

# Hierarchical high-point Energy Flow Network for jet tagging

Wei Shen,<sup>a,b</sup> Daohan Wang<sup>c</sup> and Jin Min Yang<sup>a,b</sup>

<sup>a</sup>*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics,  
Chinese Academy of Sciences,  
Beijing 100190, P. R. China*

<sup>b</sup>*School of Physics Sciences, University of Chinese Academy of Sciences,  
Beijing 100049, P. R. China*

<sup>c</sup>*Department of Physics, Konkuk University,  
Seoul 05029, Republic of Korea*

*E-mail:* [shenwei@itp.ac.cn](mailto:shenwei@itp.ac.cn), [wdh9508@gmail.com](mailto:wdh9508@gmail.com), [jmyang@itp.ac.cn](mailto:jmyang@itp.ac.cn)

**ABSTRACT:** Jet substructure observable basis is a systematic and powerful tool for analyzing the internal energy distribution of constituent particles within a jet. In this work, we propose a novel method to insert neural networks into jet substructure basis as a simple yet efficient interpretable IRC-safe deep learning framework to discover discriminative jet observables. The Energy Flow Polynomial (EFP) could be computed with a certain summation order, resulting in a reorganized form which exhibits hierarchical IRC-safety. Thus inserting non-linear functions after the separate summation could significantly extend the scope of IRC-safe jet substructure observables, where neural networks can come into play as an important role. Based on the structure of the simplest class of EFPs which corresponds to path graphs, we propose the Hierarchical Energy Flow Networks and the Local Hierarchical Energy Flow Networks. These two architectures exhibit remarkable discrimination performance on the top tagging dataset and quark-gluon dataset compared to other benchmark algorithms even only utilizing the kinematic information of constituent particles.

**KEYWORDS:** Jets and Jet Substructure, Specific QCD Phenomenology

**ARXIV EPRINT:** [2308.08300](https://arxiv.org/abs/2308.08300)

---

## Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                                  | <b>1</b>  |
| <b>2</b> | <b>Hierarchical energy flow for observables</b>      | <b>3</b>  |
| 2.1      | Jet substructure observable basis                    | 3         |
| 2.2      | Hierarchical energy flow functions                   | 5         |
| <b>3</b> | <b>Hierarchical higher-point Energy Flow Network</b> | <b>9</b>  |
| 3.1      | Hierarchical Energy Flow Network                     | 9         |
| 3.2      | Local Hierarchical Energy Flow Network               | 10        |
| 3.3      | Model implementation                                 | 11        |
| <b>4</b> | <b>Results of jet classification</b>                 | <b>11</b> |
| 4.1      | Top tagging  | 11        |
| 4.2      | Quark/gluon discrimination                           | 12        |
| <b>5</b> | <b>Conclusions</b>                                   | <b>16</b> |

---

## 1 Introduction

Identifying deviations from the Standard Model predictions in a vast amount of high-energy event data generated at a collider like the LHC is an important task in the search of new physics. One of the most crucial objects in analyzing high-energy collision products is jets, which are collimated sprays of outgoing particles produced in a high energy collision. Jet substructure refers to a series of well-defined physical observables that measure the particle and energy distribution within a jet, reflecting the internal radiation pattern. Identifying novel jet patterns helps to separate signals of interest and provides a quantitative understanding of the underlying mechanisms.

Jet substructure basis is a set of systematic and (over-)complete basis for jet observables. Given that a jet typically consists of hundreds of constituent particles, there is a huge number of possible observable combinations that can serve as a complete basis for such a multi-body system. In practice, there are some restrictions on selecting robust and well-defined observables. To guarantee a good definition in perturbative QCD calculation as well as the robustness to experimental resolution effects, any jet substructure observable should be both infrared and collinear safe (IRC-safe) and Lorentz invariant. Previous studies have proposed a generalized form known as calorimetric-correlators (C-correlators) based on energy-weighted summation for direction-related features of all particles [1], which maintains both IRC-safety and permutation symmetry. Furthermore, any Lorentz invariant jet observable can be decomposed into a combination of Lorentz invariants of particle pairs.

For highly-boosted and narrow jets, the requirement of Lorentz invariance can be relaxed to SO(2)-rotational symmetry around the jet axis [2], and then the angular distance  $R_{ij}$  between particles  $i$  and  $j$  are often used to approximate pair-wise Lorentz invariant features. Under these conditions, most of the energy-flow based substructure basis variables like N-subjettiness [3, 4], energy correlation functions (ECFs) [5], generalized energy correlation functions (ECFGs) [6] and energy-flow polynomials (EFPs) [7] have been proposed. They focus on magnifying significant difference in radiation patterns inside a jet, especially the multi-prong structure of a highly-boosted jet [8]. Besides these substructure basis, there are also several physically defined observables, such as the jet mass, angularities [9] and planar flow [10]. Due to the large amount of data and the great freedom of features selection, machine learning is playing an increasingly important role in jet analysis in high energy collisions (for reviews on machine learning in high energy physics, see, e.g., [11–14]).

Traditional machine learning methods, such as the boost decision trees (BDTs) [15], are widely employed to analyze high-level jet features like jet mass and N-subjettiness, and linear regression is also used to find discriminative jet observables by determining the coefficients of jet substructure bases like EFPs [7]. Over the last decade, there has been a widespread adoption of deep learning techniques to improve the performance of jet tagging [16, 17]. Recently, there is an increasing emphasis on developing network architectures that prioritize both infrared and collinear (IRC) safety and physical interpretability for collision objects, rather than solely optimizing for downstream task performance. More recently, the point cloud jet representation, which treats the constituent particles within a jet as points in a point cloud, has gained significant attention and several deep learning architectures are proposed based on the point cloud representation, including Energy Flow Network [18], Energy-weighted Messaging Passing (EWMP) Neural Network [19], ParticleNet [20], ABCNet [21] and LorentzNet [22]. The Energy Flow Network is an energy-weighted deep set network serving as an IRC-safe backbone model on the point cloud representation of jets [23]. It parameterizes angular filters with trainable neural networks and has been successfully tested on many jet tagging tasks [18]. Meanwhile, the Energy-weighted Messaging Passing (EWMP) Neural Network [19] maps kinematic features of particles to nodes and distances between particles to edges in a graph, which highlights that defining only the Radius Neighbor can make the algorithm IRC-safe, and compares the tagging performance at different aggregation radius settings. Additionally, the ParticleNet [20] architecture has been proposed to process the local structure of jets permutations invariantly by aggregating the information of K-Nearest Neighbor (KNN) through a convolution block while the ABCNet employs the attention mechanism to extract the local structure of jets. The LorentzNet puts greater emphasis on integrating inductive biases derived from physics principles into its architectural design, employing a highly efficient Minkowski dot product attention mechanism. All these architectures exhibit excellent performance when applied to top tagging and quark/gluon discrimination benchmarks. Nevertheless, it is worth noting that only the EFN and the EWMP architectures maintain the IRC-safety.

In this work, we focus on inserting neural networks into jet substructure bases in an IRC-safe and rotation-invariant way to discover interpretability and discriminative jet observables from a vast amount of simulation data. The article is organized as follows. In

section 2, we give a brief introduction for some representative jet substructure observable bases, and propose to use different order of Legendre polynomials as functions of angular distance between particle pairs to achieve numerical stability. As an example, we present the distributions of 4-point Legendre-based path graph energy-flow polynomials for both top jets and QCD jets, spanning across different parameter settings. We also show how to determine the specific form of the 2-point energy-flow polynomials by employing linear regression on the coefficients of Legendre polynomials as an illustrative example. Most importantly, we point out that rewriting EFPs in a hierarchical manner would greatly reduce the computation complexity and allow for the insertion of neural network. In section 3, we introduce the Hierarchical Energy Flow Networks and Local Hierarchical Energy Flow Networks which follow the path-graph structure, and provide comprehensive details about the model implementation. In section 4, we present numerical results obtained for top tagging task and quark/gluon discrimination task, respectively. Finally, our conclusion are presented in section 5.

