the N-subjettiness observables are crucial for extracting the full physics information.

The corresponding quark/gluon and top tagging plots in figure 11 effectively tell the same story as figure 6, robustly demonstrating the linear spanning nature of the EFPs used for classification across a wide variety of kinematic configurations. As a side note, in appendix B there are sometimes cases where a linear combination of EFPs yields *improved* performance compared to a DNN on the same inputs, particularly at medium to high signal efficiencies. Since even a one-node DNN should theoretically be able to learn the linear combination of EFPs learned by the linear classifier, regimes where the linear classifier outperforms the DNN demonstrate the inherent difficulty of training complex multivariate models.

In figure 7 we directly compare the EFP classification power against the N-subjettiness basis and the 1-channel ("grayscale") and 2-channel ("color") CNNs. For operating points with $\varepsilon_s \gtrsim 0.5$, all methods except the linear N-subjettiness classifier show essentially the same performance. The worse performance of the linear EFP classifier at low signal efficiencies is expected, since the Fisher linear discriminant is not optimized for that regime. Overall, it is remarkable that similar classification performance can be achieved with these three very different learning paradigms, especially considering that the DNNs and grayscale CNN implicitly, and the color CNN explicitly, have access to non-IRC-safe information (including Sudakov-safe combinations of the IRC-safe inputs). This agreement gives evidence that the tagging techniques have approached a global bound on the maximum possible discrimination power achievable, at least in the context of parton shower simulations.

Once again, the analogous quark/gluon and top tagging plots, shown in figure 12, show very similar behavior to the W tagging case in figure 7. Linear classification with the EFPs performs similarly to the DNNs and CNNs, tending to slightly outperform at high signal efficiencies and underperform at low signal efficiencies. Ultimately, the choice of tagging method comes down to a trade off between the simplicity of the inputs and the simplicity of the training method, with the EFPs presently requiring more inputs than the

JHEP04(2018)013

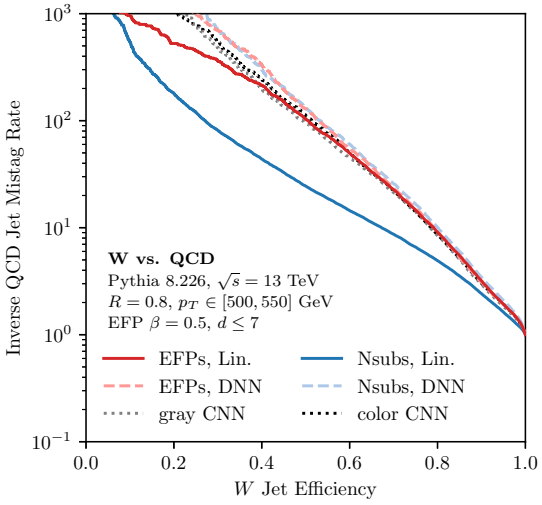


Figure 7. Inverse ROC curves for W tagging comparing six different methods: linear and DNN classification with the energy flow basis up to $d \leq 7$, linear and DNN classification with the N-subjettiness basis up to 10-body phase space, and grayscale and color jet images with CNNs. The most evident gap is between the linearly-combined N-subjettiness basis and the remaining curves, which achieve similar classification performance for medium and high signal efficiencies. See figure 12 for the corresponding quark/gluon discrimination and top tagging results.

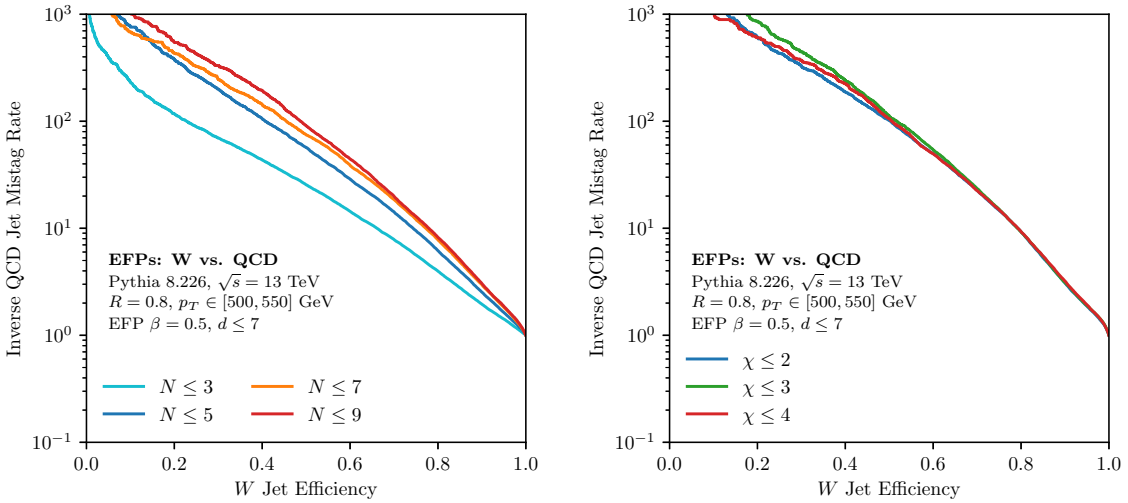
N-subjettiness basis but with the benefit of using a linear model. In the future, we plan to study ways of reducing the size of the EFP basis by exploiting linear redundancies among the EFPs and using powerful linear methods to automatically select the most important observables for a given task.

6.3 Opening the energy flow box

As argued in eq. (5.2), one of the main advantages of linear methods with the energy flow basis is that one can attempt to "open the box" and directly explore what features have been learned. We leave to future work a full exploration of this possibility, but here we attempt to probe which topological structures within the EFP basis carry the discrimination power for the different tagging problems. Since we have shown that the EFPs with $d \leq 7$ have sufficient discrimination power to qualitatively match the performance of alternative tagging methods, we will restrict to this set of observables.

In figure 8a, we vary the maximum number of vertices in the EFP graphs, where the maximum N is 14 for $d \leq 7$, finding that the performance roughly saturates at $N = 9$, highlighting the importance of higher N EFPs. The algorithmic advances described in section 4 allow for the efficient computation of these higher N EFPs, which have complexities as intractable as $\mathcal{O}(M^9)$ with the naive algorithm. Additionally, note that nearly every EFP (all except those corresponding to complete graphs) has a non-vanishing angular weighting function, which is a new feature compared to the ECFs and ECFGs (see section 2.4). In figure 8b, we vary the maximum computational complexity χ of the EFP graphs, where the maximum χ is 4 for $d \leq 7$. Remarkably, the full performance of linear

JHEP04(2018)013



(a) **(b)**

Figure 8. Inverse ROC curves for linear W tagging with the energy flow basis with $d \leq 7$, sweeping over (a) which N-point correlators and (b) observables of which VE computational complexity $\mathcal{O}(M^\chi)$ are included in the linear fit. It is clear that important information is contained in the higher N-particle correlators, which can be included because the algorithm in section 4 evades the naive $\mathcal{O}(M^N)$ scaling. Interestingly, the discrimination power appears to be almost saturated by the graphs computable in $\mathcal{O}(M^2)$. See figure 13 for the corresponding quark/gluon discrimination and top tagging results.

classification with the $d \leq 7$ EFPs can be obtained with merely those observables calculable in $\mathcal{O}(M^2)$ with VE. Thus, fortuitously for the purposes of jet tagging, it seems that restricting to the most efficiently computable EFPs (in the VE paradigm) is sufficient for extracting the near-optimal IRC-safe observable for jet discrimination. Similar results hold for quark/gluon discrimination and top tagging, shown in figure 13.

