# Hierarchical Energy Flow Networks for Jet Tagging

**Wei Shen**
shenwei@itp.ac.cn

Institute of Theoretical Physics
Chinese Academy of Sciences
Advisor: Prof. Jin Min Yang

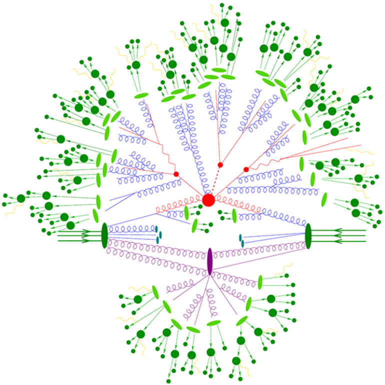
September 14, 2023

Paper: 2308.08300[Wei Shen, Daohan Wang, Jin Min Yang]

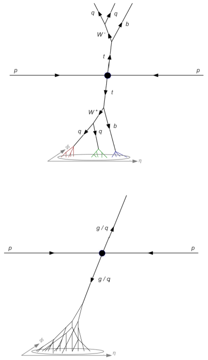
# Summary

# Introduction

- What is Jet and Jet Substructure.



Hadron-hadron collision as simulated by a Monte-Carlo event generator for high energy jets [1]



Decay sequences in $t\bar{t}$ and dijet QCD events[2]

# Calorimetric Correlators

Calorimetric Correlators[3]

- Infrared and collinear (IRC) safe

**Infrared and Collinear safety.**

Any well-defined jet observable should remain consistent after a particle is split collinearly or a soft particle is emitted.

- Collinear splittings and soft emissions occur randomly through the entire process of jet production.
- For accurate fixed-order perturbative QCD calculation.
- Robust for uncertainty of experimental detectors.

# Calorimetric Correlators

Calorimetric Correlators[3]

- Infrared and collinear (IRC) safe
- Lorentz invariant

## Lorentz invariant

Only limited observables are Lorentz invariant, for example, jet mass.

- For highly-boosted and narrow jets, the four momentum of jet and the shape of jet can be separated.
- The Lorentz invariant could be relaxed to SO(2) rotation symmetry around the jet direction.
- Decomposed into combination of energy-weighted particles pairwise correlators.

# Calorimetric Correlators

Calorimetric Correlators[3]

- Infrared and collinear (IRC) safe
- Lorentz invariant
- Permutation symmetry

## C-Correlators

Jet observable as a linear combination of C-correlators:

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

$E_i$ and $\hat{p}_i$ are respectively energy (or energy fraction) and directions of the i-th constituent particle within the jet

# Energy Correlation Functions

Energy Correlation Functions [4]

- Multiply product of angular distance between particles as $f_N$.

## energy correlation functions

$$\mathrm{ECF}(2,\beta) = \sum_{i<j\in J} E_i E_j R_{ij}^{\beta}$$

$$\mathrm{ECF}(3,\beta) = \sum_{i<j<k\in J} E_i E_j E_k R_{ij}^{\beta} R_{jk}^{\beta} R_{ki}^{\beta}$$

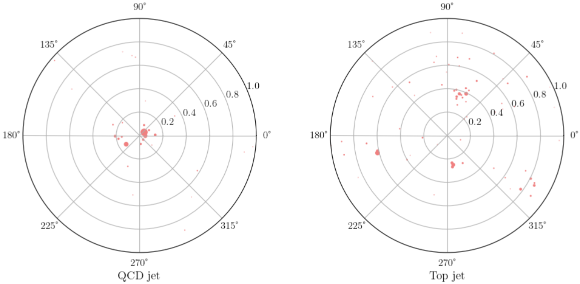
$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} E_{i_a}\right)\left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} R_{i_b i_c}\right)^{\beta}$$

where $R_{i_b i_c}$ is the angular distance between particles $i_b$ and $i_c$, $\beta$ is angular exponent. Computational complexity of $\mathrm{ECF}(N,\beta)$ of $M$ particles are $M^N/N!$.