## 2 Hierarchical energy flow for observables

### 2.1 Jet substructure observable basis

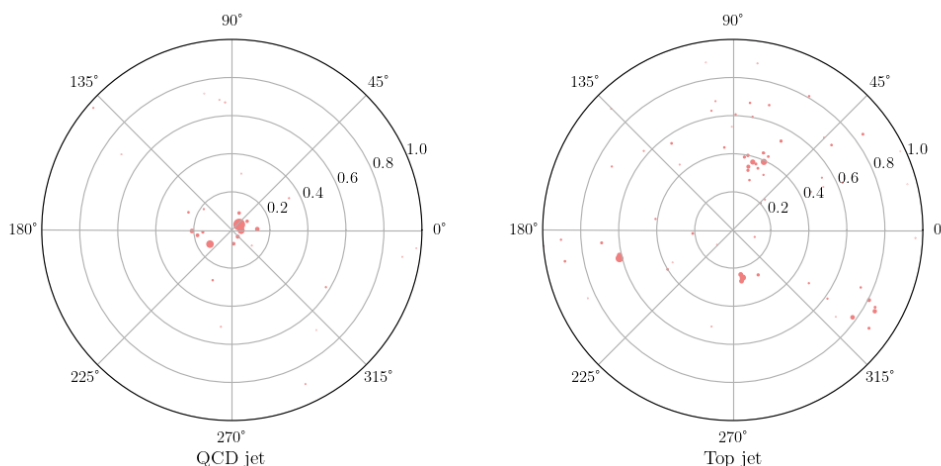
To guarantee a good definition in perturbative QCD calculation, any jet substructure observable should be both infrared and collinear safe (IRC-safe) and Lorentz invariant. These properties can be achieved by expressing an observable as a linear combination of calorimetric correlators (C-correlators) [1] with a general form

$$C_N^{f_N} = \sum_{i_1=1}^M \dots \sum_{i_N=1}^M E_{i_1} \dots E_{i_N} f_N(\{\hat{p}_{i_1}, \dots, \hat{p}_{i_N}\}), \quad (2.1)$$

where  $E_i$  and  $\hat{p}_i$  are respectively energy (or energy fraction) and direction of the  $i$ -th constituent particle within the jet,  $M$  is the total number of the constituent particles inside the jet,  $N$  is any positive integer, and  $f_N$  is any sufficiently smooth permutation invariant function of its  $N$  arguments. The multi-particle energy correlator forms are naturally derived from quantum field theory. The function  $f_N$  responsible for maintaining the calorimetric continuity depends solely on the directions of particles. Numerous efforts have been made to define various forms of function  $f_N$  as variants of the energy-flow based observables. The Deep Set Theorem [24] indicates that the functions operating on point sets should be permutation invariant to the order of objects in the set. By further enforcing the IRC-safety, any observable  $\mathcal{O}$  of a variable-length point set  $\{p_1, \dots, p_M\}$  can be approximated arbitrarily well by [18]

$$\mathcal{O}(\{p_1, \dots, p_M\}) = F\left(\sum_{i=1}^M z_i \Phi(\hat{p}_i)\right), \quad (2.2)$$

where  $z_i$  are energy fraction  $E_i/(\sum E_i)$ ,  $\Phi$  and  $F$  are arbitrary functions. Figure 1 presents a visualization of point cloud representation of a jet in polar coordinate. To enhance the flexibility of the representation, the Energy Flow Networks [18] have been introduced, which directly replace the fixed form of expansion polynomials with parameterized and



**Figure 1.** The visualizations of point cloud representation for a QCD jet and a top jet in polar coordinate. The jet radius is set to 1 and each red point denotes the position of a constituent particle within the jet.

learnable filters  $\Phi$  in neural networks. The universal approximation theorem of neural networks supports this formulation, making it more adaptable and versatile. There are also some attributes that utilize a systematic and complete basis  $\Phi^l$  for angular directions to decompose complete features of the jet global observables, including Zernike polynomials [2] and spherical harmonics [25].

For a jet with  $N$ -prong substructure, it is also efficient to define the  $N$ -subjettiness [3] to measure the energy radiation alignment with  $N$  candidate subjects:

$$\tau_N^\beta = \sum_i E_i \min\{\Delta R_{i,1}^\beta, \Delta R_{i,2}^\beta, \dots, \Delta R_{i,N}^\beta\}, \quad (2.3)$$

where  $\Delta R_{i,j}$  represents the angular separation between  $i$ -th and  $j$ -th candidate subjects in the angular plane. The parameter  $\beta$  acts as the angular exponent, enabling the capture of the  $3M - 4$  dimensional representation of the  $M$ -body phase space [26]. To determine the axes of the fixed number of candidate subjects, the exclusive- $k_T$  algorithm, similar to the weighted  $k$ -mean algorithm, is used to implement the minimization process. Additionally, several specialized approaches for evaluating  $N$ -prong structures of a jet are proposed. One method is the  $N$ -point energy correlation functions (ECFs) [5], which are defined as

$$\text{ECF}(N, \beta) = \sum_{i_1 < i_2 < \dots < i_N \in J} \left( \prod_{a=1}^N E_{i_a} \right) \left( \prod_{b=1}^{N-1} \prod_{c=b+1}^N R_{i_b i_c} \right)^\beta, \quad (2.4)$$

where  $R_{i_b i_c}$  is the angular distance between particles  $i_b$  and  $i_c$ . Another method, known as the generalized energy correlation functions (ECFGs) [6] use minimizing steps in  $N$ -subjettiness and compare its superiority. Moreover, the energy flow polynomials (EFPs) [7] provide a systematic and organized method for calculating a more (over)complete basis of

jet substructure, with a topological correspondence to multi-graphs:

$$\text{EFP}_G = \sum_{i_1=1}^M \dots \sum_{i_N=1}^M E_{i_1} \dots E_{i_N} \prod_{(k,l) \in G} R_{i_k i_l}. \quad (2.5)$$

These EFPs have been demonstrated to compose existing physics-inspired observables like jet mass, planar flow [10] and the above-mentioned ECFs (corresponding with complete graph). Compared with the moment expansion representation including jet image mentioned above, those higher-point energy flow bases are more complete and show clearer correspondence to the physically well-defined observables. Note that the use of the Euclidean distance metric in the energy flow basis indicates the rotation symmetry of a single jet. However, when we generalize this to the construction of global event features, like jet pull [27] and other color flow variables, with higher-point energy flow bases, directional functions  $f_N$  of rapidity and azimuth directions should be defined separately.

The systematic comparison studies demonstrate the discrimination performance of the energy correlation functions for  $N$ -prong jets under various angular exponent  $\beta$  [5]. In this study, we propose a method to modify  $R_{ij}^\beta$  in the EFPs to a complete and orthogonal basis, such as Legendre polynomials, to parameterize and vectorize the angular distance function relationship. This approach allows us to effectively combine the influence of different angular exponents. Furthermore, we would like to point out that it may not be necessary to strictly adhere to the formula where the function  $f_N$  in the C-correlator is solely dependent on particle directions to ensure IRC safety. By liberating from this constraint, the possible formulations of jet observables are greatly expanded to beyond the linear combination of C-correlators.

## 2.2 Hierarchical energy flow functions

Now we consider a naive 2-point energy correlation function  $\text{ECF}(2, \beta)$

$$\text{ECF}(2, \beta) = \sum_{i,j} z_i z_j R_{ij}^\beta. \quad (2.6)$$

We could generalize the function of angular distance between two particles  $R_{ij}$  as any function  $f(R_{ij})$ , and then expand it with orthogonal polynomials like Legendre polynomials  $P_\beta(\theta_{ij})$ . We set  $\theta_{ij} = R_{ij}/R_0 - 1$  to make  $\theta_{ij}$  in the domain  $[-1, +1]$  to keep numerical stability. Then the Legendre-based 2-point energy correlation function  $\text{ECF}(2, P_\beta)$  can be written as

$$\text{ECF}(2, P_\beta) = \sum_{i,j} z_i z_j P_\beta(\theta_{ij}). \quad (2.7)$$

However, it has been pointed out that the higher-point energy correlation functions pose challenges due to their power-growth computational complexity [5]. The computational complexity of  $N$ -point Energy Correlation Functions (ECFs) grows significantly, denoted as  $\mathcal{O}(M^N)$ , where  $M$  represents the number of particles inside the jet. And it was shown that linear relationships between EFPs that hold for quark and gluon jets to a specific order could be obtained by applying power counting [28]. Fortunately, some special  $N$ -point Energy

