LETTER • OPEN ACCESS

Boost invariant polynomials for efficient jet tagging

To cite this article: Jose M Munoz *et al* 2022 *Mach. Learn.: Sci. Technol.* **3** 04LT05

View the article online for updates and enhancements.

LETTER

Boost invariant polynomials for efficient jet tagging

Jose M Munoz[1,*] ![ORCID], Ilyes Batatia[2,3] and Christoph Ortner[4]

[1] EIA University, FTA Group, Antioquia, Colombia
[2] Engineering Laboratory, University of Cambridge, Cambridge CB2 1PZ, United Kingdom
[3] ENS Paris-Saclay, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
[4] Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2
* Author to whom any correspondence should be addressed.

E-mail: munozariasjm@hotmail.com

Abstract

Given the vast amounts of data generated by modern particle detectors, computational efficiency is essential for many data-analysis jobs in high-energy physics. We develop a new class of physically interpretable boost invariant polynomial (BIP) features for jet tagging that achieves such efficiency. We show that, for both supervised and unsupervised tasks, integrating BIPs with conventional classification techniques leads to models achieving high accuracy on jet tagging benchmarks while being orders of magnitudes faster to train and evaluate than contemporary deep learning systems.
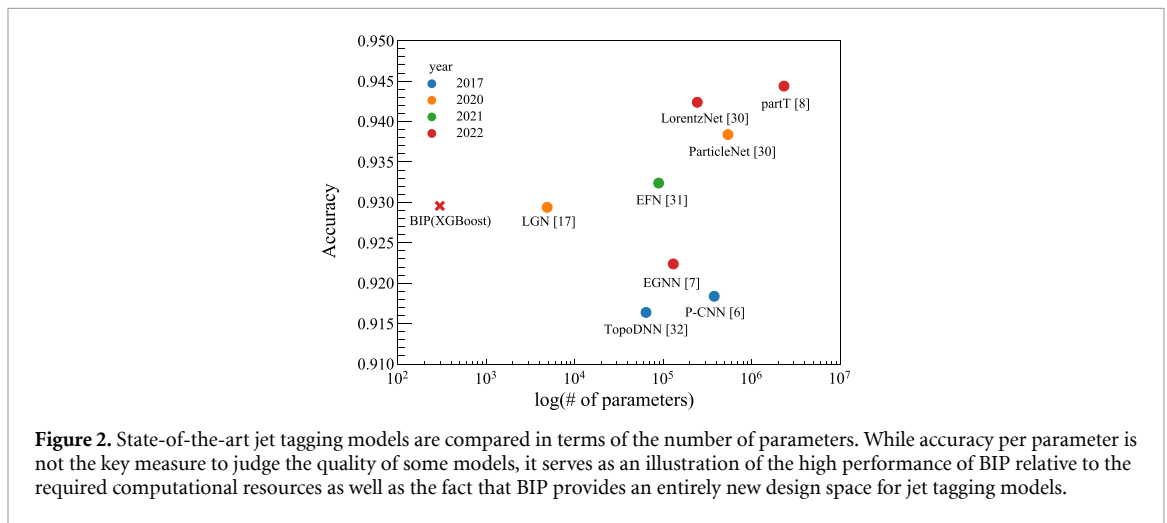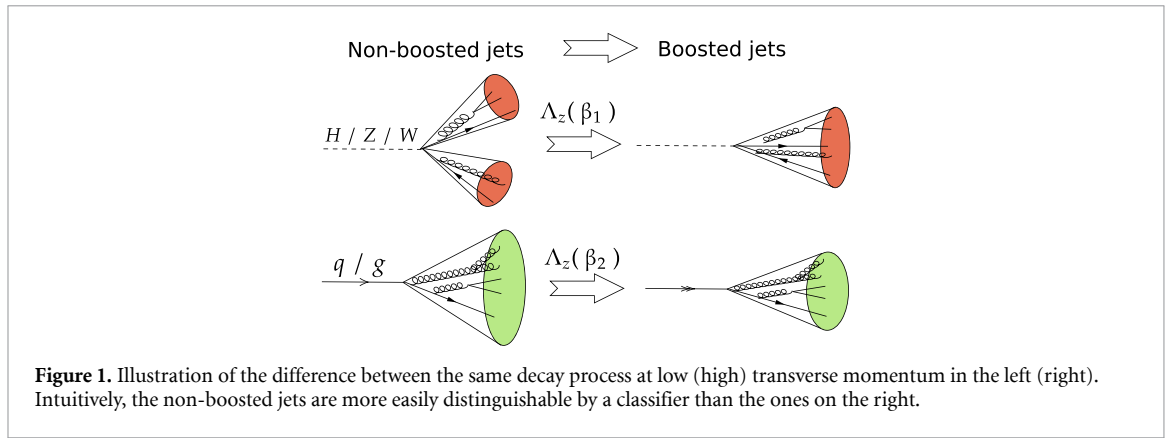
1. Introduction