# Energy Correlation Functions

Energy Correlation Functions [4]

- Multiply product of angular distance between particles as $f_N$.
- N-prong structure

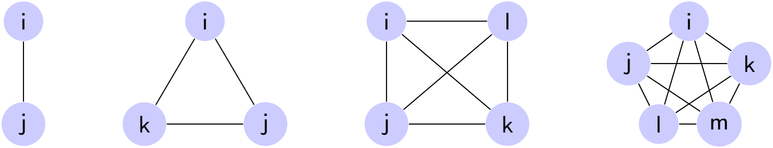



QCD jet and top jet internal radiation patterns show significant differences in the above visual representation, from a quantitative and statistical standpoint, $r_N^\beta \equiv \text{ECF}(N+1,\beta)/\text{ECF}(N,\beta)$ go to zero for $N$-subjets system.

# Energy Flow Polynomials

Energy flow polynomials[5]

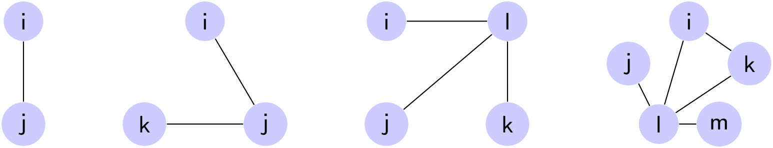
- Connected Graph Correspondence of ECFs



# Energy Flow Polynomials

Energy flow polynomials[5]

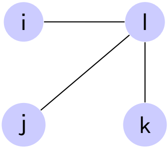
- Connected Graph Correspondence of ECFs
- More Graph Structures and Energy Flow Polynomials



# Energy Flow Polynomials

Energy flow polynomials[5]

- Connected Graph Correspondence of ECFs
- More Graph Structures and Energy Flow Polynomials
- Computational complexity analysis




**EFP$_G$**

$$\mathrm{EFP}_G = \sum_{i,j,k,l}^{M} E_i E_j E_k E_l R_{il}^{\beta} R_{jl}^{\beta} R_{kl}^{\beta}$$
$$= \sum_{l} E_l \left(\sum_i E_i R_{il}^{\beta}\right)\left(\sum_j E_j R_{jl}^{\beta}\right)\left(\sum_k E_k R_{kl}^{\beta}\right)$$

# Hierarchical Energy Flow

## Hierarchical Energy Flow [arxiv: 2308.08300]

- Embed Pairwise Correlation via Orthogonal Polynomials

### Legendre Expanded Correlators

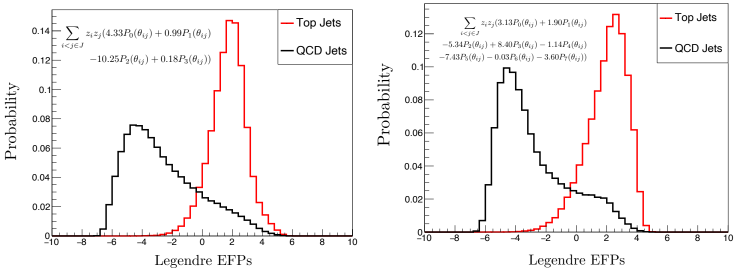
$$\sum_{i,j}^{M} E_i E_j f(R_{ij}) = \sum_{\beta=0}^{\beta_{max}} \alpha_\beta \sum_{i,j}^{M} E_i E_j \mathsf{P}_\beta(\theta_{ij})$$

where $\theta_{ij} = R_{ij}/R_0 - 1$ to make $\theta_{ij} \in [-1, +1]$ for numerical stability. We will abbreviate $\mathsf{P}_\beta(\theta_{ab})$ as $\mathsf{P}_{ab}$ in the following.

# Hierarchical Energy Flow

## Hierarchical Energy Flow [arxiv: 2308.08300]

- Embed Pairwise Correlation via Orthogonal Polynomials
- Linear Logical Regression

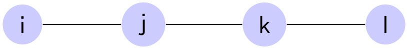



Distribution of Legendre Energy Correlation Function of Top and QCD jets.