Flow Polynomials (EFPs) associated with tree graphs could be computed by utilizing the Variable Elimination (VE) algorithm [7] with a reduced complexity of  $\mathcal{O}(M^2)$ . Here we select a subset from those special EFPs corresponding to the path graphs, which are the simplest example of tree graphs. Therefore, the path-graph corresponding EFPs can be expressed in a hierarchical summation form. For example, the 2-point, 3-point and 4-point path-graph Hierarchical Energy Flow Polynomials (HEFPs) are defined as

$$\text{HEFP}(2, \beta) = \sum_i^M z_i \left( \sum_j^M z_j P_\beta(\theta_{ij}) \right), \quad (2.8)$$

$$\text{HEFP}(3, \{\beta_1, \beta_2\}) = \sum_i^M z_i \left( \sum_j^M z_j P_{\beta_2}(\theta_{ij}) \left( \sum_k^M z_k P_{\beta_1}(\theta_{jk}) \right) \right), \quad (2.9)$$

$$\text{HEFP}(4, \{\beta_1, \beta_2, \beta_3\}) = \sum_i^M z_i \left( \sum_j^M z_j P_{\beta_3}(\theta_{ij}) \left( \sum_k^M z_k P_{\beta_2}(\theta_{jk}) \left( \sum_l^M z_l P_{\beta_1}(\theta_{kl}) \right) \right) \right) \quad (2.10)$$

The compact formulas of path-graph  $(N + 1)$ -point HEFPs are

$$\text{HEFP}(N + 1, \{\beta_1, \dots, \beta_N\}) = \sum_{i_1, i_2, \dots, i_{N+1}}^M \left( \prod_{a=1}^{N+1} z_{i_a} \right) \left( \prod_{b=1}^N P_{\beta_{i_b}}(\theta_{i_b i_{b+1}}) \right). \quad (2.11)$$

They could be seen as an energy-weighted summation of  $N$ -point particle features:

$$\text{HEFP}(N + 1, \{\beta_1, \dots, \beta_N\}) = \sum_i^M z_i \hat{p}_i^N(\{\beta_1, \dots, \beta_N\}), \quad (2.12)$$

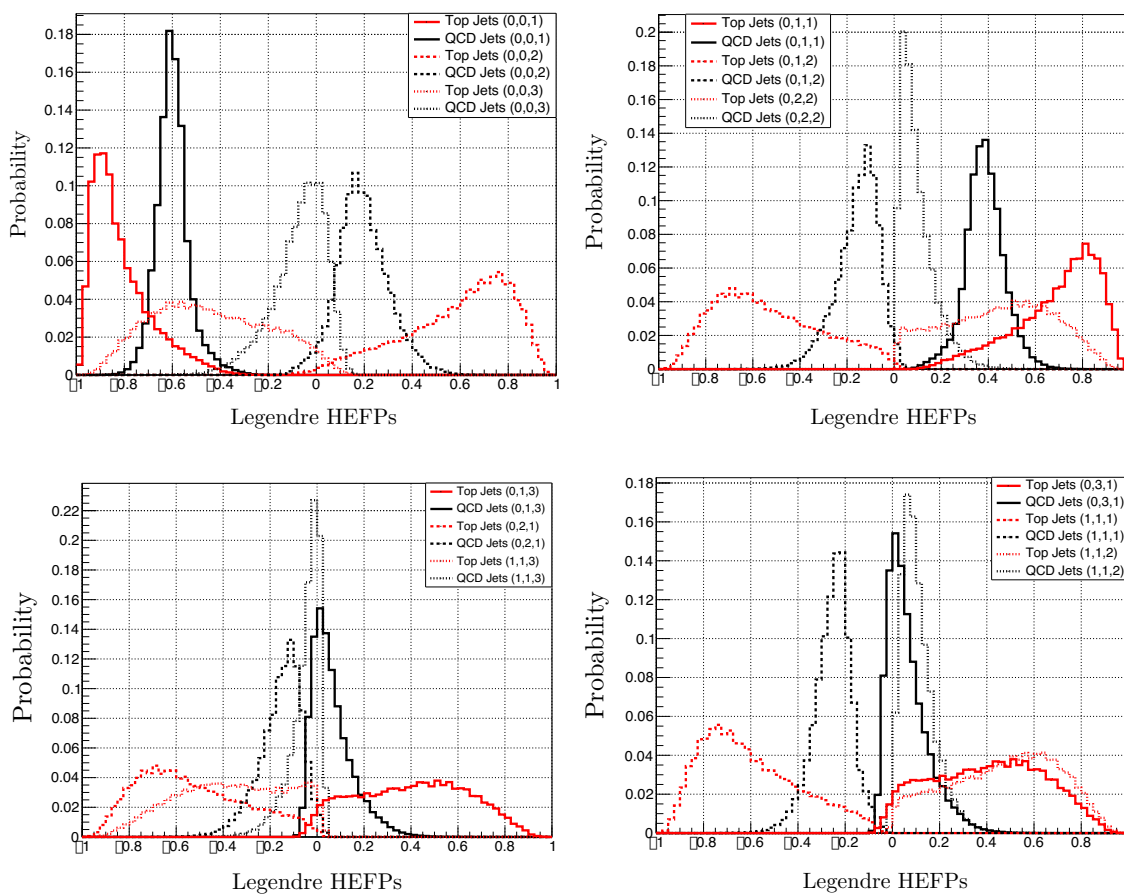
where  $\hat{p}_i^N(\{\beta_1, \dots, \beta_N\})$  are obtained through recursive relations as

$$\hat{p}_i^{t+1}(\{\beta_1, \dots, \beta_{t+1}\}) = \sum_j^M z_j \hat{p}_j^t(\{\beta_1, \dots, \beta_t\}) P_{\beta_{t+1}}(\theta_{ij}), \quad (2.13)$$

with  $(t + 1)$ -point particle features  $\hat{p}_i^{t+1}$  being an energy-weighted summation of all  $t$ -point particle features  $\hat{p}_j^t$  multiplied by a pair-wise distance related factor  $P_{\beta_{t+1}}(\theta_{ij})$ .

The above path-graph HEFPs are linear combinations of multi-graph corresponding EFPs. Nevertheless, the integration of Legendre polynomials in our approach offers several crucial advantages. These polynomials serve as a powerful and flexible tool for representing any complex angular distance functions,  $f(\theta_{ij})$ , in a compact and numerically stable manner. By expanding  $f(\theta_{ij})$  using Legendre polynomials  $P_\beta(\theta_{ij})$ , we can effectively capture the essential features of the angular distribution while controlling the model complexity. Notably, the orthogonality property of Legendre polynomials guarantees that each coefficient in the expansion corresponds to a distinct feature, enhancing the interpretability of the model. Additionally, by transforming  $R_{ij}$  to the range of  $\theta_{ij} \in [-1, +1]$ , we effectively restrict the absolute value of Legendre polynomials  $P_\beta(\theta_{ij}) \in [0, 1]$ , further improving the model's robustness.

In figure 2, we present the distributions of 4-point Legendre Hierarchical Energy Flow polynomials, denoted as  $\text{HEFP}(4, \{\beta_1, \beta_2, \beta_3\})$ , for both top jets and QCD jets, spanning across different parameter settings  $(\{\beta_1, \beta_2, \beta_3\})$ . Remarkably, the distributions of all the  $\text{HEFP}(4, \beta_1, \beta_2, \beta_3)$  of top jets and QCD jets exhibit clear discrimination. This distinct

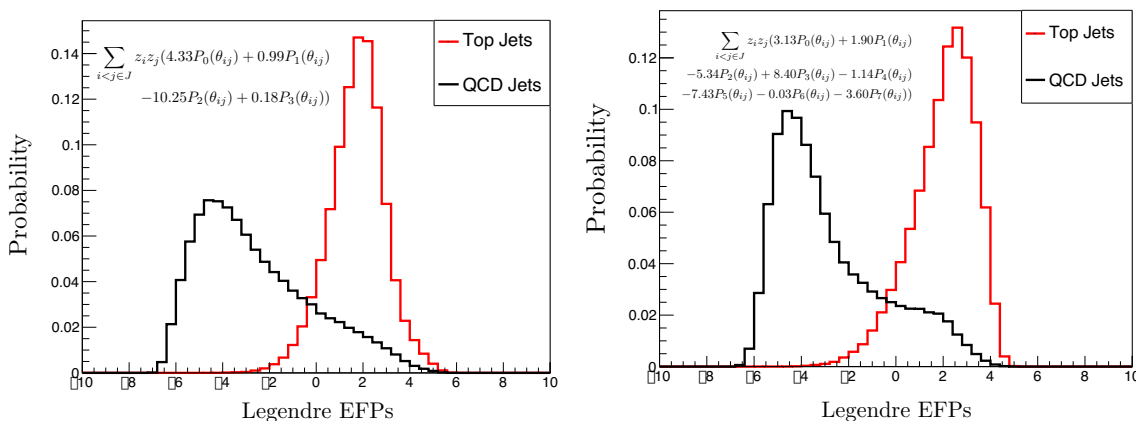


**Figure 2.** Distributions of 4-point Legendre Hierarchical Energy Flow polynomials HEFP(4,  $\{\beta_1, \beta_2, \beta_3\}$ ) under various settings of  $(\{\beta_1, \beta_2, \beta_3\})$  for top jets and QCD jets.