7 Conclusions

In this paper, we have introduced the EFPs, which linearly span the space of IRC-safe observables.[10] The core argument, presented in section 3, is that one can systematically expand an arbitrary IRC-safe observable in terms of energies and angles and read off the unique resulting analytic structures. This expansion yields a new way to understand the importance of C-correlators [74, 87–89] for IRC safety, and it enables a powerful graph-theoretic representation of the various angular structures. The multigraph correspondence makes manifest a more efficient algorithm than the naive $\mathcal{O}(M^N)$ one for computing EFPs, overcoming a primary obstacle to exploring higher-N multiparticle correlators for jet substructure.

[10] In the course of this research, we encountered a more descriptive acronym than "EFPs" albeit with an unintended biblical reference: polynomials of Energies and Angles Result in a Linear Spanning Basis for Energy Flow Observables Relevant for Extracting Substructure With Improved Nuance and Efficiency.

JHEP04(2018)013

To demonstrate the power of the energy flow basis, we performed a variety of representative regression and classification tasks for jet substructure. Crucially, linear methods were sufficient to achieve good performance with the EFPs. As a not-quite apples-to-apples comparison in three representative jet tagging applications, linear classification with 1000 EFPs achieved comparable performance to a CNN acting on a jet image with $33 \times 33 = 1089$ pixels. Because of the wide variety of linear learning methods available [122], we expect that the EFPs will be a useful starting point to explore more applications in jet substructure and potentially elsewhere in collider physics.

There are many possible refinements and extensions to the energy flow basis. In this paper, we truncated the EFPs at a fixed maximum degree d; alternatively, one could truncate the prime EFPs at a fixed d and compute all composite EFPs up to a specified cutoff. Since the EFPs yield an overcomplete basis, it could be valuable to cull the list of required multigraphs. A similar problem of overcompleteness was solved for kinematic polynomial rings in ref. [110], and that strategy may be relevant for EFPs with a suitable choice of measure. In the other direction, it may be valuable to make the energy flow basis even more redundant by including EFPs with multiple measures. With a vastly overcomplete basis, one could use techniques like lasso regression [121] to zero out unnecessary terms. While we have restricted our attention to IRC-safe observables, it would be straightforward to relax the restriction to just infrared safety. In particular, the set of IR-safe (but C-unsafe) functions in eq. (3.15) can be expanded into multigraphs that have an extra integer decoration on each vertex to indicate the energy scaling. Finally, the EFPs are based on an expansion in pairwise angles, but one could explore alternative angular expansions in terms of single particle directions or multiparticle factors.

To gain some perspective, we find it useful to discuss the EFPs in the broader context of machine learning for jet substructure. Over the past few years, there has been a surge of interest in using powerful tools from machine learning to learn useful observables from low-level or high-level representations of a jet [50, 57–63, 65–73]. The power of these machine learning methods is formidable, and techniques like neural networks and boosted decision trees have shifted the focus away from single- or few-variable jet substructure taggers to multivariate methods. On the other hand, multivariate methods can sometimes obscure the specific physics information that the model learns, leading to recent efforts to "open the box" of machine learning tools [63, 72, 113, 132, 133]. Even with an open box, though, theoretical calculations of multivariate distributions are impractical (if not impossible). Furthermore, training multivariate models is often difficult, requiring large datasets, hyperparameter tuning, and preprocessing of the data.

The EFPs represent both a continuation of and a break from these machine learning trends. The EFPs continue the trend from multivariate to hypervariate representations for jet information, with $\mathcal{O}(100)$ elements needed for effective regression and classification. On the other hand, the linear-spanning nature of the EFPs make it feasible to move away from "black box" nonlinear algorithms and return to simpler linear methods (explored previously for jet substructure in e.g. [50, 55]) without loss of generality. Armed with the energy flow basis, there is a suite of powerful tools and ideas from linear regression and classification which can now be fully utilized for jet substructure applications, with simpler training

JHEP04(2018)013

processes compared to DNNs and stronger guarantees of optimal training convergence. Multivariate methods would ideally be trained directly on data to avoid relying on imperfect simulations, as discussed in ref. [134]. The energy flow basis may be compelling for recent data-driven learning approaches [134–136] due to its completeness, the simplicity of linear learning algorithms, and a potentially lessened requirement on the size of training samples.

As with any jet observable, the impact of non-perturbative effects on the EFPs is important to understand. Even with IRC safety, hadronization modifies the distributions predicted by pQCD and therefore complicates first-principles calculations. It would be interesting to see if the shape function formalism [137, 138] could be used to predict the impact of non-perturbative contributions to EFP distributions. Alternatively, one standard tool that is used to mitigate non-perturbative effects is jet grooming [5, 105–109], which also simplifies first-principles calculations and allows for "quark" and "gluon" jets to be theoretically well-defined [139]. We leave a detailed study of the effects of non-perturbative contributions and jet grooming on EFPs to future work.

Eventually, one hopes that the EFPs will be amenable to precision theoretical calculations of jet substructure (see e.g. refs. [108, 139–148]). This is by no means obvious, since generic EFPs have different power-counting structures from the ECFs [51] or ECFGs [52]. That said, phrasing jet substructure entirely in the language of energy flow observables and energy correlations may provide interesting new theoretical avenues to probe QCD, realizing the C-correlator vision of refs. [74, 87–89]. Most IRC-safe jet observables rely on particle-level definitions and calculations, but there has been theoretical interest in directly analyzing the correlations of energy flow in specific angular directions [149–151], particularly in the context of conformal field theory [152–156]. The energy flow basis is a step towards connecting the particle-level and energy-correlation pictures, and one could even imagine that the energy flow logic could be applied directly at the path integral level. Ultimately, the structure of the EFPs is a direct consequence of IRC safety, resulting in a practical tool for jet substructure at colliders as well as a new way of thinking about the space of observables more generally.

Acknowledgments

We thank J.J. Carrasco, Yang-Ting Chien, Timothy Cohen, Kyle Cranmer, Matthew Dolan, Frédéric Dreyer, Marat Freytsis, Brian Henning, Gregor Kasieczka, Andrew Larkoski, Alex Lombardi, Ian Moult, Benjamin Nachman, Lina Necib, Miruna Oprescu, Bryan Ostdiek, Gilad Perez, Matthew Schwartz, and Jon Walsh for illuminating discussions. This work was supported by the Office of Nuclear Physics of the U.S. Department of Energy (DOE) under grant DE-SC-0011090 and the DOE Office of High Energy Physics under grant DE-SC-0012567. Cloud computing resources were provided through a Microsoft Azure for Research award.

A Energy flow and the stress-energy tensor

In this appendix, we review the connection between the energy flow of an event, as described by the stress-energy tensor, to multiparticle energy correlators [74, 87–89].