# Hierarchical Energy Flow

## Hierarchical Energy Flow [arxiv: 2308.08300]

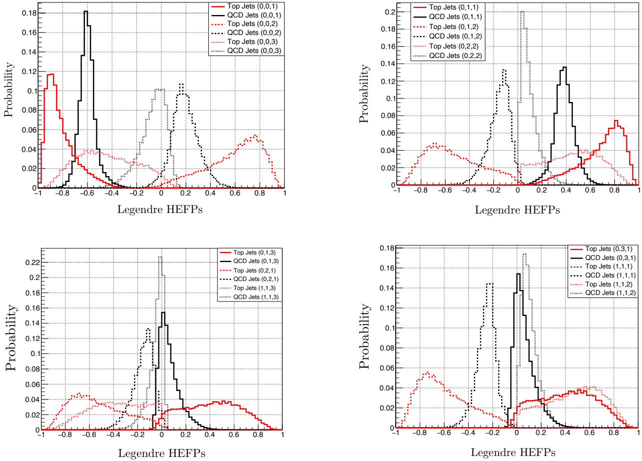
- Embed Pairwise Correlation via Orthogonal Polynomials
- Linear Logical Regression
- Hierarchical IRC-Safe and Insert Non-Linear Function




### $\mathsf{HEFP}_G$

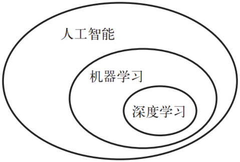
$$\mathsf{HEFP}_G = \sum_i^M E_i (\sum_j^M E_j \mathsf{P}_{ij} (\sum_k^M E_k \mathsf{P}_{jk} (\sum_l^M E_l \mathsf{P}_{kl})))$$

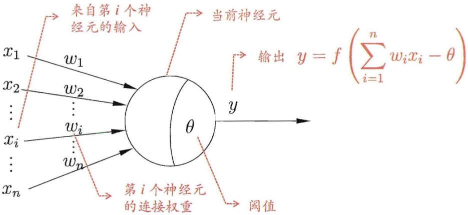
$$\rightarrow \sum_i^M E_i \Phi_3 (\sum_j^M E_j \mathsf{P}_{ij} \Phi_2 (\sum_k^M E_k \mathsf{P}_{jk} \Phi_1 (\sum_l^M E_l \mathsf{P}_{kl})))$$

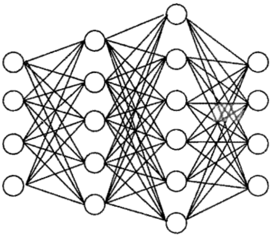


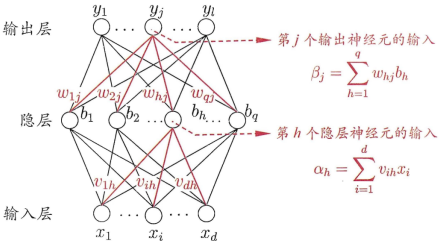

Distributions of 4-point Legendre Hierarchical Energy Flow polynomials HEFP(4, $\{\beta_1, \beta_2, \beta_3\}$) under various settings of ($\{\beta_1, \beta_2, \beta_3\}$) for top jets and QCD jets.

# A brief introduction to forward neural network

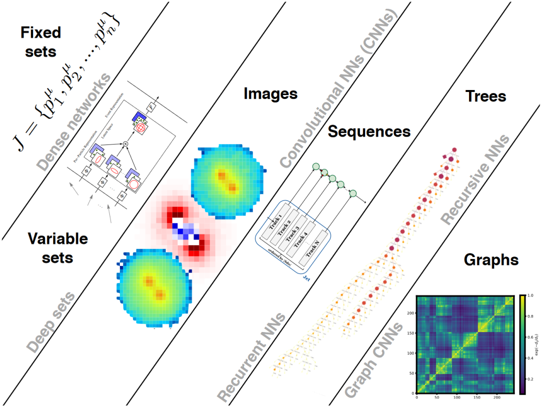




# Introduction to Neural Network for Jet




Representation of jet and neural networks architecture[6].