The study of jets generated at particle colliders is a fundamental tool for understanding subatomic interactions and probing the standard model (SM). During experiments at high-energy colliders, a collection of the detected particles is analyzed, recording approximately 10^5 events per second, from which only a small percentage contain helpful physical information. The detection of events of interest in the myriad of observations has motivated the creation of novel algorithmic approaches to perform a classification given the particles that originated the detected shower, known as *jet tagging*. The task consists of the classification of cascades of particles generated after the beams collide. The cascade of generated events, known as the jet, comprises a set of particles described by their *four-momentum* ($E, \boldsymbol{p}$) and possibly additional features as they are reconstructed in the detectors.

The difficulty in the classification arises from the similarity in the structure of detected jets at relativistic energies as measured in the laboratory reference frame. This means that the jets substructure is not fully accessible to the detector, as it would be on the center of mass of the interactions; cf figure 1.

The first approaches to tackle the jet tagging task were clustering algorithms based on features derived from quantum-chromodynamic (QCD) theory [1–4]. Recently, a range of deep learning algorithms have been proposed, including convolutional neural networks (CNNs) [5, 6] or graph neural networks (GNNs) [7–13]. Some efforts have also turned towards the usage of physically inspired features via so-called energy flow polynomials (EFPs) [14, 15], or novel QCD-inspired features [16, 16, 17]. Currently, the most accurate machine learning approaches on jet tagging benchmarks are Lorentz group equivariant message passing networks (LE-MPNNs) [7, 18, 19]. Recently, such an injection of Lorentz invariance into the models has been shown to considerably improve the performance and data efficiency of taggers [20]. LE-MPNN models are computationally highly demanding as both symmetrization to the full Lorentz group is costly, in addition to a large number of parameters and hence the need for large amounts of training data.

In this letter, we propose the boost invariant polynomials (BIPs) a new framework for the jet tagging problem: we construct N-body polynomial features that are invariant under (a) permutations of the detected particles in the jet; (b) boosts *in the mean jet direction*; and (c) rotations around the jet mean axis. By adapting ideas from the atomic cluster expansion (ACE), [21–23] we achieve this in a computationally efficient, systematic, and general way. Since the three groups (permutations, boosts, rotations) completely

Figure 1. Illustration of the difference between the same decay process at low (high) transverse momentum in the left (right). Intuitively, the non-boosted jets are more easily distinguishable by a classifier than the ones on the right.

Figure 2. State-of-the-art jet tagging models are compared in terms of the number of parameters. While accuracy per parameter is not the key measure to judge the quality of some models, it serves as an illustration of the high performance of BIP relative to the required computational resources as well as the fact that BIP provides an entirely new design space for jet tagging models.

decouple our resulting features are particularly straightforward to derive and implement. We demonstrate the expressiveness of our novel representations by using them as input features for a range of standard classifiers, for both supervised and unsupervised learning. Our emphasis is on simplicity, for example avoiding extensive hyperparameter tuning. Nevertheless, our proposed method achieves excellent accuracy at a computational cost several orders of magnitude lower than state-of-the-art LE-MPNNs, reducing the training time on a large data set to minutes and inference time per jet to tens of microseconds, all on standard CPU hardware and with a small number of parameters (figure 2). At the same time, we maintain excellent interpretability and nearly state-of-the-art accuracy in both labeled and unlabeled tasks.

2. Methodology

2.1. Boost-invariant polynomials

2.1.1. Coordinate transform

For each (detected) particle i in the jet, experiments or simulations are able to extract their four-momentum, and possibly also additional features ξ (highly application dependent) such as the particle-id, charge, mass, or flavor. A jet can then be understood as a collection of particles $\{E_i, \boldsymbol{p}_i, \xi_i\}_{i=1}^{N}$, where N is the number of detected particles in the jet. The *mean direction* of the jet is given by the mean direction of the detected particles, i.e.

$$\boldsymbol{r}_{\text{jet}} = N^{-1} \sum_i \boldsymbol{p}_i. \tag{1}$$

We transform the spatial momentum $\boldsymbol{p}_i$ to cylindrical coordinates

$$\left(p_{\perp,i}, \varphi_i, p_{\parallel,i} \right)$$

corresponding to transverse, angular, and parallel components relative to the jet axis $\boldsymbol{r}_{\text{jet}}$. We define the angle via a Householder reflection, which we detail in the SI.