JHEP04(2018)013

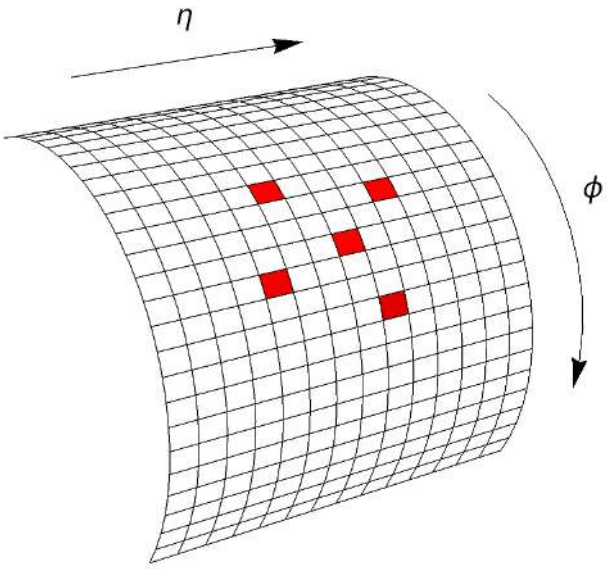


Figure 9. An example calorimeter cell configuration to measure a 5-point energy correlator. The red regions indicate the five calorimeter cells chosen to measure the energy infinitely far from the interaction. For each event, the values of the five energy deposits are multiplied together to obtain the value of the observable in eq. (A.3).

Consider an idealized hadronic calorimeter cell at position $\hat{n}$ in pseudorapidity-azimuth (η, ϕ)-space, spanning a patch of size $d\eta\, d\phi$. The *energy flow operator* $\mathcal{E}_T(\hat{n})$ corresponding to the total transverse momentum density flowing into the calorimeter cell can be written in terms of the stress-energy tensor $T_{\mu\nu}$ [87, 98, 157–159] as:

$$\mathcal{E}_T(\hat{n}) = \frac{1}{\cosh^3 \eta} \lim_{R\to\infty} R^2 \int_0^\infty dt\, \hat{n}_i\, T^{0i}(t, R\hat{n}), \tag{A.1}$$

with its action on a state $|X\rangle$ of M massless particles given by:

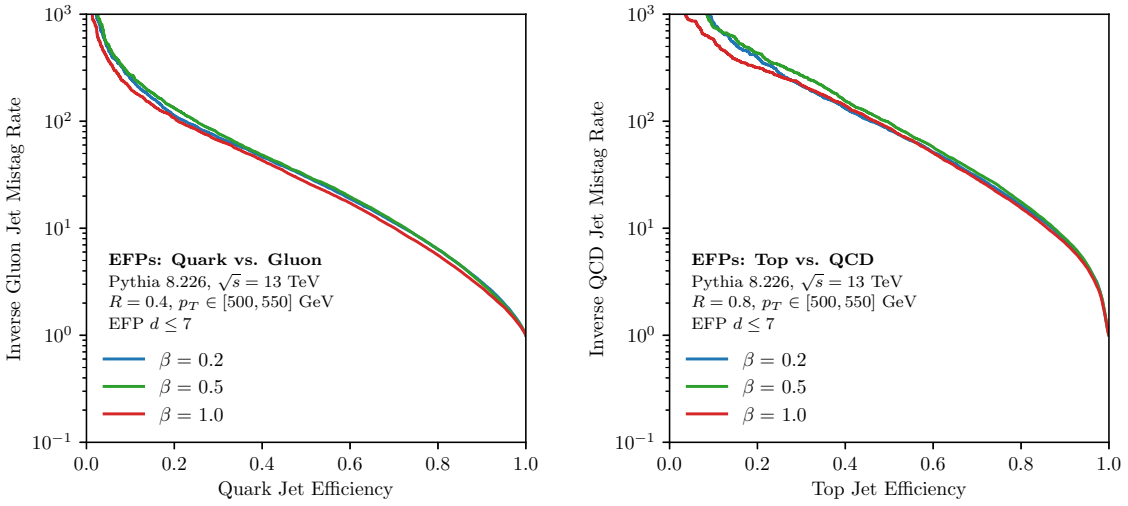
$$\mathcal{E}_T(\hat{n})\,|X\rangle = \sum_{i\in X} p_{T,i}\delta(\eta-\eta_i)\delta(\phi-\phi_i)\,|X\rangle\,. \tag{A.2}$$

Next, consider N calorimeter cells at positions $(\hat{n}_1, \cdots, \hat{n}_N)$. An illustration of an example calorimeter cell configuration is shown in figure 9. For an event X, multiply together the measured energy deposits in each of these N cells. The corresponding observable is then the energy N-point correlator as defined in refs. [149, 150]:

$$\mathcal{E}_T(\hat{n}_1)\cdots\mathcal{E}_T(\hat{n}_N)\,|X\rangle = \sum_{i_1\in X}\cdots\sum_{i_N\in X}\left[\prod_{j=1}^{N} p_{T,i_j}\delta^2(\hat{n}_j-\hat{p}_{i_j})\right]|X\rangle\,, \tag{A.3}$$

where $\hat{n}_a = (\eta_a, \phi_a)$ for the calorimeters cells and $\hat{p}_a = (\eta_a, \phi_a)$ for the particles in the event.

JHEP04(2018)013



(a) **(b)**

Figure 10. Same as figure 5, but for (a) quark/gluon discrimination and (b) top tagging. Similar to the W tagging case, the $\beta = 0.5$ choice has the best performance (marginally) for both tagging problems.

We can define a new set of observables in terms of the N-point correlators in eq. (A.3). Consider averaging eq. (A.3) over all calorimeter cells with an arbitrary angular weighting function $f_N(\hat{n}_1, \ldots, \hat{n}_N)$. The resulting observables are then of the form:

$$\mathcal{C}_N^{f_N} |X\rangle = \int d^2\hat{n}_1 \cdots d^2\hat{n}_N \, f_N(\hat{n}_1, \ldots, \hat{n}_N) \mathcal{E}_T(\hat{n}_1) \cdots \mathcal{E}_T(\hat{n}_N) |X\rangle \tag{A.4}$$

$$= \sum_{i_1 \in X} \cdots \sum_{i_N \in X} p_{T,i_1} \cdots p_{T,i_N} f_N(\hat{p}_{i_1}, \ldots, \hat{p}_{i_N}) |X\rangle \,, \tag{A.5}$$

namely, these observables $\mathcal{C}_N^{f_N}$ written in the form of eq. (A.5) are exactly the C-correlators defined in eq. (1.3). Thus the averaging process of eq. (A.4) relates the particle-level C-correlators of eq. (1.3) to the energy flow of the stress-energy tensor $T_{\mu\nu}$.

B Quark/gluon discrimination and top tagging results

In this appendix, we supplement the W tagging results of section 6 with the corresponding results for quark/gluon discrimination and top tagging. The details of the event generation are given in section 5.2.

We compare the EFP linear classification performance with $\beta = 0.2$, $\beta = 0.5$, and $\beta = 1$ in figure 10. Consistent with the W tagging case in figure 5, we find that the optimal performance is achieved with $\beta = 0.5$. We therefore use $\beta = 0.5$ for the remainder of this study.