# Minimal Neural Network Attempt

**Algorithm 1:** Minimal trainable N-point path-graph hierarchical energy flow observable of jet

**Input:** energy fraction of particles $E_i$,
Legendre embed of particles pairwise distance $\mathsf{P}_{ij}^{\beta}$

**Output:** N-point jet observable $\mathcal{O}^N$

```
Initialization t = 0, p̂_i^t = 1
for t<N do
    p̂_{i,β} ← Σ_j E_j p̂_j^t P_{ij}^β
    x_{J,β} ← Σ_i^{M_j} E_i p̂_{i,β}
    μ_{B,β} ← Σ_{J=1}^m (1/m) x_{J,β}, σ_{B,β}^2 ← (1/m) Σ_{J=1}^m (x_{J,β} − μ_{B,β})^2
    p̃_{i,β} ← (p̂_{i,β} − μ_{B,β}) / (sqrt(σ_{B,β}^2 + ε))
    p̂_i^{t+1} ← σ(p̂_i^t + Σ_β^{β_max} w_β p̃_{i,β} + b)
end
O^N ← Σ_i^M E_i p̂_i^N
```

Initialization $t = 0, \hat{p}_i^t = 1$

**for** $t<N$ **do**

$\quad \hat{p}_{i,\beta} \leftarrow \sum_j E_j \hat{p}_j^t \mathsf{P}_{ij}^{\beta}$

$\quad x_{J,\beta} \leftarrow \sum_i^{M_j} E_i \hat{p}_{i,\beta}$

$\quad \mu_{B,\beta} \leftarrow \sum_{J=1}^{m} \frac{1}{m} x_{J,\beta}, \sigma_{B,\beta}^2 \leftarrow \frac{1}{m} \sum_{J=1}^{m} (x_{J,\beta} - \mu_{B,\beta})^2$

$\quad \tilde{p}_{i,\beta} \leftarrow (\hat{p}_{i,\beta} - \mu_{B,\beta})/(\sqrt{\sigma_{B,\beta}^2 + \epsilon})$

$\quad \hat{p}_i^{t+1} \leftarrow \sigma(\hat{p}_i^t + \sum_{\beta}^{\beta_{max}} w_\beta \tilde{p}_{i,\beta} + b)$

**end**

$\mathcal{O}^N \leftarrow \sum_i^M E_i \hat{p}_i^N$

# Minimal Neural Network Attempt

- Multiple Observables Ensemble

  normalize different channel of learned observables and then weighted vote

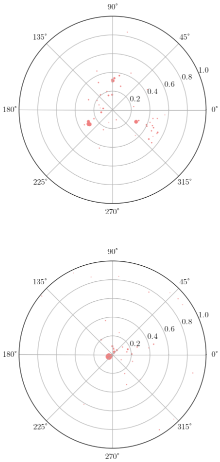
- Count Trainable Parameters $\sum_k^C E_k \times (\beta_{max} + 1) + (C + 1)$

  where $E_k$ is the number of pairwise direction correlator of $k$-th channel HEFPs (the number of edges) and $C$ is the number of channels, $\beta_{max}$ is the truncated order of Legendre polynomials.

- Tagging performance of minimal HEFN (AUC, $N = 4$, $\beta_{max} = 8$, $C = 16$)

| model | #params | IRC-safe | Top vs QCD | Quark vs Gluon |
|---|---|---|---|---|
| EFN[7] | 86.0k | ✓ | 0.9759 | 0.8824 |
| EFP[5] | 1k | ✓ | 0.980 | 0.8919 |
| P-CNN[8] | 34.8k | | 0.9803 | 0.9002 |
| ParticleNet-Lite[8] | 26k | | 0.9844 | 0.9116 |
| PELICAN$_{25/15}$ [9] | 11k | | 0.9858 | - |
| **minimal HEFN** | **593** | ✓ | **0.9819** | **0.8967** |

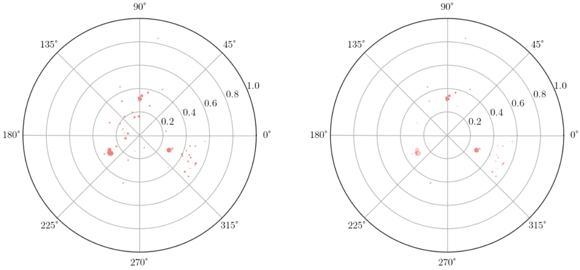
# What the machine learned?