Recall that we wish to impose boost invariance in the jet direction which effectively transforms a jet into its center-of-mass frame; cf figure 1. To give the boost operation a particularly simple form, we introduce the (regularized) rapidity and transverse energy,

$$y_i = \frac{1}{2}\log\left(\frac{\delta_1 + E_i + p_{\|i}}{\delta_1 + E_i - p_{\|,i}}\right),$$

$$E_{\perp,i} = \sqrt{m_i^2 + p_{\perp,i}^2}, \tag{2}$$

where $\delta_1 > 0$ is a regularization parameter to ensure that y_i is well-defined as $E_i - p_{\|,i} \to 0$. (We take a fixed value of 10^{-4} throughout). Provided that $\mathbf{r}_{\text{jet}} \neq 0$, the mapping $(E_i, \mathbf{p}_i) \mapsto (E_{\perp,i}, p_{\perp,i}, \varphi_i, y_i)$ is injective, i.e. it is a genuine coordinate transformation. In the coordinates $(E_{\perp,i}, p_{\perp,i}, \varphi_i, y_i)$ the effect of a boost Λ_β in the direction $\hat{\mathbf{r}}_{\text{jet}}$ of the jet, and a rotation $R_{\Delta\varphi}$ about the boost axis becomes

$$R_{\Delta\varphi}\Lambda_\beta\left(E_{\perp,i}, p_{\perp,i}, \varphi_i, y_i\right)$$
$$= \left(E_{\perp,i}, p_{\perp,i}, \varphi_i + \Delta\varphi, y_i + \tanh^{-1}\beta_i\right). \tag{3}$$

Thus, we only consider the product of two one-dimensional translation groups which will make it particularly straightforward to construct invariant features in a systematic way. We refer to the SI for further explanation of this selection.

2.1.2. Many-body polynomial expansion

The ACE approach [22, 24] was proposed as a complete set of polynomial basis functions invariant to rotations and permutations to parameterize a many-body expansion of local interatomic interaction for molecular simulation. Our polynomial construction takes heavy inspiration from the ACE expansion, applying analogous techniques *globally* rather than locally, and adapting them to the different coordinate systems and symmetry groups that arise in the jet tagging context.

First, we expand the coordinates of each jet into power sum polynomial type permutation invariant features

$$A_{nlk} = \sum_{i=1}^{N} Q_n(p_{\perp,i}, E_{\perp,i}, \xi_i)e^{il\varphi_i}e^{-\lambda k y_i}, \tag{4}$$

where we have canonically chosen trigonometric and Morse polynomials (with $\lambda > 0$, though we use $\lambda = 1$ throughout) to embed the angle and rapidity, ensuring the simplest representation of the rotation and boost groups. The features $(p_{\perp,i}, E_{\perp,i}, \xi_i)$ are already invariant, hence we can embed them using a general basis Q_n, for which there is considerable design freedom. This freedom makes it possible to account for any additional detected features, ξ_i, of particles in a jet. We discuss concrete choices in § 2.3.1.

We seek permutation-, boost- and rotation-invariant features of jets. All A_{nkl} features are permutation invariant, but only the A_{n00} features are also invariant under rotations and boosts. We can generate a much richer set of permutation-, rotation- and boost-invariant polynomials by forming the product basis

$$\mathbf{A}_{\mathbf{nlk}} = \prod_{t=1}^{\nu} A_{n_t l_t k_t}, \qquad \text{where}$$

$$\mathbf{nlk} = (n_1 l_1 k_1, \ldots, n_\nu l_\nu k_\nu) \quad \text{and } \nu > 0. \tag{5}$$

Only the basis functions satisfying the constraints

$$\sum_t l_t = \sum_t k_t = 0 \tag{6}$$

which encode, respectively, rotation and boost invariance are retained. The *correlation order* ν indicates how many particles are directly interacting in such a feature. We proceed to further inspect the impact of this parameter in section 3.2.

With a suitable choice of Q_n, the $\mathbf{A}_{\mathbf{nlk}}$ features form a complete basis of invariant polynomials (cf SI), hence any (smooth) property of a jet that satisfies the same invariance can be represented to within arbitrary accuracy as a linear combination,

$$f\left(\{E_i, \mathbf{p}_i, \xi_i\}_i\right) = \sum_{\mathbf{nlk}} w_{\mathbf{nlk}}\mathbf{A}_{\mathbf{nlk}}, \tag{7}$$

where w_{nlk} are the model parameters (or, weights). We call such linear models invariant polynomials. The specific selection of feature multi-indices nlk is again application-dependent; we present a simple and general strategy in § 2.3.2. Equation (7) expresses the fact that the BIP features form a complete linear basis of the space of permutation-, boost-, and rotation- invariant set functions.