In figure 11 we compare the linear and nonlinear performances of the energy flow basis and the N-subjettiness basis. There is a clear gap between the linear and nonlinear N-subjettiness classifiers, whereas no such gap exists for the EFPs. Interestingly, the linear

JHEP04(2018)013

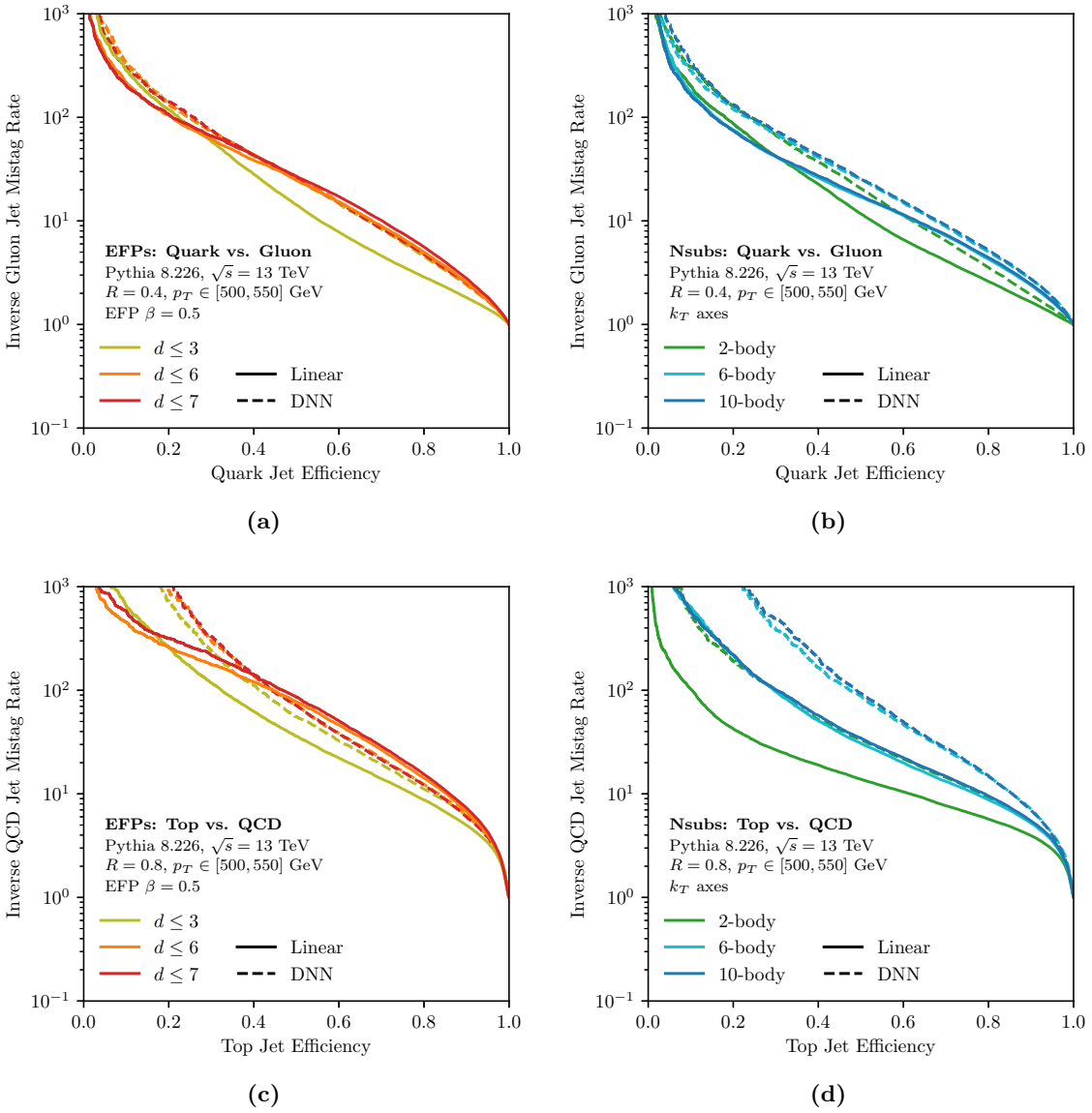
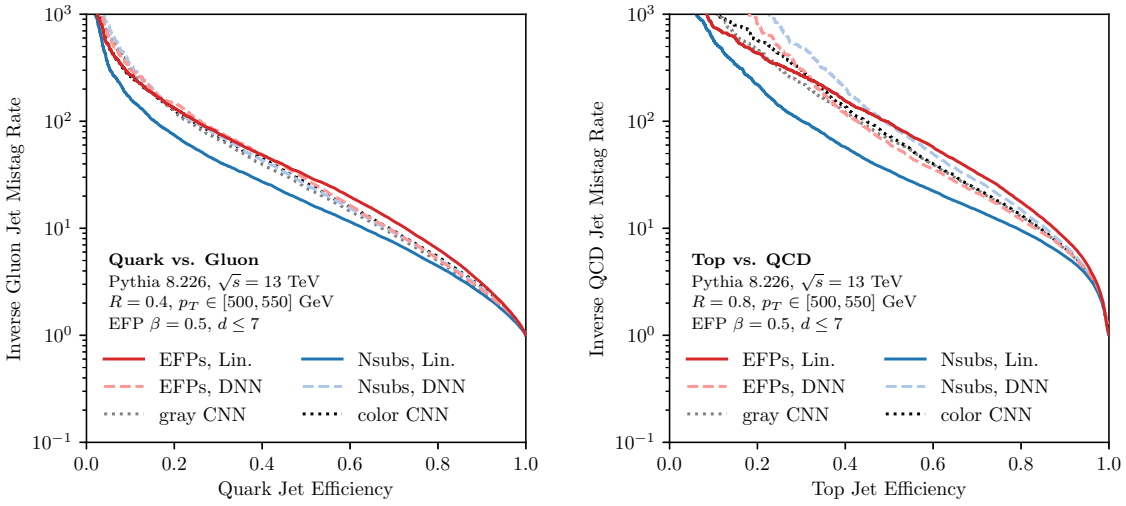


Figure 11. Same as figure 6, but for quark/gluon discrimination (top) and top tagging (bottom). As in the W tagging case, the linear combinations of EFPs can be seen to approach (or even exceed) the nonlinear combinations, particularly for higher signal efficiencies.

classifier of EFPs tends to outperform the DNN at medium and high signal efficiencies, indicating the difficulty of training high-dimensional neural networks. This behavior was not seen in figure 6, most likely because the achievable efficiency is overall higher in the W tagging case.

A summary of the six tagging methods is shown in figure 12, comparing linear and nonlinear combinations of the energy flow basis and N-subjettiness basis to grayscale and color jet images. As in figure 7, linear combinations of EFP tend to match or outperform the other methods, especially at high signal efficiencies.

JHEP04(2018)013



(a) **(b)**

Figure 12. Same as figure 7 for (a) quark/gluon discrimination and (b) top tagging. As in the W tagging case, the linear classification with EFPs can match (or even outperform) the other methods at high signal efficiencies.