Trained minimal HEFN knows the "importance score" $s$ of particles relative to the jet, where $s = \text{sigmoid}((\hat{p}_i^N - \sum_i E_i \hat{p}_i^N) * y)$, and y=-1 for QCD jet, y=+1 for Top jet.

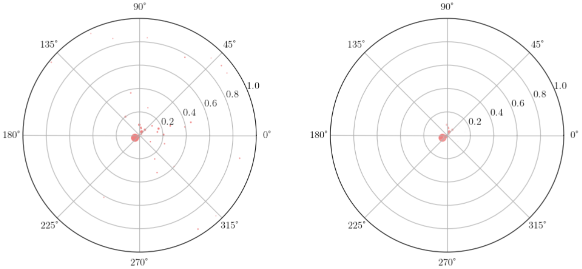
# What the machine learned?

- 3-prong substructure of Top Jet



# What the machine learned?

- 3-prong substructure of Top Jet
- Singularity structure of QCD Jet



# Hierarchical Energy Flow Nets

- More Observables
- Deeper and Wider Non-Linear $\Phi$
- Multi Layer Classifier Head
- Different $\beta_{max}$ Setting

**HEFNs correspond with path-graph**

$$\hat{p}_i^{t+1} = \sum_{k=1}^{t} \Phi^{a,k}\left(\sum_{j}^{M} z_j \Phi^{b,k}(\hat{p}_j^k) \otimes \mathsf{P}_\beta(\theta_{ij})\right)$$

$$\text{HEFN}_G(N+1) = F\left(\sum_{i}^{M} z_i \sum_{k=1}^{N} \Phi^{a,k}\left(\sum_{j}^{M} z_j \Phi^{b,k}(\hat{p}_j^k) \otimes \mathsf{P}_\beta(\theta_{ij})\right)\right)$$

# Hierarchical Energy Flow Nets Lite

- Legendre Polynomials $\rightarrow$ One-Hot Embedding
- Global Summation $\rightarrow$ Neighbor Aggregation
- Different $k_{max}, r_{max}$ Setting

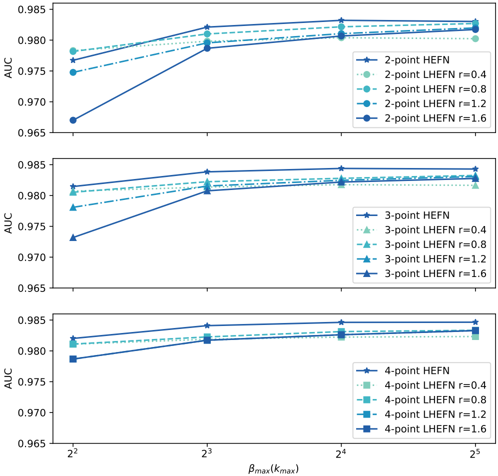
## HEFNs Lite

$$\mathsf{P}_\beta(\theta_{ij}) \rightarrow A_{ij}^k = 1 \text{ if } (k-1)\epsilon \leq R_{ij} < k\epsilon$$