2.2. Interpretation

Two intuitive interpretations of the features (5) are related, respectively, to signal to process and to the many-body expansion. In the context of molecular simulation, analogous connections were explored in detail in [25]. For the sake of a more succinct notation, we now identify $v = (n, l, k)$ and $x_i = (E_i, \boldsymbol{p}_i, \xi_i)$.

Signal processing interpretation: Instead of a set of particles, a jet can also be identified with a density

$$\rho(x) = \sum_{i=1}^{N} \delta(x - x_i),$$

which can be thought of as a signal. Defining the one-particle basis function $\phi_v(x) = Q_n(p_\perp, E_\perp, \xi)e^{il\varphi}e^{-ky}$ the features A_v can be written as a projection of the signal onto that basis,

$$A_v = \langle \phi_v | \rho \rangle.$$

That is, the features A_v *represent* the signal ρ. Invariant representation can be obtained by taking the projected ν-correlations,

$$\langle \phi_{v_1} \otimes \ldots \otimes \phi_{v_\nu} \, | \, \rho \otimes \ldots \otimes \rho \rangle$$
$$= \prod_t \langle \phi_{v_t} | \rho \rangle = \prod_t A_{v_t} = \boldsymbol{A_v},$$

and then averaging them over rotations and boosts. In our current setting, this simply results in the constraint in equation (6). For this reason, we often call the features (5) symmetry-adapted ν-correlations.

Many-body expansion interpretation: Let f be a property of a jet that is invariant under permutations, rotations about the jet direction, and boosts in the jet direction. Then we can approximate it to within arbitrary accuracy using a many-body expansion,

$$f(\{x_i\}_i) = f_0 + \sum_i f_1(x_i) + \sum_{i_1, i_2} f_2(x_{i_1}, x_{i_2})$$
$$+ \ldots + \sum_{i_1, \ldots, i_{\bar{\nu}}} f_{\bar{\nu}}(x_{i_1}, \ldots, x_{i_{\bar{\nu}}}). \tag{8}$$

Crucially, we include self-interaction in this expansion by allowing the indices $i_1, i_2, \ldots$ to be unordered and repeated. Expanding each f_ν in terms of the tensor-product basis $\phi_{v_1} \otimes \ldots \otimes \phi_{v_\nu}$ and reorganizing the summation (see the SI in IV for the details) results exactly in equation (7) with the ν-correlation features $\boldsymbol{A_v}$ arising exactly from the expansion of f_ν. Thus, we can alternatively interpret equation (7) as an efficient linear parametrization of the many-body expansion in equation (8) and the ν-correlation features as natural basis functions for the ν-body term.
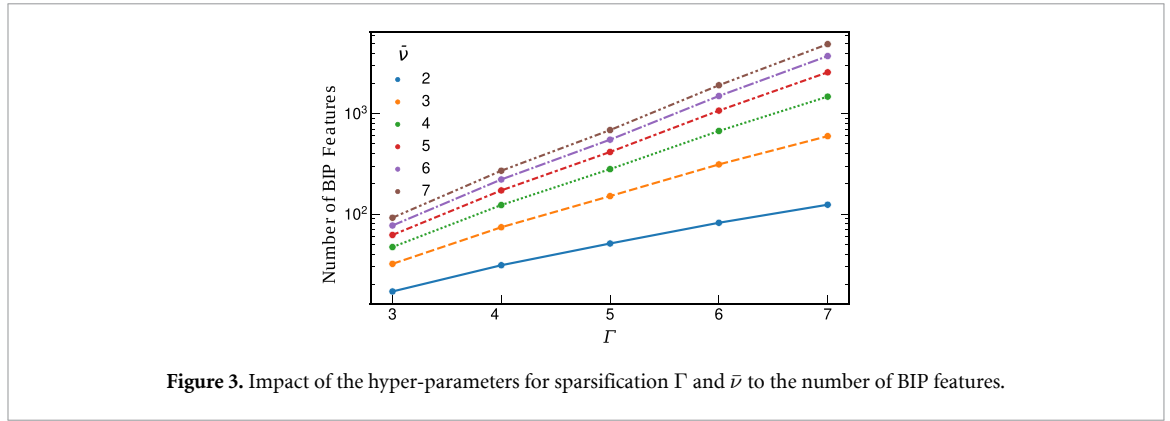
2.3. Jet tagging with BIPs

The BIP basis is a *complete linear basis* and therefore expressive enough to contain information about the jet's substructure, regardless of the frame observing the interactions. Any classification technique, either linear or nonlinear, can be used to produce a probability score. To that end, we discuss how to select a finite subset of the BIP features, and then explain how we will use them for jet tagging in supervised and non-supervised manners. We emphasize that we will employ no hyper-parameter tuning for the BIP methods that we report.