JHEP04(2018)013

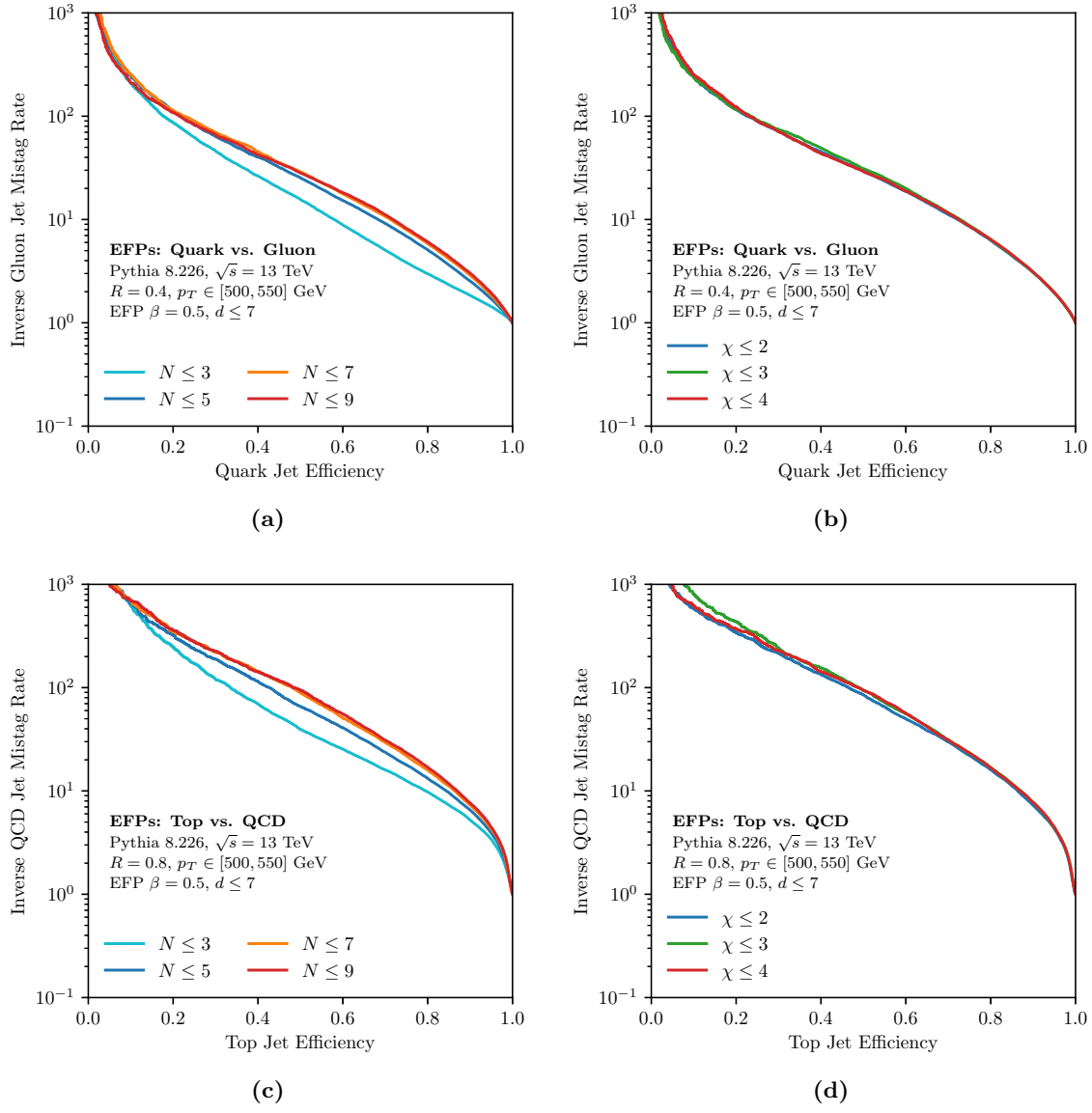


Figure 13. Same as figure 8 but for (top) quark/gluon discrimination and (bottom) top tagging.

JHEP04(2018)013

Finally, we truncate the set of EFPs with $d \leq 7$ by the number of vertices N and by the VE computational complexity χ in figure 13. As in figure 8, the higher N-particle correlators contribute to the classification performance up to at least $N = 7$, whereas the higher-complexity EFPs beyond $\chi = 2$ do not significantly contribute to the classification performance.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] M.H. Seymour, *Tagging a heavy Higgs boson*, in *ECFA Large Hadron Collider Workshop*, Aachen Germany, 4–9 October 1990, proceedings 2, (1991), pg. 557 [INSPIRE].

[2] M.H. Seymour, *Searches for new particles using cone and cluster jet algorithms: a comparative study*, *Z. Phys.* **C 62** (1994) 127 [INSPIRE].

[3] J.M. Butterworth, B.E. Cox and J.R. Forshaw, *WW scattering at the CERN LHC*, *Phys. Rev.* **D 65** (2002) 096014 [hep-ph/0201098] [INSPIRE].

[4] J.M. Butterworth, J.R. Ellis and A.R. Raklev, *Reconstructing sparticle mass spectra using hadronic decays*, *JHEP* **05** (2007) 033 [hep-ph/0702150] [INSPIRE].

[5] J.M. Butterworth, A.R. Davison, M. Rubin and G.P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [arXiv:0802.2470] [INSPIRE].

[6] CMS collaboration, *Search for a Higgs boson in the decay channel $H \to ZZ^{(*)} \to q\bar{q}\ell^-\ell^+$ in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **04** (2012) 036 [arXiv:1202.1416] [INSPIRE].

[7] CMS collaboration, *Search for a Standard Model-like Higgs boson decaying into $WW \to \ell\nu q\bar{q}$ in pp collisions at $\sqrt{s} = 8$ TeV*, CMS-PAS-HIG-13-008, CERN, Geneva Switzerland, (2013).

[8] ATLAS collaboration, *Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV*, *Phys. Lett.* **B 750** (2015) 475 [arXiv:1506.05629] [INSPIRE].

[9] ATLAS collaboration, *Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, *Phys. Rev.* **D 93** (2016) 052003 [arXiv:1509.05190] [INSPIRE].

[10] ATLAS collaboration, *Performance of jet substructure techniques in early $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, ATLAS-CONF-2015-035, CERN, Geneva Switzerland, (2015).

[11] ATLAS collaboration, *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS detector*, *Phys. Rev.* **D 93** (2016) 032009 [arXiv:1510.03818] [INSPIRE].

[12] ATLAS collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C 76** (2016) 154 [arXiv:1510.05821] [INSPIRE].

JHEP04(2018)013

[13] ATLAS collaboration, *Studies of b-tagging performance and jet substructure in a high p_{T} $g \to b\bar{b}$ rich sample of large-R jets from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, ATLAS-CONF-2016-002, CERN, Geneva Switzerland, (2016).

[14] ATLAS collaboration, *Boosted Higgs ($\to b\bar{b}$) boson identification with the ATLAS detector at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2016-039, CERN, Geneva Switzerland, (2016).

[15] ATLAS collaboration, *Discrimination of light quark and gluon jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, ATLAS-CONF-2016-034, CERN, Geneva Switzerland, (2016).

[16] CMS collaboration, *Measurement of the $t\bar{t}$ production cross section at 13 TeV in the all-jets final state*, CMS-PAS-TOP-16-013, CERN, Geneva Switzerland, (2016).

[17] ATLAS and CMS collaborations, G. Rauco, *Distinguishing quark and gluon jets at the LHC*, in *Proceedings, Parton Radiation and Fragmentation from LHC to FCC-ee*, CERN, Geneva Switzerland, 22–23 November 2016 [INSPIRE].

[18] ATLAS collaboration, *Performance of top quark and W boson tagging in run 2 with ATLAS*, ATLAS-CONF-2017-064, CERN, Geneva Switzerland, (2017).

[19] CMS collaboration, *Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair*, *Phys. Rev. Lett.* **120** (2018) 071802 [arXiv:1709.05543] [INSPIRE].

[20] CMS collaboration, *Search for BSM $t\bar{t}$ production in the boosted all-hadronic final state*, CMS-PAS-EXO-11-006, CERN, Geneva Switzerland, (2011).

[21] CMS collaboration, *Search for anomalous $t\bar{t}$ production in the highly-boosted all-hadronic final state*, *JHEP* **09** (2012) 029 [*Erratum ibid.* **03** (2014) 132] [arXiv:1204.2488] [INSPIRE].

[22] ATLAS, CMS collaboration, S. Fleischmann, *Boosted top quark techniques and searches for $t\bar{t}$ resonances at the LHC*, *J. Phys. Conf. Ser.* **452** (2013) 012034 [INSPIRE].

[23] ATLAS and CMS collaborations, J. Pilot, *Boosted top quarks, top pair resonances, and top partner searches at the LHC*, *EPJ Web Conf.* **60** (2013) 09003 [INSPIRE].