separation underscores the discriminative power of the HEFP approach and highlights its potential significance in distinguishing top jets from QCD jets in high energy physics analyses. Moreover, since  $P_0(\theta_{ij}) = 1$ , HEFP(4,  $\{0, \beta_2, \beta_3\}$ ) will degenerate into 3-point HEFP(3,  $\{\beta_2, \beta_3\}$ ), while HEFP(4,  $\{0, 0, \beta_3\}$ ) will degenerate into 2-point HEFP(2,  $\{\beta_3\}$ ). Therefore, the  $(N + 1)$ -point HEFPs encompass all the HEFPs ranging from 2-point to  $N$ -point. As  $\sum_i \beta_i$  increases, the HEFPs gradually move closer to the  $y$ -axis, indicating an increasing fraction of HEFPs that approach zero. This is primarily due to the cancellation effects observed in the higher order components. Consequently, for a specific truncation order  $\beta_{\max}$ , any EFP observable with a path graph structure can be accurately expanded into a linear combination of  $\beta_{\max}^N$  HEFPs. For instance, we can easily employ the Linear Logistic Regression to identify the specific form of the 2-point HEFP(2, f) with the optimal classification performance. The number of undetermined parameters  $\{\alpha_i\}$  is  $\beta_{\max}$ , and we define the probability of the input jet being tagged as the top jet as

$$p = \sigma \left( \sum_{\beta} \alpha_{\beta} \sum_{i,j} z_i z_j P_{\beta}(\theta_{ij}) \right), \tag{2.14}$$





**Figure 3.** Normalized distributions of the 2-point Legendre Energy Flow Polynomials for top jets and QCD jets that achieved the best classification performance. The corresponding truncation orders  $\beta_{\max}$  are 4 and 8, respectively. The specific form of the 2-point Legendre Energy Flow Polynomials is also indicated on the figure.

where  $\sigma(t) = 1/(1+e^{-t})$  is the logical function. The objective function of logistic regression is

$$J(\alpha) = -[y \log(p) + (1 - y) \log(1 - p)], \quad (2.15)$$

where  $y = 1$  corresponds to the top jet and  $y = 0$  corresponds to the QCD jet. In figure 3, we present the normalized distributions of the 2-point Legendre Energy Flow Polynomials for top jets and QCD jets that achieved the optimal classification performance. The truncation orders  $\beta_{\max}$  for top jets and QCD jets are 4 and 8, respectively. Additionally, we illustrate the specific form of the 2-point Legendre Energy Flow Polynomials in this figure, which serves as a bridge between physical interpretability and machine learning. As shown in figure 3, the distinctive form of the 2-point Legendre Energy Flow Polynomials, obtained through linear regression, proves to be highly effective in discriminating between top jets and QCD jets.

Furthermore, we find that the above HEFP(N,  $P_\beta$ ) obtained under a fixed energy-weighted summation order exhibits IRC-safety after each node elimination. This special characteristic, which we called hierarchical IRC-safety, indicates that inserting arbitrary functions (including non-linear function) after each energy-weighted summation is allowed, where the parameterized and trainable neural networks can come into and play an important role. For example,

$$\mathcal{O}^{(2)} = \sum_i^M z_i \Phi \left( \sum_j^M z_j P_\beta(\theta_{ij}) \right) \quad (2.16)$$

where 2 indicates that the features are 2-point energy flow observable. After performing variable elimination algorithm, any EFPs can be computed via a hierarchical summation like

$$\begin{aligned} & \sum_{i,j,k,l}^M z_i z_j z_k z_l f_1(\theta_{ij}) f_2(\theta_{jk}) f_3(\theta_{jl}) f_4(\theta_{kl}) \\ &= \sum_{\beta_1, \beta_2, \beta_3, \beta_4}^{\beta_{\max}} \alpha_{\beta_1} \alpha_{\beta_2} \alpha_{\beta_3} \alpha_{\beta_4} \left( \sum_i^M z_i \left( \sum_j^M z_j P_{\beta_1}(\theta_{ij}) \left( \sum_k^M z_k P_{\beta_2}(\theta_{jk}) \left( \sum_l^M z_l P_{\beta_3}(\theta_{kl}) P_{\beta_4}(\theta_{jl}) \right) \right) \right) \right), \end{aligned}$$

where  $\alpha_\beta$  are linear coefficients of orthogonal bases with different orders. Moreover, an arbitrary function  $\Phi$  can be inserted IRC-safely between each energy-weighted summation and the next layer calculation, which can be effectively parameterized and trained by using neural networks. In the proceeding section, we will introduce a novel backbone model that utilizes neural network parameterization to reconstruct the path-graph energy flow polynomials while simultaneously maintaining interpretability, IRC-safety, and rotational invariance. In this work, we only consider the path-graph corresponding EFPs. For simplicity, all the Hierarchical Energy Flow functions specifically refer to path-graph Hierarchical Energy flow functions in the following sections. However, it is essential to highlight that this framework actually can be generalized to any other complex graph structures.

### 3 Hierarchical higher-point Energy Flow Network

#### 3.1 Hierarchical Energy Flow Network

As mentioned in the preceding section, inserting an arbitrary function after the energy-weighted summation of Hierarchical Energy Flow Polynomials does not break the IRC safety. Thus we can parameterize these functions with neural networks. For instance, in the case of 2-point HEFN, the 1-point features of  $i$ -th particle  $\hat{p}_i^1$  is embedded initially as

$$\hat{p}_i^1 = \Phi\left(\sum_j^M z_j P_\beta(\theta_{ij})\right), \tag{3.1}$$

where  $\hat{p}_i^1$  are invariant under translation, rotation, or reflection of the jet in the angular direction. Consequently, the jet observables in the latent space are also IRC-safe and can be expressed as

$$\mathcal{O}^{(2)} = F\left(\sum_i^M z_i \Phi\left(\sum_j^M z_j P_\beta(\theta_{ij})\right)\right), \tag{3.2}$$

which are natural generalizations of EFNs when considering 2-point energy flow functions. We introduce two MLP modules  $\Phi$  and  $F$  as EFNs do. The first MLP  $\Phi : \mathcal{R}^{\beta_{\max}} \rightarrow \mathcal{R}^l$  maps 1-point particle features with dimension  $\beta_{\max}$ , the truncated order of the Legendre polynomials, into a latent space with dimension  $l$ , while the 2-point jet observables are energy-weighted summation of all particle features. The second MLP can be viewed as a discriminant of jet features. We find better tagging performance in some public datasets compared with EFNs, as discussed in section 4.

In the following steps, we extend the scenario to  $N$ -point HEFN. We incorporate neural networks into Hierarchical Energy Flow Polynomials to obtain Hierarchical Energy Flow Networks (HEFNs). For  $t$ -point particle features at latent space  $\hat{p}_i^t$ , the next higher-point particle features are obtained through recursive relations as

$$\hat{p}_i^{t+1} = \Phi^a\left(\sum_j^M z_j \Phi^b(\hat{p}_j^t) \otimes P_\beta(\theta_{ij})\right) + \hat{p}_i^t. \tag{3.3}$$

The second MLP  $\Phi^b : \mathcal{R}^l \rightarrow \mathcal{R}^d$  maps  $t$ -point particle features  $\hat{p}_j^t$  into dimension- $d$ . And we fix the dimension of hidden space  $l' = d \times \beta_{\max}$  to control model parameter capability for

different  $\beta_{\max}$  setting. The first MLP  $\Phi^a : \mathcal{R}^{l'} \rightarrow \mathcal{R}^l$  maps the hidden  $(t + 1)$ -point particle features back to latent space. Besides, the residual connection we introduce here makes

$$\hat{p}_i^{t+1} = \sum_{k=1}^t \Phi^{a,k} \left( \sum_j^M z_j \Phi^{b,k}(\hat{p}_j^k) \otimes P_\beta(\theta_{ij}) \right), \quad (3.4)$$

which could let the final jet observables contain a mix of multiple high-point energy flow functions

$$\mathcal{O}^{(N+1)} = F \left( \sum_i^M z_i \sum_{k=1}^N \Phi^{a,k} \left( \sum_j^M z_j \Phi^{b,k}(\hat{p}_j^k) \otimes P_\beta(\theta_{ij}) \right) \right). \quad (3.5)$$

It is worth noting that since the inserted nonlinear functions are not only about angular information, the learned observables cannot be expanded as a linear combination of C-correlators nor EFPs. In the generalized scenario of an  $N$ -point Hierarchical Energy Flow Network before the final discriminant  $F$ , the architecture consists of total  $N$   $\Phi^a$  and  $(N - 1)$   $\Phi^b$  since we only apply  $\Phi^a$  to get  $\hat{p}_i^1$ . Both  $\Phi^a$  and  $\Phi^b$  are MLPs with BatchNorm-ReLU-FullyConnected structures. Note that when applying Batch Normalization to particle-level features, we start by obtaining jet-level observables through energy-weighted summation. Subsequently, we calculate the mean and variance of jet features in mini-batches. This ensures that Batch Normalization maintains the statistical robustness of neural network inputs, even under particle splitting and soft radiation conditions.