2.3.1. The invariant embedding Q_n

There is significant freedom in the design of the embedding Q_n of the invariant features $p_{\perp,i}, E_{\perp,i}$ and ξ_i, and for the most challenging data analysis tasks or in the absence of a clear intuition we advocate that it is chosen trainable, e.g. a classical multilayer perceptron (MLP). However, we found that much simpler specifications may often suffice.

We focus on the case when only the four-momentum is detected (thus ignoring ξ_i) as in the top-tagging benchmark; cf § 3.1. In this case, we choose

$$Q_n(E_{\perp,i}, p_{\perp,i}) = B_n(\tilde{p}_{\perp,i})\log(1 + E_{\perp,i}), \tag{9}$$

Figure 3. Impact of the hyper-parameters for sparsification Γ and $\bar{\nu}$ to the number of BIP features.

where B_n are the Bessel polynomials applied to the log-transverse momentum

$$\tilde{p}_{\perp,i} = A \log \left(\frac{p_{\perp,i}}{\sum_i p_{\perp,i}} + \delta_2 \right) + B. \tag{10}$$

Here, δ_2 is another regularization parameter that ensures $\tilde{p}_{\perp,j}$ remains bounded as $p_{\perp,i} \to 0$. (We choose $\delta_2 = 10^{-2}$ throughout). Also, A, B defines an affine transformation to ensure that $\tilde{p}_{\perp,i}$ belongs to the domain of orthogonality of the Bessel polynomials. The logarithmic transformation is suggested by analyzing the distribution of the $p_{\perp,i}$ in the top-tagging dataset. The factor $\log(1 + E_{\perp,i})$ imposes a form of infrared safety (cf SI) ensuring that particles with low transverse energy do not contribute significantly to the features. We explain in the SI that the resulting embedding Q_n is not theoretically complete as it does not give full flexibility to dependence on $E_{\perp,i}$. We found empirically that providing additional flexibility led to overfitting, and speculate that enough information about the energy of a particle may already be contained in the rapidity variable.

As an example, how to incorporate additional particle features ξ_i into the embedding Q_n we consider the case when the particle's charge q_i is known. We then set $\xi_i = q_i$ and incorporate it through a one-hot embedding,

$$Q_{nq}(E_{\perp,i}, p_{\perp,i}, q_i) = B_n(\tilde{p}_{\perp,i}) \log(1 + E_{\perp,i}) \delta_{q,q_i},$$

also changing n to a multi-index (n, q).

2.3.2. A priori sparse feature selection

The infinite set of possible BIP features in (5) is indexed by a high-dimensional multi-index. We use a common sparse grid technique to select which features to employ for classification tasks. First, we fix an upper bound $\bar{\nu}$ on the correlation order, which is a measure of how strongly correlated groups of particles are. Secondly, we specify a *level* Γ, which is primarily an approximation theoretic parameter. We now select all features ***nlk*** satisfying

$$\sum_{t=1}^{\nu} |l_t| + |k_t| + n_t \leqslant \Gamma \quad \text{and} \quad \nu \leqslant \bar{\nu}. \tag{11}$$

In the limit as $\bar{\nu} \to \infty$ and $\Gamma \to \infty$, we recover *all* possible features, and in this limit our model becomes universal. The choice of both parameters determines the total number of features in the basis, as shown in figure 3. We label the resulting model a BIP $(\bar{\nu}, \Gamma, \text{method})$ where 'method' stands for the technique used to create a classifier from the features and which we detail next.

2.3.3. Supervised learning

To classify jets, we arrange the BIP basis as a vector and use it as input into standard classification schemes. For each data point, the label contains the expected classification to a selection of standard classifiers without any modification from the default values in the published implementation in [26]. Our method of choice is ensemble learning (gradient boosting), but we also test BIP features in conjunction with neural networks (a standard multi-layer perceptron), and linear classifiers (with logistic regression and support vector machine). We observe that the accuracy does not change significantly between all of these classifiers. A further examination of this phenomenon and a description of the models is given in the SI § C.

Table 1. Performance comparison between BIP classifiers and a range of other classifiers taken from [7, 32, 33]. We report the results as BIP($\bar{\nu} = 3, \Gamma = 6$, architecture); in all cases, 312 BIP features are used, and the results are evaluated over 3 folds.