[24] ATLAS collaboration, *Performance of boosted top quark identification in 2012 ATLAS data*, ATLAS-CONF-2013-084, CERN, Geneva Switzerland, (2013).

[25] CMS collaboration, *Search for pair-produced vector-like quarks of charge $-1/3$ decaying to bH using boosted Higgs jet-tagging in pp collisions at $\sqrt{s} = 8$ TeV*, CMS-PAS-B2G-14-001, CERN, Geneva Switzerland, (2014).

[26] CMS collaboration, *Search for top-Higgs resonances in all-hadronic final states using jet substructure methods*, CMS-PAS-B2G-14-002, CERN, Geneva Switzerland, (2014).

[27] CMS collaboration, *Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure*, *JHEP* **06** (2015) 080 [arXiv:1503.01952] [INSPIRE].

[28] ATLAS collaboration, *Search for diboson resonances in the $\nu\nu qq$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2015-068, CERN, Geneva Switzerland, (2015).

[29] ATLAS collaboration, *Search for diboson resonances in the $\ell\ell qq$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2015-071, CERN, Geneva Switzerland, (2015).

JHEP04(2018)013

[30] CMS collaboration, *Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, *JHEP* **02** (2016) 145 [arXiv:1506.01443] [INSPIRE].

[31] CMS collaboration, *Search for* $t\bar{t}H$ *production in the* $H \to b\bar{b}$ *decay channel with* $\sqrt{s} = 13$ *TeV pp collisions at the CMS experiment*, CMS-PAS-HIG-16-004, CERN, Geneva Switzerland, (2016).

[32] CMS collaboration, *Search for dark matter in association with a boosted top quark in the all hadronic final state*, CMS-PAS-EXO-16-017, CERN, Geneva Switzerland, (2016).

[33] CMS collaboration, *Search for top quark-antiquark resonances in the all-hadronic final state at* $\sqrt{s} = 13$ *TeV*, CMS-PAS-B2G-15-003, CERN, Geneva Switzerland, (2015).

[34] CMS collaboration, *Searches for invisible Higgs boson decays with the CMS detector*, CMS-PAS-HIG-16-016, CERN, Geneva Switzerland, (2016).

[35] ATLAS collaboration, *Searches for heavy diboson resonances in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *JHEP* **09** (2016) 173 [arXiv:1606.04833] [INSPIRE].

[36] CMS collaboration, *Search for dark matter in proton-proton collisions at* 8 *TeV with missing transverse momentum and vector boson tagged jets*, *JHEP* **12** (2016) 083 [*Erratum ibid.* **08** (2017) 035] [arXiv:1607.05764] [INSPIRE].

[37] CMS collaboration, *Search for new physics in a boosted hadronic monotop final state using* $12.9\,fb^{-1}$ *of* $\sqrt{s} = 13$ *TeV data*, CMS-PAS-EXO-16-040, CERN, Geneva Switzerland, (2016).

[38] CMS collaboration, *Search for dark matter in final states with an energetic jet, or a hadronically decaying W or Z boson using* $12.9\,fb^{-1}$ *of data at* $\sqrt{s} = 13$ *TeV*, CMS-PAS-EXO-16-037, CERN, Geneva Switzerland, (2016).

[39] ATLAS collaboration, *Search for dark matter produced in association with a hadronically decaying vector boson in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B 763** (2016) 251 [arXiv:1608.02372] [INSPIRE].

[40] CMS collaboration, *Search for high-mass* $Z\gamma$ *resonances in proton-proton collisions at* $\sqrt{s} = 8$ *and* 13 *TeV using jet substructure techniques*, *Phys. Lett.* **B 772** (2017) 363 [arXiv:1612.09516] [INSPIRE].

[41] CMS collaboration, *Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **11** (2017) 085 [arXiv:1706.03408] [INSPIRE].

[42] CMS collaboration, *Search for massive resonances decaying into WW, WZ, ZZ, qW and qZ with dijet final states at* $\sqrt{s} = 13$ *TeV*, arXiv:1708.05379 [INSPIRE].

[43] CMS collaboration, *Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **01** (2018) 097 [arXiv:1710.00159] [INSPIRE].

[44] A. Abdesselam et al., *Boosted objects: a probe of beyond the Standard Model physics*, *Eur. Phys. J.* **C 71** (2011) 1661 [arXiv:1012.5412] [INSPIRE].

[45] A. Altheimer et al., *Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks*, *J. Phys.* **G 39** (2012) 063001 [arXiv:1201.0008] [INSPIRE].

JHEP04(2018)013

[46] A. Altheimer et al., *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd–27th of July 2012*, *Eur. Phys. J.* **C 74** (2014) 2792 [arXiv:1311.2708] [INSPIRE].

[47] D. Adams et al., *Towards an understanding of the correlations in jet substructure*, *Eur. Phys. J.* **C 75** (2015) 409 [arXiv:1504.00679] [INSPIRE].

[48] 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 [INSPIRE].

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

[50] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-images: computer vision inspired techniques for jet tagging*, *JHEP* **02** (2015) 118 [arXiv:1407.5675] [INSPIRE].

[51] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy correlation functions for jet substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [INSPIRE].

[52] I. Moult, L. Necib and J. Thaler, *New angles on energy correlation functions*, *JHEP* **12** (2016) 153 [arXiv:1609.07483] [INSPIRE].

[53] A.J. Larkoski, I. Moult and D. Neill, *Power counting to better jet observables*, *JHEP* **12** (2014) 009 [arXiv:1409.6298] [INSPIRE].

[54] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [INSPIRE].

[55] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [arXiv:1108.2701] [INSPIRE].

[56] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *N-jettiness: an inclusive event shape to veto jets*, *Phys. Rev. Lett.* **105** (2010) 092002 [arXiv:1004.2489] [INSPIRE].

[57] K. Datta and A. Larkoski, *How much information is in a jet?*, *JHEP* **06** (2017) 073 [arXiv:1704.08249] [INSPIRE].

[58] J. Gallicchio, J. Huth, M. Kagan, M.D. Schwartz, K. Black and B. Tweedie, *Multivariate discrimination and the Higgs + W/Z search*, *JHEP* **04** (2011) 069 [arXiv:1010.3698] [INSPIRE].

[59] J. Gallicchio and M.D. Schwartz, *Quark and gluon tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [arXiv:1106.3076] [INSPIRE].

[60] J. Gallicchio and M.D. Schwartz, *Quark and gluon jet substructure*, *JHEP* **04** (2013) 090 [arXiv:1211.7038] [INSPIRE].

[61] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, *Nature Commun.* **5** (2014) 4308 [arXiv:1402.4735] [INSPIRE].

[62] P. Baldi, P. Sadowski and D. Whiteson, *Enhanced Higgs boson to $\tau^+\tau^-$ search with deep learning*, *Phys. Rev. Lett.* **114** (2015) 111801 [arXiv:1410.3469] [INSPIRE].

[63] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [arXiv:1511.05190] [INSPIRE].

[64] J. Barnard, E.N. Dawe, M.J. Dolan and N. Rajcic, *Parton shower uncertainties in jet substructure analyses with deep neural networks*, *Phys. Rev.* **D 95** (2017) 014018 [arXiv:1609.00607] [INSPIRE].

JHEP04(2018)013

[65] P.T. Komiske, E.M. Metodiev and M.D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110 [arXiv:1612.01551] [INSPIRE].

[66] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning top taggers or the end of QCD?*, *JHEP* **05** (2017) 006 [arXiv:1701.08784] [INSPIRE].