In our study, we keep the dimension of particle features at latent space  $l = 256$  and hidden space  $l' = d \times \beta_{\max} = 1024$ . The first MLP  $\Phi^a$  is stacked with two BN-ReLU-FC blocks with  $(1024, 256, 256)$  nodes, while the second MLP  $\Phi^b$  is a single BN-ReLU-FC block with  $(256, d)$  nodes. We selected  $\beta_{\max}$  values of 4, 8, 16, and 32 as the truncated orders of the Legendre polynomial to investigate their impact on the model's performance. Additionally, we systematically compared the performance of up to 2, 3, and 4-point HEFN (Hierarchically Energy Flow Network) on top tagging dataset and quark/gluon discrimination dataset. All the results will be presented in section 4.

### 3.2 Local Hierarchical Energy Flow Network

To further reduce the computational complexity of the model while preserving essential information, we ignore the contribution of distant particles and adopt the neighbour aggregation approach instead of global summation. In other words, we require  $f(R_{ij}) = 0$  if  $R_{ij} > r_{\max}$ . Besides, it is also possible to pixelate the distance between particles to aggregate particles in different regions [29]. Based on the neighbour aggregation, we propose the second strategy, the Local Hierarchical Energy Flow Network (LHEFN).

In our case, we choose to use the query ball approach to define neighborhood for IRC-safety during end-to-end training [19]. To combine features from different scales, we group multi-scale neighborhoods to construct the graph structure  $A_{ij}^k$  as

$$A_{ij}^k = \begin{cases} 1 & \text{if } (k - 1)\epsilon \leq R_{ij} < k\epsilon \\ 0 & \text{else} \end{cases} \quad (3.6)$$

where  $\epsilon$  measures the resolution of our implementation. The value of  $r_{\max} = k_{\max}\epsilon$  represents the maximum scope radius.

We initialize 1-point particle features similar with HEFN:

$$\hat{p}_i^1 = \Phi\left(\sum_j^M z_j A_{ij}^k\right). \tag{3.7}$$

Currently, we continue to iteratively update the particle features following eq. (3.3). However, instead of relying on pair-wise features computed using Legendre polynomials, we now adopt  $A_{ij}^k$  to replace them. Note that  $A_{ij}^k$  are symmetric edge interaction tensors resulting in a radial single square wave form. Since the edge embeddings are binary functions, the nodes can be updated using a simple weighted summation operator, significantly reducing computational complexity. The update step of  $t$ -point particle features can be expressed as

$$\hat{p}_i^{t+1} = \Phi([\hat{p}_{i,1}^t, \hat{p}_{i,2}^t, \dots, \hat{p}_{i,k_{\max}}^t]), \tag{3.8}$$

where  $\hat{p}_{i,k}^t = \sum_j^{A_{ij}^k=1} z_j \hat{p}_j^t$  is the energy-weighted sum of all particle features between distances  $k\epsilon$  and  $(k+1)\epsilon$  away from particle  $i$ . This computation efficiently aggregates relevant information from neighboring particles, taking into account their energy contributions. The neural network parameterized function  $\Phi$  plays a crucial role in adaptively selecting the aggregation radius, and combines features of particles on different neighbor regions. Note that this approach may exhibit limited flexibility in expressing edge information, resulting in potentially worse performance compared to the Legendre polynomial encoding. However, the aforementioned formula significantly reduces the computational complexity of  $O(d \cdot M^2 \cdot \beta)$  compared with eq. (3.4) by eliminating the need for the direct product of edge features and particle features. The model gains a remarkable boost in speed, making it more practical for real applications. To ensure fair evaluations, we reset  $r_{\max} = 0.5R_0, R_0, 1.5R_0$  and  $2R_0$  and selected  $k_{\max}$  of 4, 8, 16, and 32 for testing purposes. Additionally, we maintained consistency in the remaining trainable parameter capacity and hyperparameter settings throughout the experiments.

### 3.3 Model implementation

The model architecture is implemented in the PYTORCH deep learning framework with the CUDA platform. We adopt the binary cross-entropy as the loss function. To optimize the model parameters, we employ the Adam optimizer [30] with an initial learning rate of 0.001 and momentum set to 0.8, which is determined based on the gradients calculated on a mini-batch of 128 training examples. The network is trained up to 50 epochs, with a cosine decay learning rate scheduler. In addition, we employ the early-stopping technique to prevent over-fitting.

## 4 Results of jet classification

### 4.1 Top tagging

We perform the top tagging analysis utilizing a benchmark dataset [31], containing hadronic tops as the signal and QCD di-jets as the background. The event generation is performed

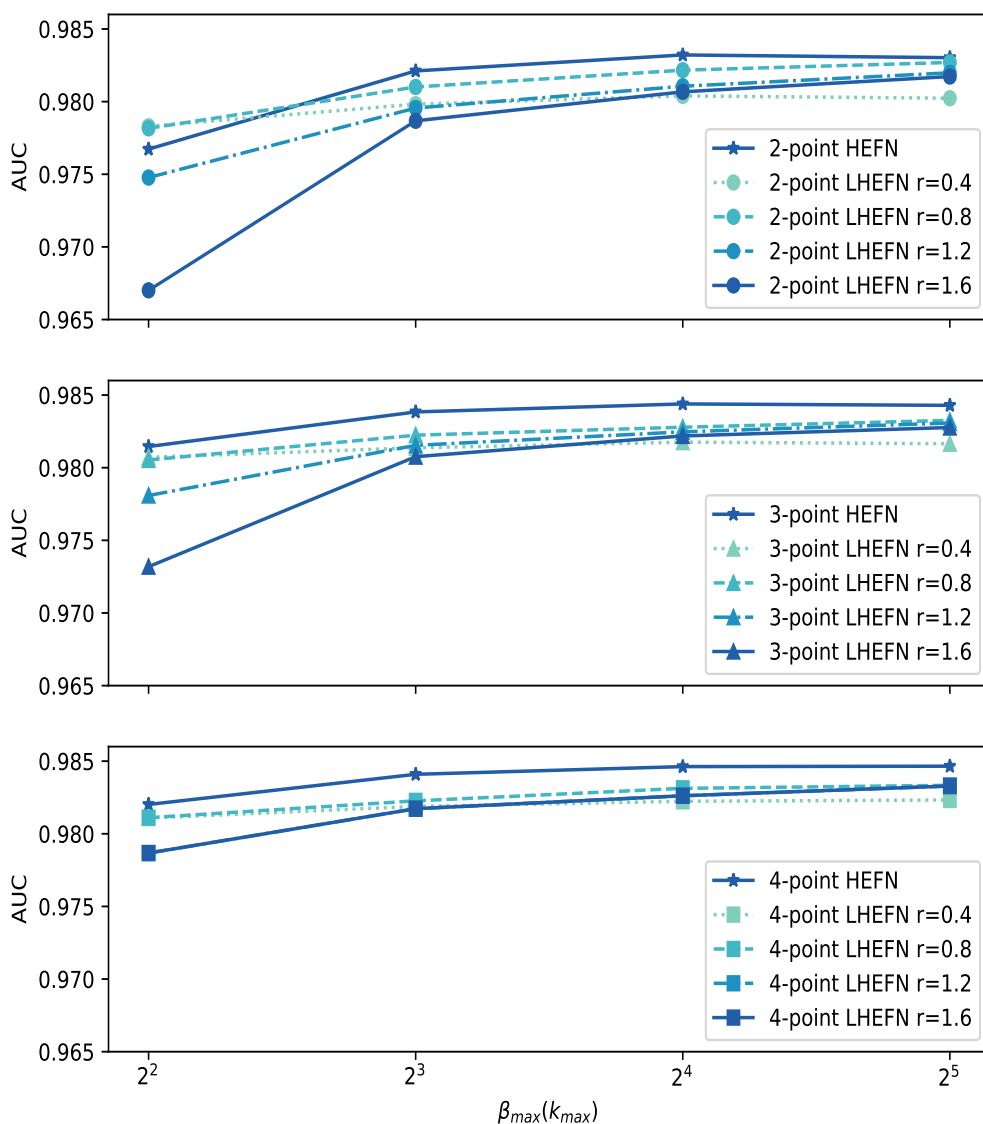
by Pythia8 [32], while the detector effect is simulated by Delphes [33]. The particle-flow constituents are clustered into jets using the anti-kT algorithm [34] with  $R_0 = 0.8$  as the radius parameter. Our analysis focuses on jets with transverse momentum  $p_T \in [550, 650]$  GeV and rapidity  $|y| < 2$ . The dataset consists of 1.2M training events, 400k validation events, and 400k test events. Only the energy-momentum 4-vectors for each particle inside the jets are contained. Both the HEFN architecture and the LHEFN architecture utilize only  $p_T$ ,  $\eta$ , and  $\phi$  of all constituent particles inside each jet. To ensure dataset quality, jets with a particle count of less than 5 were removed to prevent the influence of extreme examples on the datasets (0.01% for the top tagging dataset).