$$\hat{p}_{i,k}^t = \sum_j^{A_{ij}^k} z_j \hat{p}_j^t$$

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

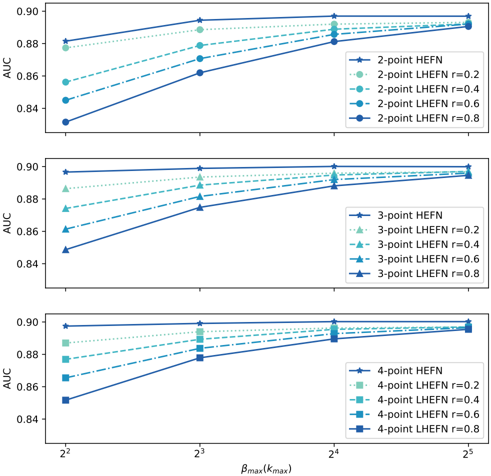
# Top vs QCD


0.985
0.980
0.975
0.970
0.965
AUC
2-point HEFN
2-point LHEFN r=0.4
2-point LHEFN r=0.8
2-point LHEFN r=1.2
2-point LHEFN r=1.6
3-point HEFN
3-point LHEFN r=0.4
3-point LHEFN r=0.8
3-point LHEFN r=1.2
3-point LHEFN r=1.6
4-point HEFN
4-point LHEFN r=0.4
4-point LHEFN r=0.8
4-point LHEFN r=1.2
4-point LHEFN r=1.6
$2^2$
$2^3$
$2^4$
$2^5$
$\beta_{max}(k_{max})$

# Top vs QCD

| | Accuracy | AUC | $1/\epsilon_B$ $(\epsilon_S = 0.5)$ | $1/\epsilon_B$ $(\epsilon_S = 0.3)$ |
|---|---|---|---|---|
| ResNeXt-50 [10] | 0.936 | 0.9837 | 302±5 | 1147±58 |
| P-CNN [10] | 0.930 | 0.9803 | 201±4 | 759±24 |
| PFN [11] | - | 0.9819 | 247±3 | 888±17 |
| ParticleNet-Lite [10] | 0.937 | 0.9844 | 325±5 | 1262±49 |
| ParticleNet [10] | 0.940 | 0.9858 | 397±7 | 1615±93 |
| JEDI-net [12] | 0.9263 | 0.9786 | - | 590.4 |
| JEDI-net with $\sum O$ [12] | 0.9300 | 0.9807 | - | 774.6 |
| SPCT [13] | 0.928 | 0.9799 | 201±9 | 725±54 |
| PCT [13] | 0.940 | 0.9855 | 392±7 | 1533±101 |
| LorentzNet [14] | 0.942 | 0.9868 | 498±18 | 2195±173 |
| ParT [15] | 0.940 | 0.9858 | 413±16 | 1602±81 |
| **HEFN** | **0.9375** | **0.9846** | **343±6** | **1262±51** |
| **LHEFN** | **0.9337** | **0.9833** | **271±5** | **935±21** |

# Quark vs Gluon



# Quark vs Gluon

| | Accuracy | AUC | $1/\epsilon_B$ $(\epsilon_S = 0.5)$ | $1/\epsilon_B$ $(\epsilon_S = 0.3)$ |
|---|---|---|---|---|
| ResNeXt-50 [10] | 0.821 | 0.9060 | 30.9 | 80.8 |
| P-CNN [10] | 0.827 | 0.9002 | 34.7 | 91.0 |
| PFN [11] | - | 0.9005 | 34.7$\pm$0.4 | - |
| ParticleNet-Lite [10] | 0.835 | 0.9079 | 37.1 | 94.5 |
| ParticleNet [10] | 0.840 | 0.9116 | 39.8$\pm$0.2 | 98.6$\pm$1.3 |
| ABCNet [16] | 0.840 | 0.9126 | 42.6$\pm$0.4 | 118.4$\pm$1.5 |
| SPCT [13] | 0.815 | 0.8910 | 31.6$\pm$0.3 | 93.0$\pm$1.2 |
| PCT [13] | 0.841 | 0.9140 | 43.2$\pm$0.7 | 118.0$\pm$2.2 |
| LorentzNet [14] | 0.844 | 0.9156 | 42.4$\pm$0.4 | 110.2$\pm$1.3 |
| ParT [15] | 0.849 | 0.9203 | 47.9$\pm$0.5 | 129.5$\pm$0.9 |
| **HEFN** | **0.8264** | **0.9002** | **33.5$\pm$0.3** | **86.2$\pm$0.8** |
| **LHEFN** | **0.8213** | **0.8969** | **31.3$\pm$0.2** | **82.5$\pm$0.7** |

# Conclusion

## conclusion

- IRC-safe, Rotational Invariant, Permutation Symmetry
- Effcient and Explainablity
- Great Potential to Explore

# Perspectives

## Technology Iteration

- More Graph Structures
- More Efficient Design
- Memory Management
- Other Machine Learning Methods
- ...