[67] L.G. Almeida, M. Backović, M. Cliche, S.J. Lee and M. Perelstein, *Playing tag with ANN: boosted top identification with pattern recognition*, *JHEP* **07** (2015) 086 [arXiv:1501.05968] [INSPIRE].

[68] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet substructure classification in high-energy physics with deep neural networks*, *Phys. Rev.* **D 93** (2016) 094034 [arXiv:1603.09349] [INSPIRE].

[69] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet flavor classification in high-energy physics with deep neural networks*, *Phys. Rev.* **D 94** (2016) 112002 [arXiv:1607.08633] [INSPIRE].

[70] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-aware recursive neural networks for jet physics*, arXiv:1702.00748 [INSPIRE].

[71] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet constituents for deep neural network based top quark tagging*, arXiv:1704.02124 [INSPIRE].

[72] A. Butter, G. Kasieczka, T. Plehn and M. Russell, *Deep-learned top tagging with a Lorentz layer*, arXiv:1707.08966 [INSPIRE].

[73] J.A. Aguilar-Saavedra, J.H. Collins and R.K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163 [arXiv:1709.01087] [INSPIRE].

[74] F.V. Tkachov, *Measuring multi-jet structure of hadronic energy flow or what is a jet?*, *Int. J. Mod. Phys.* **A 12** (1997) 5411 [hep-ph/9601308] [INSPIRE].

[75] A.J. Larkoski, J. Thaler and W.J. Waalewijn, *Gaining (mutual) information about quark/gluon discrimination*, *JHEP* **11** (2014) 129 [arXiv:1408.3122] [INSPIRE].

[76] M.H. Stone, *The generalized Weierstrass approximation theorem*, *Math. Magazine* **21** (1948) 237.

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

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

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

[80] S. Weinberg, *The quantum theory of fields. Volume* 1*: foundations*, Cambridge University Press, Cambridge U.K., (2005) [INSPIRE].

[81] CTEQ collaboration, R. Brock et al., *Handbook of perturbative QCD: version* 1.0, *Rev. Mod. Phys.* **67** (1995) 157 [INSPIRE].

[82] A.J. Larkoski and J. Thaler, *Unsafe but calculable: ratios of angularities in perturbative QCD*, *JHEP* **09** (2013) 137 [arXiv:1307.1699] [INSPIRE].

[83] A.J. Larkoski, S. Marzani and J. Thaler, *Sudakov safety in perturbative QCD*, *Phys. Rev.* **D 91** (2015) 111501 [arXiv:1502.01719] [INSPIRE].

JHEP04(2018)013

[84] W.J. Waalewijn, *Calculating the charge of a jet*, *Phys. Rev.* **D 86** (2012) 094030 [arXiv:1209.3019] [INSPIRE].

[85] H.-M. Chang, M. Procura, J. Thaler and W.J. Waalewijn, *Calculating track-based observables for the LHC*, *Phys. Rev. Lett.* **111** (2013) 102002 [arXiv:1303.6637] [INSPIRE].

[86] B.T. Elder, M. Procura, J. Thaler, W.J. Waalewijn and K. Zhou, *Generalized fragmentation functions for fractal jet observables*, *JHEP* **06** (2017) 085 [arXiv:1704.05456] [INSPIRE].

[87] N.A. Sveshnikov and F.V. Tkachov, *Jets and quantum field theory*, *Phys. Lett.* **B 382** (1996) 403 [hep-ph/9512370] [INSPIRE].

[88] P.S. Cherzor and N.A. Sveshnikov, *Jet observables and energy momentum tensor*, in *Quantum field theory and high-energy physics. Proceedings, Workshop, QFTHEP'97*, Samara Russia, 4–10 September 1997, pg. 402 [hep-ph/9710349] [INSPIRE].

[89] F.V. Tkachov, *A theory of jet definition*, *Int. J. Mod. Phys.* **A 17** (2002) 2783 [hep-ph/9901444] [INSPIRE].

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

[91] S.S. Mandal, S. Mukherjee and K. Ray, *Determination of many-electron basis functions for a quantum Hall ground state using Schur polynomials*, *Annals Phys.* **390** (2018) 236 [arXiv:1708.07658] [INSPIRE].

[92] 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** (2015) 025005 [arXiv:1409.1581] [INSPIRE].

[93] A.J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination at the Large Hadron Collider*, arXiv:1708.06760 [INSPIRE].

[94] N.J. Sloane, *The on-line encyclopedia of integer sequences*, in *Towards mechanized mathematical assistants*, Springer, Berlin Heidelberg Germany, (2007), pg. 130.

[95] F. Harary and E.M. Palmer, *Graphical enumeration*, Elsevier, The Netherlands, (2014).

[96] J. Gallicchio and M.D. Schwartz, *Seeing in color: jet superstructure*, *Phys. Rev. Lett.* **105** (2010) 022001 [arXiv:1001.5027] [INSPIRE].

[97] G.P. Salam and D. Wicke, *Hadron masses and power corrections to event shapes*, *JHEP* **05** (2001) 061 [hep-ph/0102343] [INSPIRE].

[98] V. Mateu, I.W. Stewart and J. Thaler, *Power corrections to event shapes with mass-dependent operators*, *Phys. Rev.* **D 87** (2013) 014025 [arXiv:1209.3781] [INSPIRE].

[99] I.W. Stewart, F.J. Tackmann, J. Thaler, C.K. Vermilion and T.F. Wilkason, *XCone: N-jettiness as an exclusive cone jet algorithm*, *JHEP* **11** (2015) 072 [arXiv:1508.01516] [INSPIRE].

[100] C.F. Berger, T. Kucs and G.F. Sterman, *Event shape/energy flow correlations*, *Phys. Rev.* **D 68** (2003) 014012 [hep-ph/0303051] [INSPIRE].

[101] 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** (2009) 074017 [arXiv:0807.0234] [INSPIRE].

[102] S.D. Ellis, C.K. Vermilion, J.R. Walsh, A. Hornig and C. Lee, *Jet shapes and jet algorithms in SCET*, *JHEP* **11** (2010) 101 [arXiv:1001.0014] [INSPIRE].

JHEP04(2018)013

[103] A.J. Larkoski, D. Neill and J. Thaler, *Jet shapes with the broadening axis*, *JHEP* **04** (2014) 017 [arXiv:1401.2158] [INSPIRE].

[104] J. Thaler and L.-T. Wang, *Strategies to identify boosted tops*, *JHEP* **07** (2008) 092 [arXiv:0806.0023] [INSPIRE].

[105] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev.* **D 80** (2009) 051501 [arXiv:0903.5081] [INSPIRE].

[106] 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** (2010) 094023 [arXiv:0912.0033] [INSPIRE].

[107] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, *JHEP* **02** (2010) 084 [arXiv:0912.1342] [INSPIRE].

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

[109] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146 [arXiv:1402.2657] [INSPIRE].

[110] B. Henning, X. Lu, T. Melia and H. Murayama, *Operator bases, S-matrices and their partition functions*, *JHEP* **10** (2017) 199 [arXiv:1706.08520] [INSPIRE].

[111] M. Cacciari and G.P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett.* **B 641** (2006) 57 [hep-ph/0512210] [INSPIRE].

[112] K.P. Murphy, *Machine learning: a probabilistic perspective*, MIT press, U.S.A., (2012).

[113] P.T. Komiske, E.M. Metodiev, B. Nachman and M.D. Schwartz, *Pileup mitigation with machine learning (PUMML)*, *JHEP* **12** (2017) 051 [arXiv:1707.08600] [INSPIRE].