Figure 4 shows the area under the ROC curve (AUC) for the top tagging task of 2-point, 3-point, and 4-point HEFN as a function of  $\beta_{\max}$ , along with the AUC for the top tagging task of 2-point, 3-point, and 4-point LHEFN with varying  $r$  values ( $r = 0.4, 0.8, 1.2, 1.6$ ) as a function of  $k_{\max}$ . The results displayed in figure 4 exhibit a notable trend where the Area Under the Curve (AUC) of HEFN consistently increases with  $N$ . This observation indicates that increasing the complexity of the network by adding more points (2-point, 3-point, and 4-point configurations) can enhance the performance of HEFN. Additionally, the AUC remains almost unchanged during the transition from 3-point to 4-point configurations, which indicates that the 4-point HEFN achieves the optimal performance. For all the considered 2-point, 3-point, and 4-point HEFN architectures, we notice a consistent improvement in AUC as we increase the truncated orders of the Legendre Polynomials  $\beta_{\max}$ . Similar behavior is observed in the case of 2-point, 3-point, and 4-point LHEFN models with varying  $r$  values (0.4, 0.8, 1.2, and 1.6). As the parameter  $k_{\max}$  increases, the AUC of LHEFN also demonstrates a notable enhancement, signifying that refining the neighbor division leads to a higher AUC. Another noteworthy discovery is that setting the value of  $r$  empirically to half the radius of the candidate jets yields optimal performance for LHEFN. Although it is worth noting that HEFN outperforms LHEFN slightly across almost all parameter settings, the LHEFN is notably faster compared to HEFN. The choice between the two models may depend on specific application requirements, with HEFN offering higher AUC and LHEFN presenting a notable speed advantage.

In table 1 we provide detailed results of accuracy, AUC, and background rejection for the optimal parameter settings ( $N = 4, \beta_{\max} = 16$ ) and ( $N = 4, k_{\max} = 16$ ) of both HEFN and LHEFN. Additionally, we present the performance achieved by various classification algorithms on the top tagging dataset, facilitating a comprehensive comparison. From table 1, it is evident that by solely utilizing the  $p_T$ ,  $\eta$ , and  $\phi$  information of all constituent particles, both HEFN and LHEFN deliver comparable performance to existing model architectures. Moreover, both HEFN and LHEFN enable the reconstruction of energy correlation-based observables while maintaining interpretability, IRC-safety, and rotational invariance, which other models can not preserve.

## 4.2 Quark/gluon discrimination

The Quark-Gluon benchmark dataset [18], generated using Pythia8 without detector simulation, consists of quark-initiated samples  $q\bar{q} \rightarrow Z \rightarrow \nu\bar{\nu} + (u, d, s)$  as the signal and gluon-initiated data  $q\bar{q} \rightarrow Z \rightarrow \nu\bar{\nu} + g$  as the background. For jet clustering, we use the



**Figure 4.** Area under the ROC curve (AUC) for top tagging task of 2-point, 3-point, and 4-point HEFN as a function of  $\beta_{\max}$ , and AUC for top tagging task of 2-point, 3-point, and 4-point LHEFN with varying  $r$  values ( $r = 0.4, 0.8, 1.2, 1.6$ ) as a function of  $k_{\max}$ , where  $r$  represents the maximum scope radius for LHEFN. The  $x$ -axis is shared for both  $\beta_{\max}$  and  $k_{\max}$ .

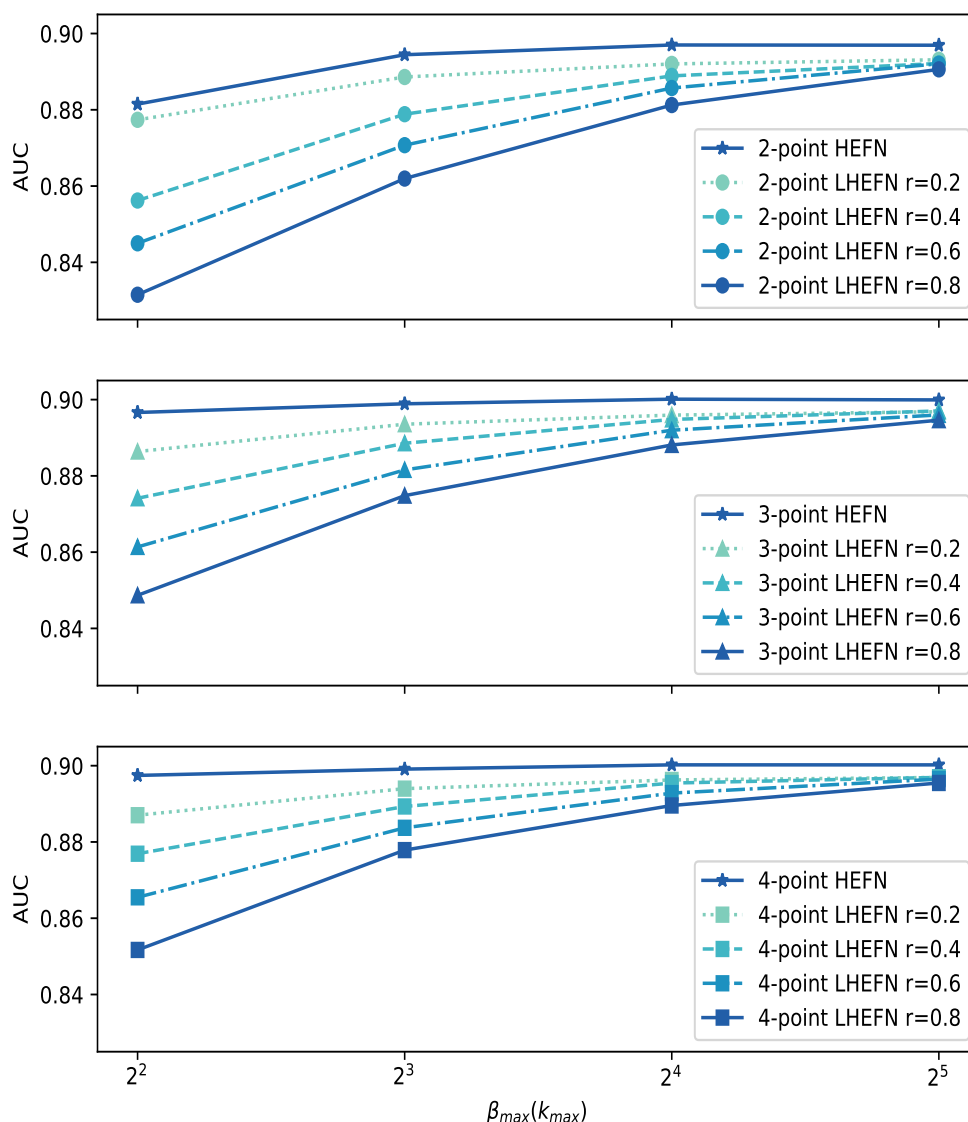
anti- $k_T$  algorithm with  $R_0 = 0.4$ . Our analysis focuses on selecting jets with transverse momentum  $p_T \in [500, 550]$  GeV and rapidity  $|y| < 1.7$ . Each particle in the dataset is characterized by its four-momentum and particle identification (PID) information. However, in this study, both the HEFN and LHEFN architectures utilize only  $p_T$ ,  $\eta$ , and  $\phi$  of all the constituent particles inside each jet. The official dataset comprises of a total of 2 million events. Among these, 1.6 million events are used for training, while 200k events each are allocated for validation and testing. Besides, to ensure dataset quality, jets with a particle count of less than 5 are removed to avoid any adverse influence of extreme examples on the datasets. The proportion of these filtered samples relative to the entire dataset is minimal, representing only 0.01% of the Quark/Gluon Discrimination dataset.