Architecture	#Params	Accuracy	AUC	Rej$_{30}$%
partT [8]	2.14M	0.940	0.986	1602 ± 81
EGNN [7]	120k	0.922	0.970	540 ± 49
PCT [33]	139.3k	0.940	0.986	1533 ± 101
EFN [34]	82k	0.927	0.979	888 ± 17
ParticleNet [33]	498k	0.938	0.985	1298 ± 46
LGN [19]	4.5k	0.929	0.964	435 ± 95
P-CNN [6]	348k	0.918	0.980	732 ± 24
TopoDNN [35]	59k	0.916	0.972	295 ± 5
Supervised				
BIP(MLP)	4k	0.931	0.981	853 ± 68
BIP(XGBoost)	312	0.929	0.978	600 ± 47
BIP(LogReg)	312	0.927	0.977	576 ± 34
BIP(SVM)	312	0.927	—	—
Unsupervised				
BIP(UMAP+GMM)	5	0.864	0.898	—
BIP(UMAP+KNN)	2	0.845	—	—

2.3.4. Unsupervised learning

Following proposals in [27, 28], we study unsupervised learning using the BIP method. Unsupervised learning is of interest for several reasons: it can identify deviations between observed and simulated data and could therefore be employed to detect physics beyond the SM. It avoids bias towards the detector's nuisance parameters, which heavily increase systematic uncertainties. Finally, it helps reduce the need for training data in order to perform phenomenological analyses.

Since higher-dimensional spaces tend to make distance metrics asymptotically indistinguishable [29], we first perform a uniform manifold approximation and projection (UMAP) developed in [30] for dimensionality reduction inspired by the t-distributed stochastic neighbor embedding (t-SNE) approach used in [31]. After the embedding has been projected, we show the expressiveness of this ultra-compact feature set by training a Gaussian mixture model (GMM) and a *k*-means clustering algorithm. Further details are given in the SI, § C.

3. Results

3.1. Top tagging benchmark

The *top tagging dataset* was proposed in [32]. It consists of 1.2 M training, 400 k validation, and 400 k. Each data point represents a jet whose origin is a top quark, a light quark, or a gluon; each of them, contains only the kinematic information of up to 200 particles, but a mean of only 30. The events were generated at 14 TeV, also including a simplified simulation of the ATLAS detector in Delphes. After that, the jets were processed and cuts were added on $\|\eta\| < 0.2$ and $\Delta R = 0.8$ for the top quark ones; considering only the range $p_T \in [550, 650]$ GeV.

Table 1 shows that our BIP embedding, together with a linear classifier, can reach excellent accuracy using several orders of magnitude fewer parameters resulting in a total training time of under 50 s. The unsupervised setting shows BIP's full expressivity, enabling the GMM classifier to reach excellent accuracy with only five parameters.

We show a selection of BIP model results for jet tagging, comparing them to previous approaches. We compare the classification performance of a range of models via the *accuracy, area under the curve* (AUC), and *background rejection rate at* 30% *signal efficiency* (Rej$_{30}$%) measures and contrast this against the number of parameters employed. We also present the background rejection rate as a function of signal efficiency in figure 4 for the two best-performing out-of-the-box classifiers on the BIPs features.

3.2. Versatility and efficiency

Our model is highly computationally efficient; to demonstrate this we performed a performance benchmark on an AMD EPYC-Rome Processor using a fully serial framework and with no GPU usage. The computational pipeline involves custom data processing and internal transformations explained in section 2. This stage of the computation, including the Householder transformation, takes between 1.9 and 6.3 μs per jet at 68% C.L.

The density projections A_{nlk}, as well as the ν-correlations (5), are both fast to evaluate, even without the optimal algorithm proposed in [23], meaning that no usage of a GPU is required to obtain excellent
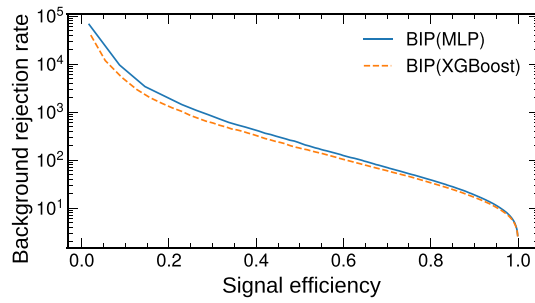
Figure 4. Comparison of the background rejection rate against signal efficiency between the MLP and the XGBoost models fitted on the BIP($\bar{\nu} = 3, \Gamma = 6,$) basis.
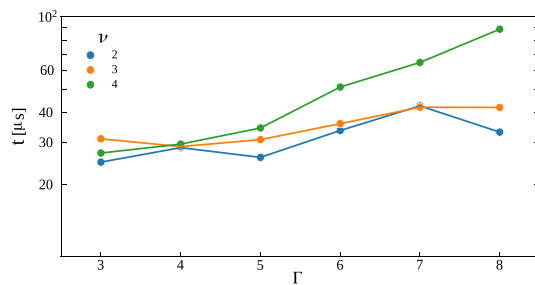
Figure 5. Performance of the construction of the BIP features at different correlation order ν and level Γ, the standard deviation is calculated by performing the transformation for all the training sets in a sequential mode with no parallelization.
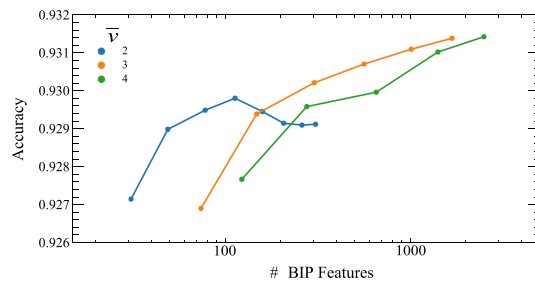
Figure 6. The accuracy as a function of the sparsification parameters, where increasing the number of features is obtained via increasing Γ. See the Text for more details.