## Application Perspective

- Lorentz Invariant
- Particles Class Information
- Event-Level: Jet Charge, Jet Pull, ...
- Jet Substructure Calculation
- ...

# Reference

[1] S. Höche, *Introduction to parton-shower event generators*, 2015. arXiv: 1411.4085 [hep-ph].

[2] J. Thaler and K. V. Tilburg, "Identifying boosted objects with n-subjettiness," *Journal of High Energy Physics*, vol. 2011, no. 3, 2011. DOI: 10.1007/jhep03(2011)015. [Online]. Available: https://doi.org/10.1007%2Fjhep03%282011%29015.

[3] F. V. Tkachov, "Measuring multijet structure of hadronic energy flow or, what is a jet?" *International Journal of Modern Physics A*, vol. 12, no. 30, pp. 5411–5529, 1997. DOI: 10.1142/s0217751x97002899. [Online]. Available: https://doi.org/10.1142%2Fs0217751x97002899.

[4] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy correlation functions for jet substructure," *Journal of High Energy Physics*, vol. 2013, no. 6, 2013. DOI: 10.1007/jhep06(2013)108. [Online]. Available: https://doi.org/10.1007%2Fjhep06%282013%29108.

[5] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy flow polynomials: A complete linear basis for jet substructure," *Journal of High Energy Physics*, vol. 2018, no. 4, 2018. DOI: 10.1007/jhep04(2018)013. [Online]. Available: https://doi.org/10.1007%2Fjhep04%282018%29013.

[6] 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.*, vol. 841, pp. 1–63, 2020. DOI: 10.1016/j.physrep.2019.11.001. arXiv: 1709.04464 [hep-ph].

[7] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy flow networks: Deep sets for particle jets," *Journal of High Energy Physics*, vol. 2019, no. 1, 2019. DOI: 10.1007/jhep01(2019)121. [Online]. Available: https://doi.org/10.1007%2Fjhep01%282019%29121.

[8] H. Qu and L. Gouskos, "Jet tagging via particle clouds," *Physical Review D*, vol. 101, no. 5, 2020. DOI: 10.1103/physrevd.101.056019. [Online]. Available: https://doi.org/10.1103%2Fphysrevd.101.056019.

[9] A. Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, and X. Liu, *Explainable equivariant neural networks for particle physics: Pelican*, 2023. arXiv: 2307.16506 [hep-ph].

[10] H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds," *Phys. Rev. D*, vol. 101, no. 5, p. 056 019, 2020. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[11] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy Flow Networks: Deep Sets for Particle Jets," *JHEP*, vol. 01, p. 121, 2019. DOI: 10.1007/JHEP01(2019)121. arXiv: 1810.05165 [hep-ph].

[12] E. A. Moreno, O. Cerri, J. M. Duarte, *et al.*, "JEDI-net: a jet identification algorithm based on interaction networks," *Eur. Phys. J. C*, vol. 80, no. 1, p. 58, 2020. DOI: 10.1140/epjc/s10052-020-7608-4. arXiv: 1908.05318 [hep-ex].

[13] V. Mikuni and F. Canelli, "Point cloud transformers applied to collider physics," *Mach. Learn. Sci. Tech.*, vol. 2, no. 3, p. 035 027, 2021. DOI: 10.1088/2632-2153/ac07f6. arXiv: 2102.05073 [physics.data-an].

[14] S. Gong, Q. Meng, J. Zhang, *et al.*, "An efficient Lorentz equivariant graph neural network for jet tagging," *JHEP*, vol. 07, p. 030, 2022. DOI: 10.1007/JHEP07(2022)030. arXiv: 2201.08187 [hep-ph].

[15] H. Qu, C. Li, and S. Qian, "Particle Transformer for Jet Tagging,", Feb. 2022. arXiv: 2202.03772 [hep-ph].

[16] V. Mikuni and F. Canelli, "ABCNet: An attention-based method for particle tagging," *Eur. Phys. J. Plus*, vol. 135, no. 6, p. 463, 2020. DOI: 10.1140/epjp/s13360-020-00497-3. arXiv: 2001.05311 [physics.data-an].

# Thanks

shenwei@itp.ac.cn