|                             | Accuracy      | AUC           | $1/\epsilon_B (\epsilon_S = 0.5)$ | $1/\epsilon_B (\epsilon_S = 0.3)$ |
|-----------------------------|---------------|---------------|-----------------------------------|-----------------------------------|
| ResNeXt-50 [20]             | 0.936         | 0.9837        | 302±5                             | 1147±58                           |
| P-CNN [20]                  | 0.930         | 0.9803        | 201±4                             | 759±24                            |
| PFN [18]                    | -             | 0.9819        | 247±3                             | 888±17                            |
| ParticleNet-Lite [20]       | 0.937         | 0.9844        | 325±5                             | 1262±49                           |
| ParticleNet [20]            | 0.940         | 0.9858        | 397±7                             | 1615±93                           |
| JEDI-net [35]               | 0.9263        | 0.9786        | -                                 | 590.4                             |
| JEDI-net with $\sum O$ [35] | 0.9300        | 0.9807        | -                                 | 774.6                             |
| SPCT [36]                   | 0.928         | 0.9799        | 201±9                             | 725±54                            |
| PCT [36]                    | 0.940         | 0.9855        | 392±7                             | 1533±101                          |
| LorentzNet [22]             | 0.942         | 0.9868        | 498±18                            | 2195±173                          |
| ParT [37]                   | 0.940         | 0.9858        | 413±16                            | 1602±81                           |
| <b>HEFN</b>                 | <b>0.9375</b> | <b>0.9846</b> | <b>343±6</b>                      | <b>1262±51</b>                    |
| <b>LHEFN</b>                | <b>0.9337</b> | <b>0.9833</b> | <b>271±5</b>                      | <b>935±21</b>                     |

**Table 1.** Performance comparison of HEFN and LHEFN with Existing Classification Algorithms on the Top Tagging Dataset. The uncertainty is calculated by taking the standard deviation of 5 training runs with different random weight initialization.

Figure 5 shows the area under the ROC curve (AUC) for the quark/gluon discrimination task of 2-point, 3-point, and 4-point HEFN as a function of  $\beta_{\max}$ , along with the AUC for the quark/gluon discrimination task of 2-point, 3-point, and 4-point LHEFN with  $r = 0.2, 0.4, 0.6, 0.8$  as a function of  $k_{\max}$ , where  $r$  is the maximum scope radius of LHEFN. The results presented in figure 5 show a clear pattern: the Area Under the Curve (AUC) of HEFN consistently increases with  $N$ , which suggests that the performance of HEFN benefits from the increased complexity brought by adding more points. Moreover, during the shift from 3-point to 4-point configurations, the AUC remains relatively stable, implying that the 4-point HEFN reaches its optimal performance at this stage. Regarding the 2-point HEFN architecture, we observe that increasing the truncated orders of the Legendre Polynomials  $\beta_{\max}$  leads to a consistent improvement in the Area Under the Curve (AUC). However, for the 3-point and 4-point HEFN architectures, the AUC remains unchanged as  $\beta_{\max}$  increases. Different from HEFN, in the case of 2-point, 3-point, and 4-point LHEFN models with  $r = 0.2, 0.4, 0.6, 0.8$ , as the parameter  $k_{\max}$  increases, there is a notable enhancement in the AUC of LHEFN, indicating that refining the neighbor division can result in higher AUC. Another noteworthy finding is that setting the value of  $r_{\max}$  to half the radius of the candidate jets results in optimal performance for LHEFN. Although it is worth noting that HEFN performs slightly better than LHEFN across all parameter settings, LHEFN stands out for its significantly faster performance compared to HEFN. As a result, the selection between the two models may depend on the specific needs of the application, where HEFN is preferred for achieving higher AUC, while LHEFN is a more suitable choice when speed is a critical factor.

In table 2, we present the results of accuracy, area under the curve (AUC), and background rejection for the parameter settings ( $N = 4, \beta_{\max} = 16$ ) and ( $N = 4, k_{\max} = 16$ )



**Figure 5.** Area under the ROC curve (AUC) for quark/gluon discrimination task of 2-point, 3-point, and 4-point HEFN as a function of  $\beta_{\max}$ , and AUC for top tagging task of 2-point, 3-point, and 4-point LHEFN with varying  $r$  values ( $r = 0.2, 0.4, 0.6, 0.8$ ) as a function of  $k_{\max}$ , where  $r$  represents the maximum scope radius for LHEFN. The  $x$ -axis is shared for both  $\beta_{\max}$  and  $k_{\max}$ .

of HEFN and LHEFN which correspond to the optimal performance. Additionally, we present the performance attained by several classification algorithms on the quark and gluon dataset, enabling a comprehensive comparison. From table 1, we can see clearly that both HEFN and LHEFN can achieve competitive performance compared with the existing architectures even with only the  $p_T$ ,  $\eta$ , and  $\phi$  information of all constituent particles. More importantly, both HEFN and LHEFN can reconstruct energy correlation-based observables, setting them apart from other models that fail to preserve crucial features including interpretability, IRC-safety, and rotational invariance.



|                       | Accuracy      | AUC           | $1/\epsilon_B$ ( $\epsilon_S = 0.5$ ) | $1/\epsilon_B$ ( $\epsilon_S = 0.3$ ) |
|-----------------------|---------------|---------------|---------------------------------------|---------------------------------------|
| ResNeXt-50 [20]       | 0.821         | 0.9060        | 30.9                                  | 80.8                                  |
| P-CNN [20]            | 0.827         | 0.9002        | 34.7                                  | 91.0                                  |
| PFN [18]              | -             | 0.9005        | $34.7\pm 0.4$                         | -                                     |
| ParticleNet-Lite [20] | 0.835         | 0.9079        | 37.1                                  | 94.5                                  |
| ParticleNet [20]      | 0.840         | 0.9116        | $39.8\pm 0.2$                         | $98.6\pm 1.3$                         |
| ABCNet [21]           | 0.840         | 0.9126        | $42.6\pm 0.4$                         | $118.4\pm 1.5$                        |
| SPCT [36]             | 0.815         | 0.8910        | $31.6\pm 0.3$                         | $93.0\pm 1.2$                         |
| PCT [36]              | 0.841         | 0.9140        | $43.2\pm 0.7$                         | $118.0\pm 2.2$                        |
| LorentzNet [22]       | 0.844         | 0.9156        | $42.4\pm 0.4$                         | $110.2\pm 1.3$                        |
| ParT [37]             | 0.849         | 0.9203        | $47.9\pm 0.5$                         | $129.5\pm 0.9$                        |
| <b>HEFN</b>           | <b>0.8264</b> | <b>0.9002</b> | <b><math>33.5\pm 0.3</math></b>       | <b><math>86.2\pm 0.8</math></b>       |
| <b>LHEFN</b>          | <b>0.8213</b> | <b>0.8969</b> | <b><math>31.3\pm 0.2</math></b>       | <b><math>82.5\pm 0.7</math></b>       |

**Table 2.** Performance comparison of HEFN and LHEFN with Existing Classification Algorithms on the Quark/Gluon Discrimination Dataset. The uncertainty is calculated by taking the standard deviation of 5 training runs with different random weight initialization. Note that in the HEFN and LHEFN networks, we did not utilize any PID information.

## 5 Conclusions

In this work, we introduced a novel IRC-safe deep learning framework for analyzing high-point energy flow observables in Jet Substructure analysis. We utilized a class of high-point energy flow functions from Energy Flow Polynomials as a comprehensive basis for observables, enabling hierarchical energy-weighted summation. By incorporating orthogonal polynomials to quantify the correlation between particles based on angular distance and employing neural networks as parametrized functions, we achieved enhanced flexibility in representation without compromising IRC safety.