performance, as shown in figure 5. While the number of BIP features increases rapidly with the correlation order, $\bar{\nu}$, the computational expense of embedding the jets scales only linearly; see [23] for a deeper analysis of this fact. The implementation of the BIP embeddings is publicly available [36].

The performance depends on the correlation order $\bar{\nu}$ and level Γ parameters of the BIP framework, which gives a variable basis size as shown in figure 3 and allows us to trade efficiency against the computational cost (cf figure 6). As the level and correlation order of the BIP basis increase, the jet descriptor becomes a complete linear basis, as explained theoretically in section 2.3.2. The well-known curse of dimensionality still arises but only for large correlation order ν and large levels Γ; this is a consequence of the total degree approach combined with the exploitation of symmetries and is explained in detail in [22].

As a general rule, the accuracy vs computational cost trade-off can be tuned according to the application. For instance, for trigger-level applications or prototyping for model searches, computational performance may be the main requirement, while the trade-off might be slanted toward accuracy when final statistical analyses are conducted.

4. Conclusions and outlook

We introduced the BIPs model, a systematic, interpretable, and highly efficient jet tagging architecture employing polynomial features that are invariant under permutations, rotations about the jet direction, and boosts in the jet direction. The framework draws many ideas from the ACE model [21, 22] to achieve its

generality and efficiency. While our approach does share concepts with EFPs [14] due to the fact that both are building polynomial features, our architecture is entirely different, employing different coordinate systems, invariances, and without the requirement for pairwise metrics.

We speculate that due to the simplicity of its architecture it might be easily implemented using field-programmable gate arrays as dedicated hardware for employing our framework in an experimental setting. In addition, our construction is highly versatile and enables us to easily incorporate additional measured particle properties (e.g. charge, spin, flavor).

Despite the simplicity of our approach, we achieve close to state-of-the-art accuracy while gaining several orders of magnitude of speedup in the training and inference stages. It is maybe particularly remarkable that we achieved those results without employing any hyper-parameter tuning.

There are numerous improvements to our models that can still be explored, including extensive hyperparameter tuning, as well as employing alternative classifiers. For example, preliminary experiments suggest that basic hyperparameter tuning leads to modest accuracy improvements (ca. 0.1% improved accuracy and 10% increase in background rejection). However, for bigger gains, we believe that more significant changes to the proposed architecture of the models are needed. For example, a promising pathway is to extend our BIPs model to the recently proposed framework for equivariant higher-order message passing [37, 38] that has proven highly successful for modeling inter-atomic interactions [39, 40], typically outperforming other approaches despite employing much shallower architectures. This work suggests that the BIP model could also be extended to a geometric deep-learning framework which would naturally lead to the automated discovery of the embedding Q_n, and generally open up further model tuning possibilities that will likely significantly improve the already excellent accuracy we obtain with BIP models.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://zenodo.org/record/7271316.