[114] L. de Oliveira, M. Paganini and B. Nachman, *Learning particle physics by example: location-aware generative adversarial networks for physics synthesis*, *Comput. Softw. Big Sci.* **1** (2017) 4 [arXiv:1701.05927] [INSPIRE].

[115] M. Paganini, L. de Oliveira and B. Nachman, *Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters*, *Phys. Rev. Lett.* **120** (2018) 042003 [arXiv:1705.02355] [INSPIRE].

[116] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA* 6.4 *physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[117] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA* 8.1, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[118] T. Sjöstrand et al., *An introduction to PYTHIA* 8.2, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[119] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[120] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[121] R. Tibshirani, *Regression shrinkage and selection via the lasso: a retrospective*, *J. Roy. Statist. Soc.* **B 73** (2011) 273.

[122] C.M. Bishop, *Pattern recognition and machine learning*, Springer, New York U.S.A., (2006).

JHEP04(2018)013

[123] F. Pedregosa et al., *Scikit-learn: machine learning in Python*, *J. Machine Learning Res.* **12** (2011) 2825 [arXiv:1201.0490] [INSPIRE].

[124] J.R. Andersen et al., *Les Houches* 2015*: physics at TeV colliders Standard Model working group report*, in 9th *Les Houches Workshop on Physics at TeV Colliders (PhysTeV* 2015*)*, FERMILAB-CONF-16-175-PPD-T, Les Houches France, 1–19 June 2015 [arXiv:1605.04692] [INSPIRE].

[125] P. Gras et al., *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091 [arXiv:1704.03878] [INSPIRE].

[126] R.A. Fisher, *The use of multiple measurements in taxonomic problems*, *Ann. Human Genetics* **7** (1936) 179.

[127] V. Nair and G.E. Hinton, *Rectified linear units improve restricted Boltzmann machines*, in *Proceedings of the* 27th *international conference on machine learning (ICML-*10*)*, (2010), pg. 807.

[128] K. He, X. Zhang, S. Ren and J. Sun, *Delving deep into rectifiers: surpassing human-level performance on imagenet classification*, in *Proceedings of the IEEE international conference on computer vision*, IEEE, (2015), pg. 1026.

[129] D. Kingma and J. Ba, *Adam: a method for stochastic optimization*, arXiv:1412.6980.

[130] F. Chollet, *Keras*, https://github.com/fchollet/keras, (2017).

[131] J. Bergstra et al., *Theano: a CPU and GPU math compiler in Python*, in *Proc.* 9th *Python in Science Conf.*, (2010), pg. 1.

[132] S. Chang, T. Cohen and B. Ostdiek, *What is the machine learning?*, *Phys. Rev.* **D 97** (2018) 056009 [arXiv:1709.10106] [INSPIRE].

[133] E.M. Metodiev and J. Thaler, *On the topic of jets*, arXiv:1802.00008 [INSPIRE].

[134] P.T. Komiske, E.M. Metodiev, B. Nachman and M.D. Schwartz, *Learning to classify from impure samples*, arXiv:1801.10158 [INSPIRE].

[135] L.M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, *Weakly supervised classification in high energy physics*, *JHEP* **05** (2017) 145 [arXiv:1702.00414] [INSPIRE].

[136] E.M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174 [arXiv:1708.02949] [INSPIRE].

[137] G.P. Korchemsky and G.F. Sterman, *Power corrections to event shapes and factorization*, *Nucl. Phys.* **B 555** (1999) 335 [hep-ph/9902341] [INSPIRE].

[138] G.P. Korchemsky and S. Tafat, *On power corrections to the event shape distributions in QCD*, *JHEP* **10** (2000) 010 [hep-ph/0007005] [INSPIRE].

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

[140] I. Feige, M.D. Schwartz, I.W. Stewart and J. Thaler, *Precision jet substructure from boosted event shapes*, *Phys. Rev. Lett.* **109** (2012) 092001 [arXiv:1204.3898] [INSPIRE].

[141] M. Dasgupta, A. Fregoso, S. Marzani and A. Powling, *Jet substructure with analytical methods*, *Eur. Phys. J.* **C 73** (2013) 2623 [arXiv:1307.0013] [INSPIRE].

JHEP04(2018)013

[142] A.J. Larkoski, I. Moult and D. Neill, *Toward multi-differential cross sections: measuring two angularities on a single jet*, *JHEP* **09** (2014) 046 [arXiv:1401.4458] [INSPIRE].

[143] M. Procura, W.J. Waalewijn and L. Zeune, *Resummation of double-differential cross sections and fully-unintegrated parton distribution functions*, *JHEP* **02** (2015) 117 [arXiv:1410.6483] [INSPIRE].

[144] A.J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination*, *JHEP* **05** (2016) 117 [arXiv:1507.03018] [INSPIRE].

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

[146] M. Dasgupta, A. Powling, L. Schunk and G. Soyez, *Improved jet substructure methods: Y-splitter and variants with grooming*, *JHEP* **12** (2016) 079 [arXiv:1609.07149] [INSPIRE].

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

[148] A.J. Larkoski, I. Moult and D. Neill, *Factorization and resummation for groomed multi-prong jet shapes*, *JHEP* **02** (2018) 144 [arXiv:1710.00014] [INSPIRE].

[149] C.L. Basham, L.S. Brown, S.D. Ellis and S.T. Love, *Energy correlations in electron-positron annihilation: testing QCD*, *Phys. Rev. Lett.* **41** (1978) 1585 [INSPIRE].

[150] C.L. Basham, L.S. Brown, S.D. Ellis and S.T. Love, *Energy correlations in electron-positron annihilation in quantum chromodynamics: asymptotically free perturbation theory*, *Phys. Rev.* **D 19** (1979) 2018 [INSPIRE].

[151] A.V. Belitsky, G.P. Korchemsky and G.F. Sterman, *Energy flow in QCD and event shape functions*, *Phys. Lett.* **B 515** (2001) 297 [hep-ph/0106308] [INSPIRE].

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

[153] O.T. Engelund and R. Roiban, *Correlation functions of local composite operators from generalized unitarity*, *JHEP* **03** (2013) 172 [arXiv:1209.0227] [INSPIRE].

[154] A. Zhiboedov, *On conformal field theories with extremal a/c values*, *JHEP* **04** (2014) 038 [arXiv:1304.6075] [INSPIRE].

[155] A.V. Belitsky, S. Hohenegger, G.P. Korchemsky, E. Sokatchev and A. Zhiboedov, *From correlation functions to event shapes*, *Nucl. Phys.* **B 884** (2014) 305 [arXiv:1309.0769] [INSPIRE].

[156] A.V. Belitsky, S. Hohenegger, G.P. Korchemsky, E. Sokatchev and A. Zhiboedov, *Event shapes in $N = 4$ super-Yang-Mills theory*, *Nucl. Phys.* **B 884** (2014) 206 [arXiv:1309.1424] [INSPIRE].

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

[158] C. Lee and G.F. Sterman, *Momentum flow correlations from event shapes: factorized soft gluons and soft-collinear effective theory*, *Phys. Rev.* **D 75** (2007) 014022 [hep-ph/0611061] [INSPIRE].

[159] C.W. Bauer, S.P. Fleming, C. Lee and G.F. Sterman, *Factorization of e^+e^- event shape distributions with hadronic final states in soft collinear effective theory*, *Phys. Rev.* **D 78** (2008) 034027 [arXiv:0801.4569] [INSPIRE].

JHEP04(2018)013