Our approach demonstrated both interpretability and remarkable discrimination performance in top tagging dataset and quark-gluon dataset. To strike a balance between computational complexity and model performance, we efficiently reduced global summation to local aggregation and encoded multi-scale neighborhoods with one-hot encoding. This enabled the creation of Local Hierarchical Energy Flow Networks (LHEFN) based on local energy flow, showing comparable performance to the HEFN.

Overall, our IRC-safe deep learning framework can provide a simple yet powerful tool for analyzing high-point energy flow observables in Jet Substructure analysis, with potential applications in various high-energy collision scenarios. The combination of comprehensive observables, orthogonal polynomials, and neural networks allows us to achieve excellent performance while maintaining IRC safety, rotation invariance and interpretability, making it a valuable addition to the field of jet substructure analysis.

Furthermore, there are several potential improvements that could be explored in future research. Firstly, in this study we relaxed the Lorentz invariance to  $SO(2)$ -rotational symmetry around the jet axis. To ensure strict preservation of Lorentz symmetry, we could consider replacing the angular distance  $\theta_{ij}$  with a Lorentz invariant quantity, such as

invariant mass, within the Hierarchical Energy Flow Polynomials parametrization. Secondly, for the quark/gluon discrimination task, we did not utilize the Particle Identification (PID) information. Incorporating the PID information into the model architecture could potentially lead to performance improvements, as it provides valuable additional information for distinguishing gluon jet from light quark jet. Thirdly, in our approach, we solely reconstructed the Energy Flow Polynomials observables with path graph structures. It would be interesting to explore more complex graph structures and investigate the reconstruction of EFP observables corresponding to different graph configurations, as this could offer further insights into the energy distribution within the jets. These possible improvements merit further investigation in future studies. By addressing these aspects, we can enhance the capabilities of our model and potentially uncover new insights into jet substructure analysis.

## Acknowledgments

We thank Lei Wu, Chih-Ting Lu and Jeonghyeon Song for valuable discussions. This work was supported by the National Natural Science Foundation of China (NNSFC) under grant Nos.11821505 and 12075300, by Peng-Huan-Wu Theoretical Physics Innovation Center (12047503), by the CAS Center for Excellence in Particle Physics (CCEPP), and by the Key Research Program of the Chinese Academy of Sciences, Grant NO. XDPB15.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] 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](#)].
- [2] G. Gur-Ari, M. Papucci and G. Perez, *Classification of Energy Flow Observables in Narrow Jets*, [arXiv:1101.2905](#).
- [3] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with  $N$ -subjettiness*, *JHEP* **03** (2011) 015 [[arXiv:1011.2268](#)] [[INSPIRE](#)].
- [4] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing  $N$ -subjettiness*, *JHEP* **02** (2012) 093 [[arXiv:1108.2701](#)] [[INSPIRE](#)].
- [5] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [[arXiv:1305.0007](#)] [[INSPIRE](#)].
- [6] I. Moutl, L. Necib and J. Thaler, *New Angles on Energy Correlation Functions*, *JHEP* **12** (2016) 153 [[arXiv:1609.07483](#)] [[INSPIRE](#)].
- [7] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow polynomials: A complete linear basis for jet substructure*, *JHEP* **04** (2018) 013 [[arXiv:1712.07124](#)] [[INSPIRE](#)].
- [8] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, Springer (2019) [[DOI:10.1007/978-3-030-15709-8](#)] [[INSPIRE](#)].

- [9] C.F. Berger and L. Magnea, *Scaling of power corrections for angularities from dressed gluon exponentiation*, *Phys. Rev. D* **70** (2004) 094010 [[hep-ph/0407024](#)] [[INSPIRE](#)].
- [10] L.G. Almeida et al., *Substructure of high- $p_T$  Jets at the LHC*, *Phys. Rev. D* **79** (2009) 074017 [[arXiv:0807.0234](#)] [[INSPIRE](#)].
- [11] K. Albertsson et al., *Machine Learning in High Energy Physics Community White Paper*, *J. Phys. Conf. Ser.* **1085** (2018) 022008 [[arXiv:1807.02876](#)] [[INSPIRE](#)].
- [12] M. Abdughani et al., *Supervised deep learning in high energy phenomenology: a mini review*, *Commun. Theor. Phys.* **71** (2019) 955 [[arXiv:1905.06047](#)] [[INSPIRE](#)].
- [13] T. Plehn, A. Butter, B. Dillon and C. Krause, *Modern Machine Learning for LHC Physicists*, [arXiv:2211.01421](#) [[INSPIRE](#)].
- [14] T. Cheng, *Bridging Machine Learning and Sciences: Opportunities and Challenges*, [arXiv:2210.13441](#) [[INSPIRE](#)].
- [15] B.P. Roe et al., *Boosted decision trees, an alternative to artificial neural networks*, *Nucl. Instrum. Meth. A* **543** (2005) 577 [[physics/0408124](#)] [[INSPIRE](#)].
- [16] A.J. Larkoski, I. Moult and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, *Phys. Rept.* **841** (2020) 1 [[arXiv:1709.04464](#)] [[INSPIRE](#)].
- [17] S. Thais et al., *Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges*, [arXiv:2203.12852](#).
- [18] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [[arXiv:1810.05165](#)] [[INSPIRE](#)].
- [19] P. Konar, V.S. Ngairangbam and M. Spannowsky, *Energy-weighted message passing: an infra-red and collinear safe graph neural network algorithm*, *JHEP* **02** (2022) 060 [[arXiv:2109.14636](#)] [[INSPIRE](#)].
- [20] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019 [[arXiv:1902.08570](#)] [[INSPIRE](#)].
- [21] V. Mikuni and F. Canelli, *ABCNet: An attention-based method for particle tagging*, *Eur. Phys. J. Plus* **135** (2020) 463 [[arXiv:2001.05311](#)] [[INSPIRE](#)].
- [22] S. Gong et al., *An efficient Lorentz equivariant graph neural network for jet tagging*, *JHEP* **07** (2022) 030 [[arXiv:2201.08187](#)] [[INSPIRE](#)].
- [23] C.R. Qi, H. Su, K. Mo and L.J. Guibas, *Pointnet: Deep learning on point sets for 3d classification and segmentation*, [arXiv:1612.00593](#).
- [24] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov and A. Smola, *Deep Sets*, [arXiv:1703.06114](#).
- [25] G.C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in  $e^+e^-$  Annihilation and Other Processes*, *Phys. Rev. Lett.* **41** (1978) 1581 [[INSPIRE](#)].
- [26] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073 [[arXiv:1704.08249](#)] [[INSPIRE](#)].
- [27] J. Gallicchio and M.D. Schwartz, *Seeing in Color: Jet Superstructure*, *Phys. Rev. Lett.* **105** (2010) 022001 [[arXiv:1001.5027](#)] [[INSPIRE](#)].

- [28] P. Cal, J. Thaler and W. J. Waalewijn, *Power counting energy flow polynomials*, *JHEP* **09** (2022) 021 [[arXiv:2205.06818](https://arxiv.org/abs/2205.06818)].
- [29] C.R. Qi, L. Yi, H. Su and L.J. Guibas, *PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space*, <https://arxiv.org/abs/1706.02413>.
- [30] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) [[INSPIRE](#)].
- [31] L. Benato et al., *Shared Data and Algorithms for Deep Learning in Fundamental Physics*, *Comput. Softw. Big Sci.* **6** (2022) 9 [[arXiv:2107.00656](https://arxiv.org/abs/2107.00656)] [[INSPIRE](#)].
- [32] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](https://arxiv.org/abs/1410.3012)] [[INSPIRE](#)].
- [33] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[arXiv:1307.6346](https://arxiv.org/abs/1307.6346)] [[INSPIRE](#)].
- [34] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](https://arxiv.org/abs/0802.1189)] [[INSPIRE](#)].
- [35] E.A. Moreno et al., *JEDI-net: a jet identification algorithm based on interaction networks*, *Eur. Phys. J. C* **80** (2020) 58 [[arXiv:1908.05318](https://arxiv.org/abs/1908.05318)] [[INSPIRE](#)].
- [36] V. Mikuni and F. Canelli, *Point cloud transformers applied to collider physics*, *Mach. Learn. Sci. Tech.* **2** (2021) 035027 [[arXiv:2102.05073](https://arxiv.org/abs/2102.05073)] [[INSPIRE](#)].
- [37] H. Qu, C. Li and S. Qian, *Particle Transformer for Jet Tagging*, [arXiv:2202.03772](https://arxiv.org/abs/2202.03772) [[INSPIRE](#)].