Acknowledgments

ORCID iD

Jose M Munoz https://orcid.org/0000-0001-7740-2866

References

[1] Salam G P 2010 *Eur. Phys. J. C* **67** 637
[2] Larkoski A J, Moult I and Nachman B 2020 *Phys. Rept.* **841** 1
[3] Dasgupta M, Guzzi M, Rawling J and Soyez G 2018 *J. High Energy Phys.* JHEP09(2018)170
[4] Farhi E 1977 *Phys. Rev. Lett.* **39** 1587
[5] Komiske P T, Metodiev E M and Schwartz M D 2017 *J. High Energy Phys.* JHEP01(2017)110
[6] Macaluso S and Shih D 2018 *J. High Energy Phys.* JHEP10(2018)121
[7] Gong S, Meng Q, Zhang J, Qu H, Li C, Qian S, Du W, Ma Z-M and Liu T-Y 2022 An efficient Lorentz equivariant graph neural network for jet tagging *J. High Energy Phys.* JHEP07(2022)030
[8] Qu H, Li C and Qian S 2022 Particle transformer for jet tagging *Int. Conf. on Machine Learning* vol 162
[9] Shimmin C 2021 arXiv:2107.02908 [hep-ph]
[10] Ju X and Nachman B 2020 *Phys. Rev. D* **102** 075014
[11] Mikuni V and Canelli F 2020 *Eur. Phys. J. Plus* **135** 463
[12] Moreno E A, Cerri O, Duarte J M, Newman H B, Nguyen T Q, Periwal A, Pierini M, Serikova A, Spiropulu M and Vlimant J-R 2020 *Eur. Phys. J. C* **80** 58
[13] Moreno E A, Nguyen T Q, Vlimant J-R, Cerri O, Newman H B, Periwal A, Spiropulu M, Duarte J M and Pierini M 2020 *Phys. Rev. D* **102** 012010
[14] Komiske P T, Metodiev E M and Thaler J 2018 *J. High Energy Phys.* JHEP04(2018)013
[15] Romero A, Whiteson D, Fenton M, Collado J and Baldi P 2021 Safety of Quark/Gluon jet classification (arXiv:2103.09103) [hep-ph]
[16] Fedkevych O, Khosa C K, Marzani S and Sforza F 2022 Identification of b-jets using QCD-inspired observables (arXiv:2202.05082) [hep-ph]
[17] Khosa C K and Marzani S 2021 *Phys. Rev. D* **104** 055043
[18] Erdmann M, Geiser E, Rath Y and Rieger M 2019 *J. Instrum.* **14** 06006

[19] Bogatskiy A, Anderson B, Offermann J T, Roussi M, Miller D W and Kondor R 2020 Lorentz group equivariant neural network for particle physics *Int. Conf. on Machine Learning* (available at: http://proceedings.mlr.press/v119/bogatskiy20a/bogatskiy20a.pdf)

[20] Li C, Qu H, Qian S, Meng Q, Gong S, Zhang J, Liu T-Y and Li Q 2022 arXiv:2208.07814 [hep-ph]

[21] Drautz R 2019 *Phys. Rev. B* **99** 014104

[22] Bachmayr M, Csanyi G, Dusson G, Drautz R, Etter S, van der Oord C and Ortner C 2022 *J. Comp. Phys.* **454** 110946

[23] Kaliuzhnyi I and Ortner C 2022 Optimal evaluation of symmetry-adapted *n*-correlations via recursive contraction of sparse symmetric tensors (arXiv:2202.04140)

[24] Drautz R 2019 *Phys. Rev. B* **99** 014104

[25] Musil F, Grisafi A, Bartok A P, Ortner C, Csanyi G and Ceriotti M 2021 *Chem. Rev.* **121** 9759–815

[26] Pedregosa F *et al* 2011 *J. Mach. Learn. Res.* **12** 2825

[27] Komiske P T, Kryhin S and Thaler J 2022 arXiv:2205.04459 [hep-ph]

[28] Alvarez E, Spannowsky M and Szewc M 2022 *Front. Artif. Intell.* **5** 852970

[29] Thrun M C 2021 *Int. J. Comput. Intell. Appl.* **20** 2150016

[30] McInnes L, Healy J and Melville J 2018 arXiv:1802.03426

[31] Komiske P T, Metodiev E M and Thaler J 2019 *Phys. Rev. Lett.* **123** 041801

[32] Butter A, Plehn T and Winterhalder R 2019 *SciPost Phys.* **7** 014

[33] Qu H and Gouskos L 2020 *Phys. Rev. D* **101** 056019

[34] Komiske P T, Metodiev E M and Thaler J 2019 *J. High Energy Phys.* JHEP01(2019)121

[35] Pearkes J, Fedorko W, Lister A and Gay C 2017 arXiv:1704.02124 [hep-ex]

[36] Ortner C, Munoz J M and Batatia I 2022 Acesuit/bips.jl: v1.0.0 (available at: https://zenodo.org/record/7271316)

[37] Batatia I, Batzner S, Kovács D P, Musaelian A, Simm G N C, Drautz R, Ortner C, Kozinsky B and Csányi G 2022 The design space of e(3)-equivariant atom-centered interatomic potentials (arXiv:2205.06643)

[38] Bochkarev A, Lysogorskiy Y, Ortner C, Csányi G and Drautz R 2022 Multilayer atomic cluster expansion for semi-local interactions *Phys. Rev. Res.* **4** L042019

[39] Kovács D P, van der Oord C, Kucera J, Allen A E A, Cole D J, Ortner C and Csányi G 2021 *J. Chem. Theory Comput.* **17** 7696

[40] Batatia I, Kovács D P, Simm G N, Ortner C, and Csányi G 2022 arXiv:2206.07697